

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
29 July 2010 (29.07.2010)

PCT

(10) International Publication Number
WO 2010/085361 A2

(51) International Patent Classification:
H04N 7/24 (2006.01)

(21) International Application Number:
PCT/US2010/000194

(22) International Filing Date:
26 January 2010 (26.01.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/205,938 26 January 2009 (26.01.2009) US
61/269,955 1 July 2009 (01.07.2009) US

(71) Applicant (for all designated States except US): THOMSON LICENSING [FR/FR]; 1 rue Jeanne d'Arc, F-92443 Issy-les moulineaux (FR).

(72) Inventors; and

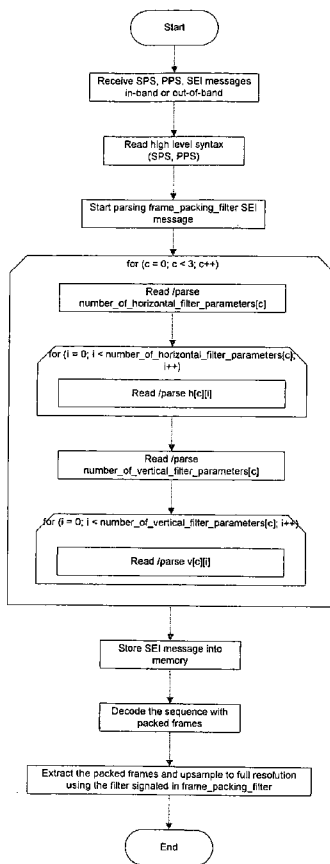
(75) Inventors/Applicants (for US only): TIAN, Dong [CN/US]; 49 Thoreau Drive, Plainsboro, NJ 08536 (US). LAI, Polin [US/US]; 4906 Quail Ridge Drive, Plainsboro, NJ 08536 (US). LUO, Jiancong [CN/US]; 3404 Hunters Glen Drive, Plainsboro, NJ 08536 (US). YIN, Peng [CN/US]; 6 John Street, Ithaca, NY 14850 (US).

(74) Agents: SHEDD, Robert, D. et al.; Thomson Licensing LLC, Two Independence Way, Suite # 200, Princeton, NJ 08540 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,

[Continued on next page]

(54) Title: FRAME PACKING FOR VIDEO CODING



(57) Abstract: Implementations are provided that relate, for example, to view tiling in video encoding and decoding. A particular implementation accesses a video picture that includes multiple pictures combined into a single picture, and accesses additional information indicating how the multiple pictures in the accessed video picture are combined. The accessed information includes spatial interleaving information and sampling information. Another implementation encodes a video picture that includes multiple pictures combined into a single picture, and generates information indicating how the multiple pictures in the accessed video picture are combined. The generated information includes spatial interleaving information and sampling information. A bitstream is formed that includes the encoded video picture and the generated information. Another implementation provides a data structure for transmitting the generated information.

Figure 36. Decoding flowchart (embodiment 1)

WO 2010/085361 A2

ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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.

ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

FRAME PACKING FOR VIDEO CODING

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of each of (1) U.S. Provisional Application Serial
5 No. 61/205,938, titled "*Spatial Interleaving*" and filed January 26, 2009, (Attorney Docket
PU090010), and (2) U.S. Provisional Application Serial No. 61/269,955, titled
"*Communicating Filter Parameters For Frame Packing Applications*" and filed July 1, 2009
(Attorney Docket PU090084). Each of these two applications is expressly incorporated by
reference herein in its entirety for all purposes.

10 This application is also related to International Application No. PCT/US2008/004747,
titled "*Tiling In Video Encoding And Decoding*" and having an International Filing Date of
April 11, 2008. This application is also identified as U.S. Patent Application Serial No.
12/450,829 (Attorney Docket No. PU070078).

15 TECHNICAL FIELD

Implementations are described that relate generally to the fields of video encoding
and/or decoding.

BACKGROUND

20 With the emergence of 3D displays in the market, including stereoscopic and auto
stereoscopic displays, there is a strong demand for more 3D content to be available. It is
typically a challenging task to code the 3D content usually involving multiple views and
possibly corresponding depth maps as well. Each frame of 3D content may require the
system to handle a huge amount of data. In typical 3D video applications, multiview video
25 signals need to be transmitted or stored efficiently due to limitations in transmission
bandwidth, storage limitations, and processing limitation, for example. Multiview Video
Coding (MVC) extends H.264/Advanced Video Coding (AVC) using high level syntax to
facilitate the coding of multiple views. This syntax aids in the subsequent handling of the 3D
images by image processors.

30 H.264/AVC, though designed ostensibly for 2D video, can also be used to transmit
stereo contents by exploiting a frame-packing technique. The technique of frame-packing is
presented simply as follows: on the encoder side, two views or pictures are generally
downsampled for packing into one single video frame, which is then supplied to a
H.264/AVC encoder for output as a bitstream; on the decoder side, the bitstream is decoded

and the recovered frame is then unpacked. Unpacking permits the extraction of the two original views from the recovered frame and generally involves an upsampling operation to restore the original size to each view so that the views can be rendered for display. This approach is able to be used for two or more views, such as with multi-view images or with depth information and the like.

Frame packing may rely on the existence of ancillary information associated with the frame and its views. Supplemental enhancement information (SEI) messages may be used to convey some frame-packing information. As an example, in a draft amendment of AVC, it has been proposed that an SEI message be used to inform a decoder of various spatial interleaving characteristics of a packed picture, including that the constituent pictures are formed by checkerboard spatial interleaving. By employing the SEI message, it is possible to encode the checkerboard interleaved picture of stereo video images using AVC directly. FIG. 26 shows a known example of checkerboard interleaving. To date, however, the SEI message contents and the contents of other high level syntaxes have been limited in conveying information relevant to pictures or views that have been subjected to frame packing.

SUMMARY

According to a general aspect, a video picture is encoded that includes multiple pictures combined into a single picture. Information is generated indicating how the multiple pictures in the accessed video picture are combined. The generated information includes spatial interleaving information and sampling information. The spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture. The sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution. The one or more parameters related to the upsampling filter include an indication of filtering direction. A bitstream is formed that includes the encoded video picture and the generated information. The generated information provides information for use in processing the encoded video picture.

According to another general aspect, a video signal or video structure includes an encoded picture section and a signaling section. The encoded picture section includes an encoding of a video picture, the video picture including multiple pictures combined into a single picture. The signaling section includes an encoding of generated information indicating how the multiple pictures in the accessed video picture are combined. The

generated information includes spatial interleaving information and sampling information. The spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture. The sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a
5 desired resolution. The one or more parameters related to the upsampling filter includes an indication of filtering direction. The generated information provides information for use in decoding the encoded video picture.

According to another general aspect, a video picture is accessed that includes multiple pictures combined into a single picture, the video picture being part of a received video
10 stream. Information is accessed that is part of the received video stream, the accessed information indicating how the multiple pictures in the accessed video picture are combined. The accessed information includes spatial interleaving information and sampling information. The spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture. The sampling information indicates one or
15 more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution. The one or more parameters related to the upsampling filter includes an indication of filtering direction. The video picture is decoded to provide a decoded representation of at least one of the multiple pictures.

The details of one or more implementations are set forth in the accompanying
20 drawings and the description below. Even if described in one particular manner, it should be clear that implementations may be configured or embodied in various manners. For example, an implementation may be performed as a method, or embodied as an apparatus configured to perform a set of operations, or embodied as an apparatus storing instructions for performing a set of operations, or embodied in a signal. Other aspects and features will become apparent
25 from the following detailed description considered in conjunction with the accompanying drawings and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a diagram showing an example of four views tiled on a single frame.

30 FIG. 2 is a diagram showing an example of four views flipped and tiled on a single frame.

FIG. 3 shows a block diagram for an implementation of a video encoder to which the present principles may be applied.

FIG. 4 shows a block diagram for an implementation of a video decoder to which the present principles may be applied.

FIG. 5 is a flow diagram for an implementation of a method for encoding pictures for a plurality of views using the MPEG-4 AVC Standard.

5 FIG. 6 is a flow diagram for an implementation of a method for decoding pictures for a plurality of views using the MPEG-4 AVC Standard.

FIG. 7 is a flow diagram for an implementation of a method for encoding pictures for a plurality of views and depths using the MPEG-4 AVC Standard.

10 FIG. 8 is a flow diagram for an implementation of a method for decoding pictures for a plurality of views and depths using the MPEG-4 AVC Standard.

FIG. 9 is a diagram showing an example of a depth signal.

FIG. 10 is a diagram showing an example of a depth signal added as a tile.

FIG. 11 is a diagram showing an example of 5 views tiled on a single frame.

15 FIG. 12 is a block diagram for an exemplary Multi-view Video Coding (MVC) encoder to which the present principles may be applied.

FIG. 13 is a block diagram for an exemplary Multi-view Video Coding (MVC) decoder to which the present principles may be applied.

20 FIG. 14 is a flow diagram for an implementation of a method for processing pictures for a plurality of views in preparation for encoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

FIG. 15 is a flow diagram for an implementation of a method for encoding pictures for a plurality of views using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

25 FIG. 16 is a flow diagram for an implementation of a method for processing pictures for a plurality of views in preparation for decoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

FIG. 17 is a flow diagram for an implementation of a method for decoding pictures for a plurality of views using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

30 FIG. 18 is a flow diagram for an implementation of a method for processing pictures for a plurality of views and depths in preparation for encoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

FIG. 19 is a flow diagram for an implementation of a method for encoding pictures for a plurality of views and depths using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

FIG. 20 is a flow diagram for an implementation of a method for processing pictures for a plurality of views and depths in preparation for decoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

FIG. 21 is a flow diagram for an implementation of a method for decoding pictures for a plurality of views and depths using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard.

FIG. 22 is a diagram showing tiling examples at the pixel level.

FIG. 23 shows a block diagram for an implementation of a video processing device to which the present principles may be applied.

FIG. 24 shows a simplified diagram of an exemplary 3D video system.

FIG. 25 shows exemplary left and right depth maps for an image from different reference views.

FIG. 26 shows an exemplary block diagram for spatially interleaving two constituent pictures into a single picture or frame using checkerboard interleaving.

FIG. 27 shows an exemplary picture of side-by-side spatial interleaving of two constituent pictures.

FIG. 28 shows an exemplary picture of top-bottom spatial interleaving of two constituent pictures.

FIG. 29 shows an exemplary picture of row-by-row spatial interleaving of two constituent pictures.

FIG. 30 shows an exemplary picture of column-by-column spatial interleaving of two constituent pictures.

FIG. 31 shows an exemplary picture of side-by-side spatial interleaving of two constituent pictures in which the right hand picture is flipped horizontally.

FIG. 32 shows an exemplary picture of top-bottom spatial interleaving of two constituent pictures in which the bottom picture is flipped vertically.

FIG. 33 shows an exemplary interleaved picture or frame in which the constituent pictures represent a layer depth video (LDV) format.

FIG. 34 shows an exemplary interleaved picture or frame in which the constituent pictures represent a 2D plus depth format.

FIGs. 35-38 show the exemplary flowcharts of different embodiments for handling the encoding and decoding of video images using SEI messages for frame packing information.

FIG. 39 shows an exemplary video transmission system to which the present principles may be applied.

FIG. 40 shows an exemplary video receiving system to which the present principles may be applied.

FIG. 41 shows an exemplary video processing device to which the present principles may be applied.

The exemplary embodiments set out herein illustrate various embodiments, and such exemplary embodiments are not to be construed as limiting the scope of this disclosure in any manner.

DETAILED DESCRIPTION

Various implementations are directed to methods and apparatus for view tiling in video encoding and decoding. It will thus be appreciated that those skilled in the art will be able to devise various arrangements that, although not explicitly described or shown herein, embody the present principles and are included within its spirit and scope.

All examples and conditional language recited herein are intended for pedagogical purposes to aid the reader in understanding the present principles and the concepts contributed by the inventor(s) to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions.

Moreover, all statements herein reciting principles, aspects, and embodiments of the present principles, as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents developed in the future, that is, any elements developed that perform the same function, regardless of structure.

Thus, for example, it will be appreciated by those skilled in the art that the block diagrams presented herein represent conceptual views of illustrative system components and/or circuitry. Similarly, it will be appreciated that any flow charts, flow diagrams, state transition diagrams, pseudocode, and the like represent various processes which may be substantially represented in computer readable media and so executed by a computer or processor, whether or not such computer or processor is explicitly shown.

The functions of the various elements shown in the figures may be provided through the use of dedicated hardware as well as hardware capable of executing software in association with appropriate software. When provided by a processor, the functions may be provided by a single dedicated processor, by a single shared processor, or by a plurality of
5 individual processors, some of which may be shared. Moreover, explicit use of the term “processor” or “controller” should not be construed to refer exclusively to hardware capable of executing software, and may implicitly include, without limitation, digital signal processor (“DSP”) hardware, read-only memory (“ROM”) for storing software, random access memory (“RAM”), and non-volatile storage.

10 Other hardware, conventional and/or custom, may also be included in the realization of various implementations. For example, any switches shown in the figures are conceptual only. Their function may be carried out through the operation of program logic, through dedicated logic, through the interaction of program control and dedicated logic, or even manually, the particular technique being selectable by the implementer as more specifically
15 understood from the context.

In the claims hereof, any element expressed as a means for performing a specified function is intended to encompass any way of performing that function including, for example, a) a combination of circuit elements that performs that function or b) software in any form, including, therefore, firmware, microcode or the like, combined with appropriate
20 circuitry for executing that software to perform the function. The present principles as defined by such claims reside in the fact that the functionalities provided by the various recited means are combined and brought together in the manner which the claims call for. It is thus regarded that any means that can provide those functionalities are equivalent to those shown herein.

25 Reference in the specification to “one embodiment” (or “one implementation”) or “an embodiment” (or “an implementation”) of the present principles means that a particular feature, structure, characteristic, and so forth described in connection with the embodiment is included in at least one embodiment of the present principles. Thus, the appearances of the phrase “in one embodiment” or “in an embodiment” appearing in various places throughout
30 the specification are not necessarily all referring to the same embodiment.

It is to be appreciated that while one or more embodiments of the present principles are described herein with respect to the MPEG-4 AVC standard, the principles described in this application are not limited to solely this standard and, thus, may be utilized with respect to other standards, recommendations, and extensions thereof, particularly video coding

standards, recommendations, and extensions thereof, including extensions of the MPEG-4 AVC standard, while maintaining the spirit of the principles of this application.

Further, it is to be appreciated that while one or more other embodiments are described herein with respect to the multi-view video coding extension of the MPEG-4 AVC standard, the present principles are not limited to solely this extension and/or this standard and, thus, may be utilized with respect to other video coding standards, recommendations, and extensions thereof relating to multi-view video coding, while maintaining the spirit of the principles of this application. Multi-view video coding (MVC) is the compression framework for the encoding of multi-view sequences. A Multi-view Video Coding (MVC) sequence is a set of two or more video sequences that capture the same scene from a different view point.

Also, it is to be appreciated that while one or more other embodiments are described herein that use depth information with respect to video content, the principles of this application are not limited to such embodiments and, thus, other embodiments may be implemented that do not use depth information, while maintaining the spirit of the present principles.

Additionally, as used herein, "high level syntax" refers to syntax present in the bitstream that resides hierarchically above the macroblock layer. For example, high level syntax, as used herein, may refer to, but is not limited to, syntax at the slice header level, Supplemental Enhancement Information (SEI) level, Picture Parameter Set (PPS) level, Sequence Parameter Set (SPS) level, View Parameter Set (VPS), and Network Abstraction Layer (NAL) unit header level.

In the current implementation of multi-video coding (MVC) based on the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Moving Picture Experts Group-4 (MPEG-4) Part 10 Advanced Video Coding (AVC) standard/International Telecommunication Union, Telecommunication Sector (ITU-T) H.264 Recommendation (hereinafter the "MPEG-4 AVC Standard"), the reference software achieves multi-view prediction by encoding each view with a single encoder and taking into consideration the cross-view references. Each view is coded as a separate bitstream by the encoder in its original resolution and later all the bitstreams are combined to form a single bitstream which is then decoded. Each view produces a separate YUV decoded output.

An exemplary video system supporting the production and use of 3D images is presented schematically in FIG. 24. The content production side of the system shows image

capture by various means including, but not limited to, stereo cameras, a depth camera, multiple cameras operating simultaneously, and conversion of 2D images to 3D images. An example of depth map information (for example, Z information) captured for left and right views of the same scene is shown in FIG. 25. Each of these approaches not only captures the video image content, but some also generate certain depth information associated with the captured video images. Once processed and coded, all this information is available to be distributed, transmitted, and ultimately rendered. Meta-data is also generated with the video content for use in the subsequent rendering of the 3D video. Rendering can take place using 2D display systems or 3D displays. The 3D displays can vary from stereo displays to multi-view 3D displays as shown in the figure.

Another approach for multi-view prediction involves grouping a set of views into pseudo views. In one example of this approach, we can tile the pictures from every N views out of the total M views (sampled at the same time) on a larger frame or a super frame with possible downsampling or other operations. Turning to FIG. 1, an example of four views tiled on a single frame is indicated generally by the reference numeral 100. All four views are in their normal orientation.

Turning to FIG. 2, an example of four views flipped and tiled on a single frame is indicated generally by the reference numeral 200. The top-left view is in its normal orientation. The top-right view is flipped horizontally. The bottom-left view is flipped vertically. The bottom-right view is flipped both horizontally and vertically. Thus, if there are four views, then a picture from each view is arranged in a super-frame like a tile. This results in a single un-coded input sequence with a large resolution.

Alternatively, we can downsample the image to produce a smaller resolution. Thus, we create multiple sequences which each include different views that are tiled together. Each such sequence then forms a pseudo view, where each pseudo view includes N different tiled views. FIG. 1 shows one pseudo-view, and FIG. 2 shows another pseudo-view. These pseudo views can then be encoded using existing video coding standards such as the ISO/IEC MPEG-2 Standard and the MPEG-4 AVC Standard.

Yet another approach for multi-view coding simply involves encoding the different views independently using a new standard and, after decoding, tiling the views as required by the player.

Further, in another approach, the views can also be tiled in a pixel wise way. For example, in a super view that is composed of four views, pixel (x, y) may be from view 0,

while pixel (x+1, y) may be from view 1, pixel (x, y+1) may be from view 2, and pixel (x+1, y+1) may be from view 3.

Many displays manufacturers use such a framework of arranging or tiling different views on a single frame and then extracting the views from their respective locations and rendering them. In such cases, there is no standard way to determine if the bitstream has such a property. Thus, if a system uses the method of tiling pictures of different views in a large frame, then the method of extracting the different views is proprietary.

However, there is no standard way to determine if the bitstream has such a property. We propose high level syntax in order to facilitate the renderer or player to extract such information in order to assist in display or other post-processing. It is also possible the sub-pictures have different resolutions and some upsampling may be needed to eventually render the view. The user may want to have the method of upsample indicated in the high level syntax as well. Additionally, parameters to change the depth focus can also be transmitted.

In an embodiment, we propose a new Supplemental Enhancement Information (SEI) message for signaling multi-view information in a MPEG-4 AVC Standard compatible bitstream where each picture includes sub-pictures which belong to a different view. The embodiment is intended, for example, for the easy and convenient display of multi-view video streams on three-dimensional (3D) monitors which may use such a framework. The concept can be extended to other video coding standards and recommendations signaling such information using high level syntax.

Moreover, in an embodiment, we propose a signaling method of how to arrange views before they are sent to the multi-view video encoder and/or decoder. Advantageously, the embodiment may lead to a simplified implementation of the multi-view coding, and may benefit the coding efficiency. Certain views can be put together and form a pseudo view or super view and then the tiled super view is treated as a normal view by a common multi-view video encoder and/or decoder, for example, as per the current MPEG-4 AVC Standard based implementation of Multi-view Video Coding. A new flag shown in Table 1 is proposed in the Sequence Parameter Set (SPS) extension of multi-view video coding to signal the use of the technique of pseudo views. The embodiment is intended for the easy and convenient display of multi-view video streams on 3D monitors which may use such a framework.

Another approach for multi-view coding involves tiling the pictures from each view (sampled at the same time) on a larger frame or a super frame with a possible downsampling operation. Turning to FIG. 1, an example of four views tiled on a single frame is indicated generally by the reference numeral 100. Turning to FIG. 2, an example of four views flipped

and tiled on a single frame is indicated generally by the reference numeral 200. Thus, if there are four views, then a picture from each view is arranged in a super-frame like a tile. This results in a single un-coded input sequence with a large resolution. This signal can then be encoded using existing video coding standards such as the ISO/IEC MPEG-2 Standard and the MPEG-4 AVC Standard.

Turning to FIG. 3, a video encoder capable of performing video encoding in accordance with the MPEG-4 AVC standard is indicated generally by the reference numeral 300.

The video encoder 300 includes a frame ordering buffer 310 having an output in signal communication with a non-inverting input of a combiner 385. An output of the combiner 385 is connected in signal communication with a first input of a transformer and quantizer 325. An output of the transformer and quantizer 325 is connected in signal communication with a first input of an entropy coder 345 and a first input of an inverse transformer and inverse quantizer 350. An output of the entropy coder 345 is connected in signal communication with a first non-inverting input of a combiner 390. An output of the combiner 390 is connected in signal communication with a first input of an output buffer 335.

A first output of an encoder controller 305 is connected in signal communication with a second input of the frame ordering buffer 310, a second input of the inverse transformer and inverse quantizer 350, an input of a picture-type decision module 315, an input of a macroblock-type (MB-type) decision module 320, a second input of an intra prediction module 360, a second input of a deblocking filter 365, a first input of a motion compensator 370, a first input of a motion estimator 375, and a second input of a reference picture buffer 380.

A second output of the encoder controller 305 is connected in signal communication with a first input of a Supplemental Enhancement Information (SEI) inserter 330, a second input of the transformer and quantizer 325, a second input of the entropy coder 345, a second input of the output buffer 335, and an input of the Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) inserter 340.

A first output of the picture-type decision module 315 is connected in signal communication with a third input of a frame ordering buffer 310. A second output of the picture-type decision module 315 is connected in signal communication with a second input of a macroblock-type decision module 320.

An output of the Sequence Parameter Set (SPS) and Picture Parameter Set (PPS) inserter 340 is connected in signal communication with a third non-inverting input of the combiner 390. An output of the SEI Inserter 330 is connected in signal communication with a second non-inverting input of the combiner 390.

5 An output of the inverse quantizer and inverse transformer 350 is connected in signal communication with a first non-inverting input of a combiner 319. An output of the combiner 319 is connected in signal communication with a first input of the intra prediction module 360 and a first input of the deblocking filter 365. An output of the deblocking filter 365 is connected in signal communication with a first input of a reference picture buffer 380.

10 An output of the reference picture buffer 380 is connected in signal communication with a second input of the motion estimator 375 and with a first input of a motion compensator 370. A first output of the motion estimator 375 is connected in signal communication with a second input of the motion compensator 370. A second output of the motion estimator 375 is connected in signal communication with a third input of the entropy coder 345.

15 An output of the motion compensator 370 is connected in signal communication with a first input of a switch 397. An output of the intra prediction module 360 is connected in signal communication with a second input of the switch 397. An output of the macroblock-type decision module 320 is connected in signal communication with a third input of the switch 397 in order to provide a control input to the switch 397. The third input of the switch 397 determines whether or not the "data" input of the switch (as compared to the control input, that is, the third input) is to be provided by the motion compensator 370 or the intra prediction module 360. The output of the switch 397 is connected in signal communication with a second non-inverting input of the combiner 319 and with an inverting input of the combiner 385.

25 Inputs of the frame ordering buffer 310 and the encoder controller 105 are available as input of the encoder 300, for receiving an input picture 301. Moreover, an input of the Supplemental Enhancement Information (SEI) inserter 330 is available as an input of the encoder 300, for receiving metadata. An output of the output buffer 335 is available as an output of the encoder 300, for outputting a bitstream.

30 Turning to FIG. 4, a video decoder capable of performing video decoding in accordance with the MPEG-4 AVC standard is indicated generally by the reference numeral 400.

The video decoder 400 includes an input buffer 410 having an output connected in signal communication with a first input of the entropy decoder 445. A first output of the

entropy decoder 445 is connected in signal communication with a first input of an inverse transformer and inverse quantizer 450. An output of the inverse transformer and inverse quantizer 450 is connected in signal communication with a second non-inverting input of a combiner 425. An output of the combiner 425 is connected in signal communication with a second input of a deblocking filter 465 and a first input of an intra prediction module 460. A second output of the deblocking filter 465 is connected in signal communication with a first input of a reference picture buffer 480. An output of the reference picture buffer 480 is connected in signal communication with a second input of a motion compensator 470.

A second output of the entropy decoder 445 is connected in signal communication with a third input of the motion compensator 470 and a first input of the deblocking filter 465. A third output of the entropy decoder 445 is connected in signal communication with an input of a decoder controller 405. A first output of the decoder controller 405 is connected in signal communication with a second input of the entropy decoder 445. A second output of the decoder controller 405 is connected in signal communication with a second input of the inverse transformer and inverse quantizer 450. A third output of the decoder controller 405 is connected in signal communication with a third input of the deblocking filter 465. A fourth output of the decoder controller 405 is connected in signal communication with a second input of the intra prediction module 460, with a first input of the motion compensator 470, and with a second input of the reference picture buffer 480.

An output of the motion compensator 470 is connected in signal communication with a first input of a switch 497. An output of the intra prediction module 460 is connected in signal communication with a second input of the switch 497. An output of the switch 497 is connected in signal communication with a first non-inverting input of the combiner 425.

An input of the input buffer 410 is available as an input of the decoder 400, for receiving an input bitstream. A first output of the deblocking filter 465 is available as an output of the decoder 400, for outputting an output picture.

Turning to FIG. 5, an exemplary method for encoding pictures for a plurality of views using the MPEG-4 AVC Standard is indicated generally by the reference numeral 500.

The method 500 includes a start block 502 that passes control to a function block 504. The function block 504 arranges each view at a particular time instance as a sub-picture in tile format, and passes control to a function block 506. The function block 506 sets a syntax element `num_coded_views_minus1`, and passes control to a function block 508. The function block 508 sets syntax elements `org_pic_width_in_mbs_minus1` and `org_pic_height_in_mbs_minus1`, and passes control to a function block 510. The function

block 510 sets a variable *i* equal to zero, and passes control to a decision block 512. The decision block 512 determines whether or not the variable *i* is less than the number of views. If so, then control is passed to a function block 514. Otherwise, control is passed to a function block 524.

5 The function block 514 sets a syntax element `view_id[i]`, and passes control to a function block 516. The function block 516 sets a syntax element `num_parts[view_id[i]]`, and passes control to a function block 518. The function block 518 sets a variable *j* equal to zero, and passes control to a decision block 520. The decision block 520 determines whether or not the current value of the variable *j* is less than the current value of the syntax element
10 `num_parts[view_id[i]]`. If so, then control is passed to a function block 522. Otherwise, control is passed to a function block 528.

 The function block 522 sets the following syntax elements, increments the variable *j*, and then returns control to the decision block 520: `depth_flag[view_id[i]][j]`;
`flip_dir[view_id[i]][j]`; `loc_left_offset[view_id[i]][j]`; `loc_top_offset[view_id[i]][j]`;
15 `frame_crop_left_offset[view_id[i]][j]`; `frame_crop_right_offset[view_id[i]][j]`;
`frame_crop_top_offset[view_id[i]][j]`; and `frame_crop_bottom_offset[view_id[i]][j]`.

 The function block 528 sets a syntax element `upsample_view_flag[view_id[i]]`, and passes control to a decision block 530. The decision block 530 determines whether or not the current value of the syntax element `upsample_view_flag[view_id[i]]` is equal to one. If so,
20 then control is passed to a function block 532. Otherwise, control is passed to a decision block 534.

 The function block 532 sets a syntax element `upsample_filter[view_id[i]]`, and passes control to the decision block 534.

 The decision block 534 determines whether or not the current value of the syntax
25 element `upsample_filter[view_id[i]]` is equal to three. If so, then control is passed to a function block 536. Otherwise, control is passed to a function block 540.

 The function block 536 sets the following syntax elements and passes control to a function block 538: `vert_dim[view_id[i]]`; `hor_dim[view_id[i]]`; and `quantizer[view_id[i]]`.

 The function block 538 sets the filter coefficients for each YUV component, and
30 passes control to the function block 540.

 The function block 540 increments the variable *i*, and returns control to the decision block 512.

 The function block 524 writes these syntax elements to at least one of the Sequence Parameter Set (SPS), Picture Parameter Set (PPS), Supplemental Enhancement Information

(SEI) message, Network Abstraction Layer (NAL) unit header, and slice header, and passes control to a function block 526. The function block 526 encodes each picture using the MPEG-4 AVC Standard or other single view codec, and passes control to an end block 599.

Turning to FIG. 6, an exemplary method for decoding pictures for a plurality of views using the MPEG-4 AVC Standard is indicated generally by the reference numeral 600.

The method 600 includes a start block 602 that passes control to a function block 604. The function block 604 parses the following syntax elements from at least one of the Sequence Parameter Set (SPS), Picture Parameter Set (PPS), Supplemental Enhancement Information (SEI) message, Network Abstraction Layer (NAL) unit header, and slice header, and passes control to a function block 606. The function block 606 parses a syntax element `num_coded_views_minus1`, and passes control to a function block 608. The function block 608 parses syntax elements `org_pic_width_in_mbs_minus1` and `org_pic_height_in_mbs_minus1`, and passes control to a function block 610. The function block 610 sets a variable `i` equal to zero, and passes control to a decision block 612. The decision block 612 determines whether or not the variable `i` is less than the number of views. If so, then control is passed to a function block 614. Otherwise, control is passed to a function block 624.

The function block 614 parses a syntax element `view_id[i]`, and passes control to a function block 616. The function block 616 parses a syntax element `num_parts_minus1[view_id[i]]`, and passes control to a function block 618. The function block 618 sets a variable `j` equal to zero, and passes control to a decision block 620. The decision block 620 determines whether or not the current value of the variable `j` is less than the current value of the syntax element `num_parts[view_id[i]]`. If so, then control is passed to a function block 622. Otherwise, control is passed to a function block 628.

The function block 622 parses the following syntax elements, increments the variable `j`, and then returns control to the decision block 620: `depth_flag[view_id[i]][j]`; `flip_dir[view_id[i]][j]`; `loc_left_offset[view_id[i]][j]`; `loc_top_offset[view_id[i]][j]`; `frame_crop_left_offset[view_id[i]][j]`; `frame_crop_right_offset[view_id[i]][j]`; `frame_crop_top_offset[view_id[i]][j]`; and `frame_crop_bottom_offset[view_id[i]][j]`.

The function block 628 parses a syntax element `upsample_view_flag[view_id[i]]`, and passes control to a decision block 630. The decision block 630 determines whether or not the current value of the syntax element `upsample_view_flag[view_id[i]]` is equal to one. If so, then control is passed to a function block 632. Otherwise, control is passed to a decision block 634.

The function block 632 parses a syntax element `upsample_filter[view_id[i]]`, and passes control to the decision block 634.

The decision block 634 determines whether or not the current value of the syntax element `upsample_filter[view_id[i]]` is equal to three. If so, then control is passed to a function block 636. Otherwise, control is passed to a function block 640.

The function block 636 parses the following syntax elements and passes control to a function block 638: `vert_dim[view_id[i]]`; `hor_dim[view_id[i]]`; and `quantizer[view_id[i]]`.

The function block 638 parses the filter coefficients for each YUV component, and passes control to the function block 640.

The function block 640 increments the variable `i`, and returns control to the decision block 612.

The function block 624 decodes each picture using the MPEG-4 AVC Standard or other single view codec, and passes control to a function block 626. The function block 626 separates each view from the picture using the high level syntax, and passes control to an end block 699.

Turning to FIG. 7, an exemplary method for encoding pictures for a plurality of views and depths using the MPEG-4 AVC Standard is indicated generally by the reference numeral 700.

The method 700 includes a start block 702 that passes control to a function block 704. The function block 704 arranges each view and corresponding depth at a particular time instance as a sub-picture in tile format, and passes control to a function block 706. The function block 706 sets a syntax element `num_coded_views_minus1`, and passes control to a function block 708. The function block 708 sets syntax elements `org_pic_width_in_mbs_minus1` and `org_pic_height_in_mbs_minus1`, and passes control to a function block 710. The function block 710 sets a variable `i` equal to zero, and passes control to a decision block 712. The decision block 712 determines whether or not the variable `i` is less than the number of views. If so, then control is passed to a function block 714. Otherwise, control is passed to a function block 724.

The function block 714 sets a syntax element `view_id[i]`, and passes control to a function block 716. The function block 716 sets a syntax element `num_parts[view_id[i]]`, and passes control to a function block 718. The function block 718 sets a variable `j` equal to zero, and passes control to a decision block 720. The decision block 720 determines whether or not the current value of the variable `j` is less than the current value of the syntax element

num_parts[view_id[i]]. If so, then control is passed to a function block 722. Otherwise, control is passed to a function block 728.

The function block 722 sets the following syntax elements, increments the variable j, and then returns control to the decision block 720: depth_flag[view_id[i]][j];
5 flip_dir[view_id[i]][j]; loc_left_offset[view_id[i]][j]; loc_top_offset[view_id[i]][j];
frame_crop_left_offset[view_id[i]][j]; frame_crop_right_offset[view_id[i]][j];
frame_crop_top_offset[view_id[i]][j]; and frame_crop_bottom_offset[view_id[i]][j].

The function block 728 sets a syntax element upsample_view_flag[view_id[i]], and passes control to a decision block 730. The decision block 730 determines whether or not the
10 current value of the syntax element upsample_view_flag[view_id[i]] is equal to one. If so, then control is passed to a function block 732. Otherwise, control is passed to a decision block 734.

The function block 732 sets a syntax element upsample_filter[view_id[i]], and passes control to the decision block 734.

15 The decision block 734 determines whether or not the current value of the syntax element upsample_filter[view_id[i]] is equal to three. If so, then control is passed to a function block 736. Otherwise, control is passed to a function block 740.

The function block 736 sets the following syntax elements and passes control to a function block 738: vert_dim[view_id[i]]; hor_dim[view_id[i]]; and quantizer[view_id[i]].

20 The function block 738 sets the filter coefficients for each YUV component, and passes control to the function block 740.

The function block 740 increments the variable i, and returns control to the decision block 712.

The function block 724 writes these syntax elements to at least one of the Sequence
25 Parameter Set (SPS), Picture Parameter Set (PPS), Supplemental Enhancement Information (SEI) message, Network Abstraction Layer (NAL) unit header, and slice header, and passes control to a function block 726. The function block 726 encodes each picture using the MPEG-4 AVC Standard or other single view codec, and passes control to an end block 799.

Turning to FIG. 8, an exemplary method for decoding pictures for a plurality of views
30 and depths using the MPEG-4 AVC Standard is indicated generally by the reference numeral 800.

The method 800 includes a start block 802 that passes control to a function block 804. The function block 804 parses the following syntax elements from at least one of the Sequence Parameter Set (SPS), Picture Parameter Set (PPS), Supplemental Enhancement

Information (SEI) message, Network Abstraction Layer (NAL) unit header, and slice header, and passes control to a function block 806. The function block 806 parses a syntax element num_coded_views_minus1, and passes control to a function block 808. The function block 808 parses syntax elements org_pic_width_in_mbs_minus1 and org_pic_height_in_mbs_minus1, and passes control to a function block 810. The function block 810 sets a variable i equal to zero, and passes control to a decision block 812. The decision block 812 determines whether or not the variable i is less than the number of views. If so, then control is passed to a function block 814. Otherwise, control is passed to a function block 824.

10 The function block 814 parses a syntax element view_id[i], and passes control to a function block 816. The function block 816 parses a syntax element num_parts_minus1[view_id[i]], and passes control to a function block 818. The function block 818 sets a variable j equal to zero, and passes control to a decision block 820. The decision block 820 determines whether or not the current value of the variable j is less than the current value of the syntax element num_parts[view_id[i]]. If so, then control is passed to a function block 822. Otherwise, control is passed to a function block 828.

The function block 822 parses the following syntax elements, increments the variable j, and then returns control to the decision block 820: depth_flag[view_id[i]][j]; flip_dir[view_id[i]][j]; loc_left_offset[view_id[i]][j]; loc_top_offset[view_id[i]][j]; 20 frame_crop_left_offset[view_id[i]][j]; frame_crop_right_offset[view_id[i]][j]; frame_crop_top_offset[view_id[i]][j]; and frame_crop_bottom_offset[view_id[i]][j].

The function block 828 parses a syntax element upsample_view_flag[view_id[i]], and passes control to a decision block 830. The decision block 830 determines whether or not the current value of the syntax element upsample_view_flag[view_id[i]] is equal to one. If so, then control is passed to a function block 832. Otherwise, control is passed to a decision block 834.

The function block 832 parses a syntax element upsample_filter[view_id[i]], and passes control to the decision block 834.

30 The decision block 834 determines whether or not the current value of the syntax element upsample_filter[view_id[i]] is equal to three. If so, then control is passed to a function block 836. Otherwise, control is passed to a function block 840.

The function block 836 parses the following syntax elements and passes control to a function block 838: vert_dim[view_id[i]]; hor_dim[view_id[i]]; and quantizer[view_id[i]].

The function block 838 parses the filter coefficients for each YUV component, and passes control to the function block 840.

The function block 840 increments the variable *i*, and returns control to the decision block 812.

5 The function block 824 decodes each picture using the MPEG-4 AVC Standard or other single view codec, and passes control to a function block 826. The function block 826 separates each view and corresponding depth from the picture using the high level syntax, and passes control to a function block 827. The function block 827 potentially performs view synthesis using the extracted view and depth signals, and passes control to an end block
10 899.

With respect to the depth used in FIGs. 7 and 8, FIG. 9 shows an example of a depth signal 900, where depth is provided as a pixel value for each corresponding location of an image (not shown). Further, FIG. 10 shows an example of two depth signals included in a tile 1000. The top-right portion of tile 1000 is a depth signal having depth values
15 corresponding to the image on the top-left of tile 1000. The bottom-right portion of tile 1000 is a depth signal having depth values corresponding to the image on the bottom-left of tile 1000.

Turning to FIG. 11, an example of 5 views tiled on a single frame is indicated generally by the reference numeral 1100. The top four views are in a normal orientation.
20 The fifth view is also in a normal orientation, but is split into two portions along the bottom of tile 1100. A left-portion of the fifth view shows the "top" of the fifth view, and a right-portion of the fifth view shows the "bottom" of the fifth view.

Turning to FIG. 12, an exemplary Multi-view Video Coding (MVC) encoder is indicated generally by the reference numeral 1200. The encoder 1200 includes a combiner
25 1205 having an output connected in signal communication with an input of a transformer 1210. An output of the transformer 1210 is connected in signal communication with an input of quantizer 1215. An output of the quantizer 1215 is connected in signal communication with an input of an entropy coder 1220 and an input of an inverse quantizer 1225. An output of the inverse quantizer 1225 is connected in signal communication with an input of an
30 inverse transformer 1230. An output of the inverse transformer 1230 is connected in signal communication with a first non-inverting input of a combiner 1235. An output of the combiner 1235 is connected in signal communication with an input of an intra predictor 1245 and an input of a deblocking filter 1250. An output of the deblocking filter 1250 is connected in signal communication with an input of a reference picture store 1255 (for view

i). An output of the reference picture store 1255 is connected in signal communication with a first input of a motion compensator 1275 and a first input of a motion estimator 1280. An output of the motion estimator 1280 is connected in signal communication with a second input of the motion compensator 1275

5 An output of a reference picture store 1260 (for other views) is connected in signal communication with a first input of a disparity estimator 1270 and a first input of a disparity compensator 1265. An output of the disparity estimator 1270 is connected in signal communication with a second input of the disparity compensator 1265.

An output of the entropy decoder 1220 is available as an output of the encoder 1200.
10 A non-inverting input of the combiner 1205 is available as an input of the encoder 1200, and is connected in signal communication with a second input of the disparity estimator 1270, and a second input of the motion estimator 1280. An output of a switch 1285 is connected in signal communication with a second non-inverting input of the combiner 1235 and with an inverting input of the combiner 1205. The switch 1285 includes a first input connected in
15 signal communication with an output of the motion compensator 1275, a second input connected in signal communication with an output of the disparity compensator 1265, and a third input connected in signal communication with an output of the intra predictor 1245.

A mode decision module 1240 has an output connected to the switch 1285 for controlling which input is selected by the switch 1285.

20 Turning to FIG. 13, an exemplary Multi-view Video Coding (MVC) decoder is indicated generally by the reference numeral 1300. The decoder 1300 includes an entropy decoder 1305 having an output connected in signal communication with an input of an inverse quantizer 1310. An output of the inverse quantizer is connected in signal communication with an input of an inverse transformer 1315. An output of the inverse
25 transformer 1315 is connected in signal communication with a first non-inverting input of a combiner 1320. An output of the combiner 1320 is connected in signal communication with an input of a deblocking filter 1325 and an input of an intra predictor 1330. An output of the deblocking filter 1325 is connected in signal communication with an input of a reference picture store 1340 (for view i). An output of the reference picture store 1340 is connected in
30 signal communication with a first input of a motion compensator 1335.

An output of a reference picture store 1345 (for other views) is connected in signal communication with a first input of a disparity compensator 1350.

An input of the entropy coder 1305 is available as an input to the decoder 1300, for receiving a residue bitstream. Moreover, an input of a mode module 1360 is also available as

an input to the decoder 1300, for receiving control syntax to control which input is selected by the switch 1355. Further, a second input of the motion compensator 1335 is available as an input of the decoder 1300, for receiving motion vectors. Also, a second input of the disparity compensator 1350 is available as an input to the decoder 1300, for receiving
5 disparity vectors.

An output of a switch 1355 is connected in signal communication with a second non-inverting input of the combiner 1320. A first input of the switch 1355 is connected in signal communication with an output of the disparity compensator 1350. A second input of the switch 1355 is connected in signal communication with an output of the motion compensator
10 1335. A third input of the switch 1355 is connected in signal communication with an output of the intra predictor 1330. An output of the mode module 1360 is connected in signal communication with the switch 1355 for controlling which input is selected by the switch 1355. An output of the deblocking filter 1325 is available as an output of the decoder 1300.

Turning to FIG. 14, an exemplary method for processing pictures for a plurality of
15 views in preparation for encoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 1400.

The method 1400 includes a start block 1405 that passes control to a function block 1410. The function block 1410 arranges every N views, among a total of M views, at a
20 particular time instance as a super-picture in tile format, and passes control to a function block 1415. The function block 1415 sets a syntax element `num_coded_views_minus1`, and passes control to a function block 1420. The function block 1420 sets a syntax element `view_id[i]` for all $(\text{num_coded_views_minus1} + 1)$ views, and passes control to a function block 1425. The function block 1425 sets the inter-view reference dependency information
25 for anchor pictures, and passes control to a function block 1430. The function block 1430 sets the inter-view reference dependency information for non-anchor pictures, and passes control to a function block 1435. The function block 1435 sets a syntax element `pseudo_view_present_flag`, and passes control to a decision block 1440. The decision block 1440 determines whether or not the current value of the syntax element
30 `pseudo_view_present_flag` is equal to true. If so, then control is passed to a function block 1445. Otherwise, control is passed to an end block 1499.

The function block 1445 sets the following syntax elements, and passes control to a function block 1450: `tiling_mode`; `org_pic_width_in_mbs_minus1`; and

org_pic_height_in_mbs_minus1. The function block 1450 calls a syntax element pseudo_view_info(view_id) for each coded view, and passes control to the end block 1499.

Turning to FIG. 15, an exemplary method for encoding pictures for a plurality of views using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 1500.

The method 1500 includes a start block 1502 that has an input parameter pseudo_view_id and passes control to a function block 1504. The function block 1504 sets a syntax element num_sub_views_minus1, and passes control to a function block 1506. The function block 1506 sets a variable i equal to zero, and passes control to a decision block 1508. The decision block 1508 determines whether or not the variable i is less than the number of sub_views. If so, then control is passed to a function block 1510. Otherwise, control is passed to a function block 1520.

The function block 1510 sets a syntax element sub_view_id[i], and passes control to a function block 1512. The function block 1512 sets a syntax element num_parts_minus1[sub_view_id[i]], and passes control to a function block 1514. The function block 1514 sets a variable j equal to zero, and passes control to a decision block 1516. The decision block 1516 determines whether or not the variable j is less than the syntax element num_parts_minus1[sub_view_id[i]]. If so, then control is passed to a function block 1518. Otherwise, control is passed to a decision block 1522.

The function block 1518 sets the following syntax elements, increments the variable j, and returns control to the decision block 1516: loc_left_offset[sub_view_id[i]][j]; loc_top_offset[sub_view_id[i]][j]; frame_crop_left_offset[sub_view_id[i]][j]; frame_crop_right_offset[sub_view_id[i]][j]; frame_crop_top_offset[sub_view_id[i]][j]; and frame_crop_bottom_offset[sub_view_id[i]][j].

The function block 1520 encodes the current picture for the current view using multi-view video coding (MVC), and passes control to an end block 1599.

The decision block 1522 determines whether or not a syntax element tiling_mode is equal to zero. If so, then control is passed to a function block 1524. Otherwise, control is passed to a function block 1538.

The function block 1524 sets a syntax element flip_dir[sub_view_id[i]] and a syntax element upsample_view_flag[sub_view_id[i]], and passes control to a decision block 1526. The decision block 1526 determines whether or not the current value of the syntax element upsample_view_flag[sub_view_id[i]] is equal to one. If so, then control is passed to a function block 1528. Otherwise, control is passed to a decision block 1530.

The function block 1528 sets a syntax element `upsample_filter[sub_view_id[i]]`, and passes control to the decision block 1530. The decision block 1530 determines whether or not a value of the syntax element `upsample_filter[sub_view_id[i]]` is equal to three. If so, the control is passed to a function block 1532. Otherwise, control is passed to a function block 1536.

The function block 1532 sets the following syntax elements, and passes control to a function block 1534: `vert_dim[sub_view_id[i]]`; `hor_dim[sub_view_id[i]]`; and `quantizer[sub_view_id[i]]`. The function block 1534 sets the filter coefficients for each YUV component, and passes control to the function block 1536.

The function block 1536 increments the variable `i`, and returns control to the decision block 1508.

The function block 1538 sets a syntax element `pixel_dist_x[sub_view_id[i]]` and the syntax element `flip_dist_y[sub_view_id[i]]`, and passes control to a function block 1540. The function block 1540 sets the variable `j` equal to zero, and passes control to a decision block 1542. The decision block 1542 determines whether or not the current value of the variable `j` is less than the current value of the syntax element `num_parts[sub_view_id[i]]`. If so, then control is passed to a function block 1544. Otherwise, control is passed to the function block 1536.

The function block 1544 sets a syntax element `num_pixel_tiling_filter_coeffs_minus1[sub_view_id[i]]`, and passes control to a function block 1546. The function block 1546 sets the coefficients for all the pixel tiling filters, and passes control to the function block 1536.

Turning to FIG. 16, an exemplary method for processing pictures for a plurality of views in preparation for decoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 1600.

The method 1600 includes a start block 1605 that passes control to a function block 1615. The function block 1615 parses a syntax element `num_coded_views_minus1`, and passes control to a function block 1620. The function block 1620 parses a syntax element `view_id[i]` for all $(\text{num_coded_views_minus1} + 1)$ views, and passes control to a function block 1625. The function block 1625 parses the inter-view reference dependency information for anchor pictures, and passes control to a function block 1630. The function block 1630 parses the inter-view reference dependency information for non-anchor pictures, and passes control to a function block 1635. The function block 1635 parses a syntax

element `pseudo_view_present_flag`, and passes control to a decision block 1640. The decision block 1640 determines whether or not the current value of the syntax element `pseudo_view_present_flag` is equal to true. If so, then control is passed to a function block 1645. Otherwise, control is passed to an end block 1699.

5 The function block 1645 parses the following syntax elements, and passes control to a function block 1650: `tiling_mode`; `org_pic_width_in_mbs_minus1`; and `org_pic_height_in_mbs_minus1`. The function block 1650 calls a syntax element `pseudo_view_info(view_id)` for each coded view, and passes control to the end block 1699.

Turning to FIG. 17, an exemplary method for decoding pictures for a plurality of
10 views using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 1700.

The method 1700 includes a start block 1702 that starts with input parameter `pseudo_view_id` and passes control to a function block 1704. The function block 1704 parses a syntax element `num_sub_views_minus1`, and passes control to a function block
15 1706. The function block 1706 sets a variable `i` equal to zero, and passes control to a decision block 1708. The decision block 1708 determines whether or not the variable `i` is less than the number of `sub_views`. If so, then control is passed to a function block 1710. Otherwise, control is passed to a function block 1720.

The function block 1710 parses a syntax element `sub_view_id[i]`, and passes control
20 to a function block 1712. The function block 1712 parses a syntax element `num_parts_minus1[sub_view_id[i]]`, and passes control to a function block 1714. The function block 1714 sets a variable `j` equal to zero, and passes control to a decision block 1716. The decision block 1716 determines whether or not the variable `j` is less than the syntax element `num_parts_minus1[sub_view_id[i]]`. If so, then control is passed to a
25 function block 1718. Otherwise, control is passed to a decision block 1722.

The function block 1718 sets the following syntax elements, increments the variable `j`, and returns control to the decision block 1716: `loc_left_offset[sub_view_id[i]][j]`;
`loc_top_offset[sub_view_id[i]][j]`; `frame_crop_left_offset[sub_view_id[i]][j]`;
`frame_crop_right_offset[sub_view_id[i]][j]`;
30 `frame_crop_top_offset[sub_view_id[i]][j]`; and `frame_crop_bottom_offset[sub_view_id[i]][j]`.

The function block 1720 decodes the current picture for the current view using multi-view video coding (MVC), and passes control to a function block 1721. The function block 1721 separates each view from the picture using the high level syntax, and passes control to an end block 1799.

The separation of each view from the decoded picture is done using the high level syntax indicated in the bitstream. This high level syntax may indicate the exact location and possible orientation of the views (and possible corresponding depth) present in the picture.

5 The decision block 1722 determines whether or not a syntax element `tiling_mode` is equal to zero. If so, then control is passed to a function block 1724. Otherwise, control is passed to a function block 1738.

The function block 1724 parses a syntax element `flip_dir[sub_view_id[i]]` and a syntax element `upsample_view_flag[sub_view_id[i]]`, and passes control to a decision block 1726. The decision block 1726 determines whether or not the current value of the syntax
10 element `upsample_view_flag[sub_view_id[i]]` is equal to one. If so, then control is passed to a function block 1728. Otherwise, control is passed to a decision block 1730.

The function block 1728 parses a syntax element `upsample_filter[sub_view_id[i]]`, and passes control to the decision block 1730. The decision block 1730 determines whether or not a value of the syntax element `upsample_filter[sub_view_id[i]]` is equal to three. If so,
15 the control is passed to a function block 1732. Otherwise, control is passed to a function block 1736.

The function block 1732 parses the following syntax elements, and passes control to a function block 1734: `vert_dim[sub_view_id[i]]`; `hor_dim[sub_view_id[i]]`; and `quantizer[sub_view_id[i]]`. The function block 1734 parses the filter coefficients for each
20 YUV component, and passes control to the function block 1736.

The function block 1736 increments the variable `i`, and returns control to the decision block 1708.

The function block 1738 parses a syntax element `pixel_dist_x[sub_view_id[i]]` and the syntax element `flip_dist_y[sub_view_id[i]]`, and passes control to a function block 1740.
25 The function block 1740 sets the variable `j` equal to zero, and passes control to a decision block 1742. The decision block 1742 determines whether or not the current value of the variable `j` is less than the current value of the syntax element `num_parts[sub_view_id[i]]`. If so, then control is passed to a function block 1744. Otherwise, control is passed to the function block 1736.

30 The function block 1744 parses a syntax element `num_pixel_tiling_filter_coeffs_minus1[sub_view_id[i]]`, and passes control to a function block 1746. The function block 1746 parses the coefficients for all the pixel tiling filters, and passes control to the function block 1736.

Turning to FIG. 18, an exemplary method for processing pictures for a plurality of views and depths in preparation for encoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 1800.

5 The method 1800 includes a start block 1805 that passes control to a function block 1810. The function block 1810 arranges every N views and depth maps, among a total of M views and depth maps, at a particular time instance as a super-picture in tile format, and passes control to a function block 1815. The function block 1815 sets a syntax element `num_coded_views_minus1`, and passes control to a function block 1820. The function block 1820 sets a syntax element `view_id[i]` for all (`num_coded_views_minus1 + 1`) depths corresponding to `view_id[i]`, and passes control to a function block 1825. The function block 1825 sets the inter-view reference dependency information for anchor depth pictures, and passes control to a function block 1830. The function block 1830 sets the inter-view reference dependency information for non-anchor depth pictures, and passes control to a function block 1835. The function block 1835 sets a syntax element `pseudo_view_present_flag`, and passes control to a decision block 1840. The decision block 1840 determines whether or not the current value of the syntax element `pseudo_view_present_flag` is equal to true. If so, then control is passed to a function block 1845. Otherwise, control is passed to an end block 1899.

20 The function block 1845 sets the following syntax elements, and passes control to a function block 1850: `tiling_mode`; `org_pic_width_in_mbs_minus1`; and `org_pic_height_in_mbs_minus1`. The function block 1850 calls a syntax element `pseudo_view_info(view_id)` for each coded view, and passes control to the end block 1899.

Turning to FIG. 19, an exemplary method for encoding pictures for a plurality of views and depths using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 1900.

25 The method 1900 includes a start block 1902 that passes control to a function block 1904. The function block 1904 sets a syntax element `num_sub_views_minus1`, and passes control to a function block 1906. The function block 1906 sets a variable `i` equal to zero, and passes control to a decision block 1908. The decision block 1908 determines whether or not the variable `i` is less than the number of `sub_views`. If so, then control is passed to a function block 1910. Otherwise, control is passed to a function block 1920.

30 The function block 1910 sets a syntax element `sub_view_id[i]`, and passes control to a function block 1912. The function block 1912 sets a syntax element

num_parts_minus1[sub_view_id[i]], and passes control to a function block 1914. The function block 1914 sets a variable j equal to zero, and passes control to a decision block 1916. The decision block 1916 determines whether or not the variable j is less than the syntax element num_parts_minus1[sub_view_id[i]]. If so, then control is passed to a function block 1918. Otherwise, control is passed to a decision block 1922.

The function block 1918 sets the following syntax elements, increments the variable j, and returns control to the decision block 1916: loc_left_offset[sub_view_id[i]][j]; loc_top_offset[sub_view_id[i]][j]; frame_crop_left_offset[sub_view_id[i]][j]; frame_crop_right_offset[sub_view_id[i]][j]; frame_crop_top_offset[sub_view_id[i]][j]; and frame_crop_bottom_offset[sub_view_id[i]][j].

The function block 1920 encodes the current depth for the current view using multi-view video coding (MVC), and passes control to an end block 1999. The depth signal may be encoded similar to the way its corresponding video signal is encoded. For example, the depth signal for a view may be included on a tile that includes only other depth signals, or only video signals, or both depth and video signals. The tile (pseudo-view) is then treated as a single view for MVC, and there are also presumably other tiles that are treated as other views for MVC.

The decision block 1922 determines whether or not a syntax element tiling_mode is equal to zero. If so, then control is passed to a function block 1924. Otherwise, control is passed to a function block 1938.

The function block 1924 sets a syntax element flip_dir[sub_view_id[i]] and a syntax element upsample_view_flag[sub_view_id[i]], and passes control to a decision block 1926. The decision block 1926 determines whether or not the current value of the syntax element upsample_view_flag[sub_view_id[i]] is equal to one. If so, then control is passed to a function block 1928. Otherwise, control is passed to a decision block 1930.

The function block 1928 sets a syntax element upsample_filter[sub_view_id[i]], and passes control to the decision block 1930. The decision block 1930 determines whether or not a value of the syntax element upsample_filter[sub_view_id[i]] is equal to three. If so, the control is passed to a function block 1932. Otherwise, control is passed to a function block 1936.

The function block 1932 sets the following syntax elements, and passes control to a function block 1934: vert_dim[sub_view_id[i]]; hor_dim[sub_view_id[i]]; and quantizer[sub_view_id[i]]. The function block 1934 sets the filter coefficients for each YUV component, and passes control to the function block 1936.

The function block 1936 increments the variable *i*, and returns control to the decision block 1908.

The function block 1938 sets a syntax element `pixel_dist_x[sub_view_id[i]]` and the syntax element `flip_dist_y[sub_view_id[i]]`, and passes control to a function block 1940. The
5 function block 1940 sets the variable *j* equal to zero, and passes control to a decision block 1942. The decision block 1942 determines whether or not the current value of the variable *j* is less than the current value of the syntax element `num_parts[sub_view_id[i]]`. If so, then control is passed to a function block 1944. Otherwise, control is passed to the function block 1936.

10 The function block 1944 sets a syntax element `num_pixel_tiling_filter_coeffs_minus1[sub_view_id[i]]`, and passes control to a function block 1946. The function block 1946 sets the coefficients for all the pixel tiling filters, and passes control to the function block 1936.

Turning to FIG. 20, an exemplary method for processing pictures for a plurality of
15 views and depths in preparation for decoding the pictures using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 2000.

The method 2000 includes a start block 2005 that passes control to a function block 2015. The function block 2015 parses a syntax element `num_coded_views_minus1`, and
20 passes control to a function block 2020. The function block 2020 parses a syntax element `view_id[i]` for all $(\text{num_coded_views_minus1} + 1)$ depths corresponding to `view_id[i]`, and passes control to a function block 2025. The function block 2025 parses the inter-view reference dependency information for anchor depth pictures, and passes control to a function block 2030. The function block 2030 parses the inter-view reference dependency
25 information for non-anchor depth pictures, and passes control to a function block 2035. The function block 2035 parses a syntax element `pseudo_view_present_flag`, and passes control to a decision block 2040. The decision block 2040 determines whether or not the current value of the syntax element `pseudo_view_present_flag` is equal to true. If so, then control is passed to a function block 2045. Otherwise, control is passed to an end block 2099.

30 The function block 2045 parses the following syntax elements, and passes control to a function block 2050: `tiling_mode`; `org_pic_width_in_mbs_minus1`; and `org_pic_height_in_mbs_minus1`. The function block 2050 calls a syntax element `pseudo_view_info(view_id)` for each coded view, and passes control to the end block 2099.

Turning to FIG. 21, an exemplary method for decoding pictures for a plurality of views and depths using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard is indicated generally by the reference numeral 2100.

The method 2100 includes a start block 2102 that starts with input parameter pseudo_view_id, and passes control to a function block 2104. The function block 2104
5 parses a syntax element num_sub_views_minus1, and passes control to a function block 2106. The function block 2106 sets a variable i equal to zero, and passes control to a decision block 2108. The decision block 2108 determines whether or not the variable i is less than the number of sub_views. If so, then control is passed to a function block 2110.
10 Otherwise, control is passed to a function block 2120.

The function block 2110 parses a syntax element sub_view_id[i], and passes control to a function block 2112. The function block 2112 parses a syntax element num_parts_minus1[sub_view_id[i]], and passes control to a function block 2114. The function block 2114 sets a variable j equal to zero, and passes control to a decision block
15 2116. The decision block 2116 determines whether or not the variable j is less than the syntax element num_parts_minus1[sub_view_id[i]]. If so, then control is passed to a function block 2118. Otherwise, control is passed to a decision block 2122.

The function block 2118 sets the following syntax elements, increments the variable j, and returns control to the decision block 2116: loc_left_offset[sub_view_id[i]][j];
20 loc_top_offset[sub_view_id[i]][j]; frame_crop_left_offset[sub_view_id[i]][j];
frame_crop_right_offset[sub_view_id[i]][j];
frame_crop_top_offset[sub_view_id[i]][j]; and frame_crop_bottom_offset[sub_view_id[i]][j].

The function block 2120 decodes the current picture using multi-view video coding (MVC), and passes control to a function block 2121. The function block 2121 separates each
25 view from the picture using the high level syntax, and passes control to an end block 2199. The separation of each view using high level syntax is as previously described.

The decision block 2122 determines whether or not a syntax element tiling_mode is equal to zero. If so, then control is passed to a function block 2124. Otherwise, control is passed to a function block 2138.

30 The function block 2124 parses a syntax element flip_dir[sub_view_id[i]] and a syntax element upsample_view_flag[sub_view_id[i]], and passes control to a decision block 2126. The decision block 2126 determines whether or not the current value of the syntax element upsample_view_flag[sub_view_id[i]] is equal to one. If so, then control is passed to a function block 2128. Otherwise, control is passed to a decision block 2130.

The function block 2128 parses a syntax element `upsample_filter[sub_view_id[i]]`, and passes control to the decision block 2130. The decision block 2130 determines whether or not a value of the syntax element `upsample_filter[sub_view_id[i]]` is equal to three. If so, the control is passed to a function block 2132. Otherwise, control is passed to a function
5 block 2136.

The function block 2132 parses the following syntax elements, and passes control to a function block 2134: `vert_dim[sub_view_id[i]]`; `hor_dim[sub_view_id[i]]`; and `quantizer[sub_view_id[i]]`. The function block 2134 parses the filter coefficients for each YUV component, and passes control to the function block 2136.

10 The function block 2136 increments the variable `i`, and returns control to the decision block 2108.

The function block 2138 parses a syntax element `pixel_dist_x[sub_view_id[i]]` and the syntax element `flip_dist_y[sub_view_id[i]]`, and passes control to a function block 2140. The function block 2140 sets the variable `j` equal to zero, and passes control to a decision
15 block 2142. The decision block 2142 determines whether or not the current value of the variable `j` is less than the current value of the syntax element `num_parts[sub_view_id[i]]`. If so, then control is passed to a function block 2144. Otherwise, control is passed to the function block 2136.

The function block 2144 parses a syntax element
20 `num_pixel_tiling_filter_co coeffs_minus1[sub_view_id[i]]`, and passes control to a function block 2146. The function block 2146 parses the coefficients for all the pixel tiling filters, and passes control to the function block 2136.

Turning to FIG. 22, tiling examples at the pixel level are indicated generally by the reference numeral 2200. FIG. 22 is described further below.

25 An application of multi-view video coding is Free-viewpoint TV (or FTV). This application requires that the user can freely move between two or more views. In order to accomplish this, the “virtual” views in between two views need to be interpolated or synthesized. There are several methods to perform view interpolation. One of the methods uses depth for view interpolation/synthesis.

30 Each view can have an associated depth signal. Thus, the depth can be considered to be another form of video signal. FIG. 9 shows an example of a depth signal 900. In order to enable applications such as FTV, the depth signal is transmitted along with the video signal. In the proposed framework of tiling, the depth signal can also be added as one of the tiles.

FIG. 10 shows an example of depth signals added as tiles. The depth signals/tiles are shown on the right side of FIG. 10.

Once the depth is encoded as a tile of the whole frame, the high level syntax should indicate which tile is the depth signal so that the renderer can use the depth signal appropriately.

In the case when the input sequence (such as that shown in FIG. 1) is encoded using a MPEG-4 AVC Standard encoder (or an encoder corresponding to a different video coding standard and/or recommendation), the proposed high level syntax may be present in, for example, the Sequence Parameter Set (SPS), the Picture Parameter Set (PPS), a slice header, and/or a Supplemental Enhancement Information (SEI) message. An embodiment of the proposed method is shown in TABLE 1 where the syntax is present in a Supplemental Enhancement Information (SEI) message.

In the case when the input sequences of the pseudo views (such as that shown in FIG. 1) is encoded using the multi-view video coding (MVC) extension of the MPEG-4 AVC Standard encoder (or an encoder corresponding to multi-view video coding standard with respect to a different video coding standard and/or recommendation), the proposed high level syntax may be present in the SPS, the PPS, slice header, an SEI message, or a specified profile. An embodiment of the proposed method is shown in TABLE 1. TABLE 1 shows syntax elements present in the Sequence Parameter Set (SPS) structure, including syntax elements proposed realized in accordance with the present principles.

TABLE 1

	C	Descriptor
seq_parameter_set_mvc_extension() {		
num_views_minus_1		ue(v)
for(i = 0; i <= num_views_minus_1; i++)		
view_id[i]		ue(v)
for(i = 0; i <= num_views_minus_1; i++) {		
num_anchor_refs_10[i]		ue(v)
for(j = 0; j < num_anchor_refs_10[i]; j++)		
anchor_ref_10[i][j]		ue(v)
num_anchor_refs_11[i]		ue(v)
for(j = 0; j < num_anchor_refs_11[i]; j++)		
anchor_ref_11[i][j]		ue(v)
}		
for(i = 0; i <= num_views_minus_1; i++) {		
num_non_anchor_refs_10[i]		ue(v)
for(j = 0; j < num_non_anchor_refs_10[i]; j++)		
non_anchor_ref_10[i][j]		ue(v)

num_non_anchor_refs_l1[i]		ue(v)
for(j = 0; j < num_non_anchor_refs_l1[i]; j++)		
non_anchor_ref_l1[i][j]		ue(v)
}		
pseudo_view_present_flag		u(1)
if(pseudo_view_present_flag) {		
tiling_mode		
org_pic_width_in_mbs_minus1		
org_pic_height_in_mbs_minus1		
for(i = 0; i < num_views_minus_1; i++)		
pseudo_view_info(i);		
}		
}		

TABLE 2 shows syntax elements for the pseudo_view_info syntax element of TABLE 1.

5

TABLE 2

pseudo_view_info(pseudo_view_id) {	C	Descriptor
num_sub_views_minus_1[pseudo_view_id]	5	ue(v)
if(num_sub_views_minus_1 != 0) {		
for(i = 0; i < num_sub_views_minus_1[pseudo_view_id]; i++) {		
sub_view_id[i]	5	ue(v)
num_parts_minus1[sub_view_id[i]]	5	ue(v)
for(j = 0; j <= num_parts_minus1[sub_view_id[i]]; j++) {		
loc_left_offset[sub_view_id[i]][j]	5	ue(v)
loc_top_offset[sub_view_id[i]][j]	5	ue(v)
frame_crop_left_offset[sub_view_id[i]][j]	5	ue(v)
frame_crop_right_offset[sub_view_id[i]][j]	5	ue(v)
frame_crop_top_offset[sub_view_id[i]][j]	5	ue(v)
frame_crop_bottom_offset[sub_view_id[i]][j]	5	ue(v)
}		
if(tiling_mode == 0) {		
flip_dir[sub_view_id[i]][j]	5	u(2)
upsample_view_flag[sub_view_id[i]]	5	u(1)
if(upsample_view_flag[sub_view_id[i]])		
upsample_filter[sub_view_id[i]]	5	u(2)
if(upsample_fiter[sub_view_id[i]] == 3) {		
vert_dim[sub_view_id[i]]	5	ue(v)
hor_dim[sub_view_id[i]]	5	ue(v)
quantizer[sub_view_id[i]]	5	ue(v)
for(yuv= 0; yuv< 3; yuv++) {		
for(y = 0; y < vert_dim[sub_view_id[i]] - 1; y++) {		
for(x = 0; x < hor_dim[sub_view_id[i]] - 1; x++)		
filter_co coeffs[sub_view_id[i]][yuv][y][x]	5	se(v)

}		
}		
}		
} // if(tiling_mode == 0)		
else if (tiling_mode == 1) {		
pixel_dist_x[sub_view_id[i]]		
pixel_dist_y[sub_view_id[i]]		
for(j = 0; j <= num_parts[sub_view_id[i]]; j++) {		
num_pixel_tiling_filter_coefs_minus1[sub_view_id[i]][j]		
for (coeff_idx = 0; coeff_idx <=		
num_pixel_tiling_filter_coefs_minus1[sub_view_id[i]][j]; j++)		
pixel_tiling_filter_coefs[sub_view_id[i]][j]		
} // for (j = 0; j <= num_parts[sub_view_id[i]]; j++)		
} // else if (tiling_mode == 1)		
} // for (i = 0; i < num_sub_views_minus_1; i++)		
} // if (num_sub_views_minus_1 != 0)		
}		

Semantics of the syntax elements presented in TABLE 1 and TABLE 2 are described below:

pseudo_view_present_flag equal to true indicates that some view is a super view of multiple sub-views.

5

tiling_mode equal to 0 indicates that the sub-views are tiled at the picture level. A value of 1 indicates that the tiling is done at the pixel level.

The new SEI message could use a value for the SEI payload type that has not been used in the MPEG-4 AVC Standard or an extension of the MPEG-4 AVC Standard. The new SEI message includes several syntax elements with the following semantics.

10

num_coded_views_minus1 plus 1 indicates the number of coded views supported by the bitstream. The value of **num_coded_views_minus1** is in the scope of 0 to 1023, inclusive.

15

org_pic_width_in_mbs_minus1 plus 1 specifies the width of a picture in each view in units of macroblocks.

20

The variable for the picture width in units of macroblocks is derived as follows:

$$\text{PicWidthInMbs} = \text{org_pic_width_in_mbs_minus1} + 1$$

The variable for picture width for the luma component is derived as follows:

$$\text{PicWidthInSamplesL} = \text{PicWidthInMbs} * 16$$

5

The variable for picture width for the chroma components is derived as follows:

$$\text{PicWidthInSamplesC} = \text{PicWidthInMbs} * \text{MbWidthC}$$

10 **org_pic_height_in_mbs_minus1** plus 1 specifies the height of a picture in each view in units of macroblocks.

The variable for the picture height in units of macroblocks is derived as follows:

$$15 \quad \text{PicHeightInMbs} = \text{org_pic_height_in_mbs_minus1} + 1$$

The variable for picture height for the luma component is derived as follows:

$$\text{PicHeightInSamplesL} = \text{PicHeightInMbs} * 16$$

20

The variable for picture height for the chroma components is derived as follows:

$$\text{PicHeightInSamplesC} = \text{PicHeightInMbs} * \text{MbHeightC}$$

25 **num_sub_views_minus1** plus 1 indicates the number of coded sub-views included in the current view. The value of **num_coded_views_minus1** is in the scope of 0 to 1023, inclusive.

30 **sub_view_id[i]** specifies the **sub_view_id** of the sub-view with decoding order indicated by *i*.

num_parts[sub_view_id[i]] specifies the number of parts that the picture of **sub_view_id[i]** is split up into.

$\text{loc_left_offset}[\text{sub_view_id}[i]][j]$ and $\text{loc_top_offset}[\text{sub_view_id}[i]][j]$ specify the locations in left and top pixels offsets, respectively, where the current part j is located in the final reconstructed picture of the view with sub_view_id equal to $\text{sub_view_id}[i]$.

5 $\text{view_id}[i]$ specifies the view_id of the view with coding order indicate by i .

$\text{frame_crop_left_offset}[\text{view_id}[i]][j]$, $\text{frame_crop_right_offset}[\text{view_id}[i]][j]$,
 $\text{frame_crop_top_offset}[\text{view_id}[i]][j]$, and $\text{frame_crop_bottom_offset}[\text{view_id}[i]][j]$
 10 specify the samples of the pictures in the coded video sequence that are part of $\text{num_part } j$
 and $\text{view_id } i$, in terms of a rectangular region specified in frame coordinates for output.

The variables CropUnitX and CropUnitY are derived as follows:

— If chroma_format_idc is equal to 0, CropUnitX and CropUnitY are derived as follows:

15 $\text{CropUnitX} = 1$

$\text{CropUnitY} = 2 - \text{frame_mbs_only_flag}$

— Otherwise (chroma_format_idc is equal to 1, 2, or 3), CropUnitX and CropUnitY are derived as follows:

$\text{CropUnitX} = \text{SubWidthC}$

20 $\text{CropUnitY} = \text{SubHeightC} * (2 - \text{frame_mbs_only_flag})$

The frame cropping rectangle includes luma samples with horizontal frame coordinates from the following:

$\text{CropUnitX} * \text{frame_crop_left_offset}$ to $\text{PicWidthInSamplesL} - (\text{CropUnitX} * \text{frame_crop_right_offset} + 1)$ and vertical frame coordinates from $\text{CropUnitY} * \text{frame_crop_top_offset}$ to $(16 * \text{FrameHeightInMbs}) - (\text{CropUnitY} * \text{frame_crop_bottom_offset} + 1)$, inclusive. The value of $\text{frame_crop_left_offset}$ shall be in the range of 0 to $(\text{PicWidthInSamplesL} / \text{CropUnitX}) - (\text{frame_crop_right_offset} + 1)$, inclusive; and the value of $\text{frame_crop_top_offset}$ shall be in the range of 0 to $(16 * \text{FrameHeightInMbs} / \text{CropUnitY}) - (\text{frame_crop_bottom_offset} + 1)$, inclusive.
 30

When chroma_format_idc is not equal to 0, the corresponding specified samples of the two chroma arrays are the samples having frame coordinates $(x / \text{SubWidthC}, y / \text{SubHeightC})$, where (x, y) are the frame coordinates of the specified luma samples.

For decoded fields, the specified samples of the decoded field are the samples that fall within the rectangle specified in frame coordinates.

5 **num_parts[view_id[i]]** specifies the number of parts that the picture of view_id[i] is split up into.

depth_flag[view_id[i]] specifies whether or not the current part is a depth signal. If depth_flag is equal to 0, then the current part is not a depth signal. If depth_flag is equal to 1, then the current part is a depth signal associated with the view identified by view_id[i].

10

flip_dir[sub_view_id[i]][j] specifies the flipping direction for the current part. flip_dir equal to 0 indicates no flipping, flip_dir equal to 1 indicates flipping in a horizontal direction, flip_dir equal to 2 indicates flipping in a vertical direction, and flip_dir equal to 3 indicates flipping in horizontal and vertical directions.

15

flip_dir[view_id[i]][j] specifies the flipping direction for the current part. flip_dir equal to 0 indicates no flipping, flip_dir equal to 1 indicates flipping in a horizontal direction, flip_dir equal to 2 indicates flipping in vertical direction, and flip_dir equal to 3 indicates flipping in horizontal and vertical directions.

20

loc_left_offset[view_id[i]][j], **loc_top_offset[view_id[i]][j]** specifies the location in pixels offsets, where the current part j is located in the final reconstructed picture of the view with view_id equals to view_id[i]

25 **upsample_view_flag[view_id[i]]** indicates whether the picture belonging to the view specified by view_id[i] needs to be upsampled. upsample_view_flag[view_id[i]] equal to 0 specifies that the picture with view_id equal to view_id[i] will not be upsampled. upsample_view_flag[view_id[i]] equal to 1 specifies that the picture with view_id equal to view_id[i] will be upsampled.

30

upsample_filter[view_id[i]] indicates the type of filter that is to be used for upsampling. upsample_filter[view_id[i]] equals to 0 indicates that the 6-tap AVC filter should be used, upsample_filter[view_id[i]] equals to 1 indicates that the 4-tap SVC filter should be used, upsample_filter[view_id[i]] 2 indicates that the bilinear filter should be used,

upsample_filter[view_id[i]] equals to 3 indicates that custom filter coefficients are transmitted. When upsample_filter[view_id[i]] is not present it is set to 0. In this embodiment, we use 2D customized filter. It can be easily extended to 1D filter, and some other nonlinear filter.

5 **vert_dim[view_id[i]]** specifies the vertical dimension of the custom 2D filter.

hor_dim[view_id[i]] specifies the horizontal dimension of the custom 2D filter.

quantizer[view_id[i]] specifies the quantization factor for each filter coefficient.

10

filter_coeffs[view_id[i]] [yuv][y][x] specifies the quantized filter coefficients. yuv signals the component for which the filter coefficients apply. yuv equal to 0 specifies the Y component, yuv equal to 1 specifies the U component, and yuv equal to 2 specifies the V component.

15

pixel_dist_x[sub_view_id[i]] and **pixel_dist_y[sub_view_id[i]]** respectively specify the distance in the horizontal direction and the vertical direction in the final reconstructed pseudo view between neighboring pixels in the view with sub_view_id equal to sub_view_id[i].

20

num_pixel_tiling_filter_coeffs_minus1[sub_view_id[i][j]] plus one indicates the number of the filter coefficients when the tiling mode is set equal to 1.

25 **pixel_tiling_filter_coeffs[sub_view_id[i][j]]** signals the filter coefficients that are required to represent a filter that may be used to filter the tiled picture.

30 Turning to FIG. 22, two examples showing the composing of a pseudo view by tiling pixels from four views are respectively indicated by the reference numerals 2210 and 2220, respectively. The four views are collectively indicated by the reference numeral 2250. The syntax values for the first example in FIG. 22 are provided in TABLE 3 below.

TABLE 3

pseudo_view_info (pseudo_view_id) {	Value
num_sub_views_minus_1[pseudo_view_id]	3
sub_view_id[0]	0
num_parts_minus1[0]	0
loc_left_offset[0][0]	0
loc_top_offset[0][0]	0
pixel_dist_x[0][0]	0
pixel_dist_y[0][0]	0
sub_view_id[1]	0
num_parts_minus1[1]	0
loc_left_offset[1][0]	1
loc_top_offset[1][0]	0
pixel_dist_x[1][0]	0
pixel_dist_y[1][0]	0
sub_view_id[2]	0
num_parts_minus1[2]	0
loc_left_offset[2][0]	0
loc_top_offset[2][0]	1
pixel_dist_x[2][0]	0
pixel_dist_y[2][0]	0
sub_view_id[3]	0
num_parts_minus1[3]	0
loc_left_offset[3][0]	1
loc_top_offset[3][0]	1
pixel_dist_x[3][0]	0
pixel_dist_y[3][0]	0

The syntax values for the second example in FIG. 22 are all the same except the following two syntax elements: loc_left_offset[3][0] equal to 5 and loc_top_offset[3][0] equal to 3.

The offset indicates that the pixels corresponding to a view should begin at a certain offset location. This is shown in FIG. 22 (2220). This may be done, for example, when two views produce images in which common objects appear shifted from one view to the other. For example, if first and second cameras (representing first and second views) take pictures of an object, the object may appear to be shifted five pixels to the right in the second view as compared to the first view. This means that pixel(i-5, j) in the first view corresponds to pixel(i, j) in the second view. If the pixels of the two views are simply tiled pixel-by-pixel, then there may not be much correlation between neighboring pixels in the tile, and spatial coding gains may be small. Conversely, by shifting the tiling so that pixel(i-5, j) from view one is placed next to pixel(i, j) from view two, spatial correlation may be increased and

spatial coding gain may also be increased. This follows because, for example, the corresponding pixels for the object in the first and second views are being tiled next to each other.

Thus, the presence of `loc_left_offset` and `loc_top_offset` may benefit the coding efficiency. The offset information may be obtained by external means. For example, the position information of the cameras or the global disparity vectors between the views may be used to determine such offset information.

As a result of offsetting, some pixels in the pseudo view are not assigned pixel values from any view. Continuing the example above, when tiling `pixel(i-5, j)` from view one alongside `pixel(i, j)` from view two, for values of $i=0 \dots 4$ there is no `pixel(i-5, j)` from view one to tile, so those pixels are empty in the tile. For those pixels in the pseudo-view (tile) that are not assigned pixel values from any view, at least one implementation uses an interpolation procedure similar to the sub-pixel interpolation procedure in motion compensation in AVC. That is, the empty tile pixels may be interpolated from neighboring pixels. Such interpolation may result in greater spatial correlation in the tile and greater coding gain for the tile.

In video coding, we can choose a different coding type for each picture, such as I, P, and B pictures. For multi-view video coding, in addition, we define anchor and non-anchor pictures. In an embodiment, we propose that the decision of grouping can be made based on picture type. This information of grouping is signaled in high level syntax.

Turning to FIG. 11, an example of 5 views tiled on a single frame is indicated generally by the reference numeral 1100. In particular, the ballroom sequence is shown with 5 views tiled on a single frame. Additionally, it can be seen that the fifth view is split into two parts so that it can be arranged on a rectangular frame. Here, each view is of QVGA size so the total frame dimension is 640x600. Since 600 is not a multiple of 16 it should be extended to 608.

For this example, the possible SEI message could be as shown in TABLE 4.

TABLE 4

<code>multiview_display_info(payloadSize) {</code>	Value
<code>num_coded_views_minus1</code>	5
<code>org_pic_width_in_mbs_minus1</code>	40
<code>org_pic_height_in_mbs_minus1</code>	30
<code>view_id[0]</code>	0
<code>num_parts[view_id[0]]</code>	1

depth_flag[view_id[0]][0]	0
flip_dir[view_id[0]][0]	0
loc_left_offset[view_id[0]][0]	0
loc_top_offset[view_id[0]][0]	0
frame_crop_left_offset[view_id[0]][0]	0
frame_crop_right_offset[view_id[0]][0]	320
frame_crop_top_offset[view_id[0]][0]	0
frame_crop_bottom_offset[view_id[0]][0]	240
upsample_view_flag[view_id[0]]	1
if(upsample_view_flag[view_id[0]]) {	
vert_dim[view_id[0]]	6
hor_dim[view_id[0]]	6
quantizer[view_id[0]]	32
for (yuv= 0; yuv< 3; yuv++) {	
for (y = 0; y < vert_dim[view_id[i]] - 1; y++) {	
for (x = 0; x < hor_dim[view_id[i]] - 1; x++)	
filter_coefs[view_id[i]] [yuv][y][x]	XX
view_id[1]	1
num_parts[view_id[1]]	1
depth_flag[view_id[0]][0]	0
flip_dir[view_id[1]][0]	0
loc_left_offset[view_id[1]][0]	0
loc_top_offset[view_id[1]][0]	0
frame_crop_left_offset[view_id[1]][0]	320
frame_crop_right_offset[view_id[1]][0]	640
frame_crop_top_offset[view_id[1]][0]	0
frame_crop_bottom_offset[view_id[1]][0]	320
upsample_view_flag[view_id[1]]	1
if(upsample_view_flag[view_id[1]]) {	
vert_dim[view_id[1]]	6
hor_dim[view_id[1]]	6
quantizer[view_id[1]]	32
for (yuv= 0; yuv< 3; yuv++) {	
for (y = 0; y < vert_dim[view_id[i]] - 1; y++) {	
for (x = 0; x < hor_dim[view_id[i]] - 1; x++)	
filter_coefs[view_id[i]] [yuv][y][x]	XX
.....(similarly for view 2,3)	

view_id[4]	4
num_parts[view_id[4]]	2
depth_flag[view_id[0]][0]	0
flip_dir[view_id[4]][0]	0
loc_left_offset[view_id[4]][0]	0
loc_top_offset[view_id[4]][0]	0
frame_crop_left_offset[view_id[4]][0]	0
frame_crop_right_offset[view_id[4]][0]	320
frame_crop_top_offset[view_id[4]][0]	480
frame_crop_bottom_offset[view_id[4]][0]	600
flip_dir[view_id[4]][1]	0
loc_left_offset[view_id[4]][1]	0
loc_top_offset[view_id[4]][1]	120
frame_crop_left_offset[view_id[4]][1]	320
frame_crop_right_offset[view_id[4]][1]	640
frame_crop_top_offset[view_id[4]][1]	480
frame_crop_bottom_offset[view_id[4]][1]	600
upsample_view_flag[view_id[4]]	1
if(upsample_view_flag[view_id[4]]) {	
vert_dim[view_id[4]]	6
hor_dim[view_id[4]]	6
quantizer[view_id[4]]	32
for (yuv= 0; yuv< 3; yuv++) {	
for (y = 0; y < vert_dim[view_id[i]] - 1; y ++) {	
for (x = 0; x < hor_dim[view_id[i]] - 1; x ++)	
filter_coefs[view_id[i]] [yuv][y][x]	XX

TABLE 5 shows the general syntax structure for transmitting multi-view information for the example shown in TABLE 4.

5

TABLE 5

	C	Descriptor
multiview_display_info(payloadSize) {		
num_coded_views_minus1	5	ue(v)
org_pic_width_in_mbs_minus1	5	ue(v)
org_pic_height_in_mbs_minus1	5	ue(v)
for(i = 0; i <= num_coded_views_minus1; i++) {		
view_id[i]	5	ue(v)
num_parts[view_id[i]]	5	ue(v)
for(j = 0; j <= num_parts[i]; j++) {		

depth_flag[view_id[i]][j]		
flip_dir[view_id[i]][j]	5	u(2)
loc_left_offset[view_id[i]][j]	5	ue(v)
loc_top_offset[view_id[i]][j]	5	ue(v)
frame_crop_left_offset[view_id[i]][j]	5	ue(v)
frame_crop_right_offset[view_id[i]][j]	5	ue(v)
frame_crop_top_offset[view_id[i]][j]	5	ue(v)
frame_crop_bottom_offset[view_id[i]][j]	5	ue(v)
}		
upsample_view_flag[view_id[i]]	5	u(1)
if(upsample_view_flag[view_id[i]])		
upsample_filter[view_id[i]]	5	u(2)
if(upsample_fiter[view_id[i]] == 3) {		
vert_dim[view_id[i]]	5	ue(v)
hor_dim[view_id[i]]	5	ue(v)
quantizer[view_id[i]]	5	ue(v)
for (yuv= 0; yuv< 3; yuv++) {		
for (y = 0; y < vert_dim[view_id[i]] - 1; y ++)		
for (x = 0; x < hor_dim[view_id[i]] - 1; x ++)		
filter_coeffs[view_id[i]] [yuv][y][x]	5	se(v)
}		
}		
}		
}		
}		

Referring to FIG. 23, a video processing device 2300 is shown. The video processing device 2300 may be, for example, a set top box or other device that receives encoded video and provides, for example, decoded video for display to a user or for storage. Thus, the device 2300 may provide its output to a television, computer monitor, or a computer or other processing device.

The device 2300 includes a decoder 2310 that receive a data signal 2320. The data signal 2320 may include, for example, an AVC or an MVC compatible stream. The decoder 2310 decodes all or part of the received signal 2320 and provides as output a decoded video signal 2330 and tiling information 2340. The decoded video 2330 and the tiling information 2340 are provided to a selector 2350. The device 2300 also includes a user interface 2360 that receives a user input 2370. The user interface 2360 provides a picture selection signal 2380, based on the user input 2370, to the selector 2350. The picture selection signal 2380 and the user input 2370 indicate which of multiple pictures a user desires to have displayed. The selector 2350 provides the selected picture(s) as an output 2390. The selector 2350 uses the picture selection information 2380 to select which of the pictures in the decoded video

2330 to provide as the output 2390. The selector 2350 uses the tiling information 2340 to locate the selected picture(s) in the decoded video 2330.

In various implementations, the selector 2350 includes the user interface 2360, and in other implementations no user interface 2360 is needed because the selector 2350 receives the user input 2370 directly without a separate interface function being performed. The selector 2350 may be implemented in software or as an integrated circuit, for example. The selector 2350 may also incorporate the decoder 2310.

More generally, the decoders of various implementations described in this application may provide a decoded output that includes an entire tile. Additionally or alternatively, the decoders may provide a decoded output that includes only one or more selected pictures (images or depth signals, for example) from the tile.

As noted above, high level syntax may be used to perform signaling in accordance with one or more embodiments of the present principles. The high level syntax may be used, for example, but is not limited to, signaling any of the following: the number of coded views present in the larger frame; the original width and height of all the views; for each coded view, the view identifier corresponding to the view; for each coded view, the number of parts the frame of a view is split into; for each part of the view, the flipping direction (which can be, for example, no flipping, horizontal flipping only, vertical flipping only or horizontal and vertical flipping); for each part of the view, the left position in pixels or number of macroblocks where the current part belongs in the final frame for the view; for each part of the view, the top position of the part in pixels or number of macroblocks where the current part belongs in the final frame for the view; for each part of the view, the left position, in the current large decoded/encoded frame, of the cropping window in pixels or number of macroblocks; for each part of the view, the right position, in the current large decoded/encoded frame, of the cropping window in pixels or number of macroblocks; for each part of the view, the top position, in the current large decoded/encoded frame, of the cropping window in pixels or number of macroblocks; and, for each part of the view, the bottom position, in the current large decoded/encoded frame, of the cropping window in pixels or number of macroblocks; for each coded view whether the view needs to be upsampled before output (where if the upsampling needs to be performed, a high level syntax may be used to indicate the method for upsampling (including, but not limited to, AVC 6-tap filter, SVC 4-tap filter, bilinear filter or a custom 1D, 2D linear or non-linear filter).

It is to be noted that the terms "encoder" and "decoder" connote general structures and are not limited to any particular functions or features. For example, a decoder may

receive a modulated carrier that carries an encoded bitstream, and demodulate the encoded bitstream, as well as decode the bitstream.

Various methods have been described. Many of these methods are detailed to provide ample disclosure. It is noted, however, that variations are contemplated that may vary one or many of the specific features described for these methods. Further, many of the features that are recited are known in the art and are, accordingly, not described in great detail.

Further, reference has been made to the use of high level syntax for sending certain information in several implementations. It is to be understood, however, that other implementations use lower level syntax, or indeed other mechanisms altogether (such as, for example, sending information as part of encoded data) to provide the same information (or variations of that information).

Various implementations provide tiling and appropriate signaling to allow multiple views (pictures, more generally) to be tiled into a single picture, encoded as a single picture, and sent as a single picture. The signaling information may allow a post-processor to pull the views/pictures apart. Also, the multiple pictures that are tiled could be views, but at least one of the pictures could be depth information. These implementations may provide one or more advantages. For example, users may want to display multiple views in a tiled manner, and these various implementations provide an efficient way to encode and transmit or store such views by tiling them prior to encoding and transmitting/storing them in a tiled manner.

Implementations that tile multiple views in the context of AVC and/or MVC also provide additional advantages. AVC is ostensibly only used for a single view, so no additional view is expected. However, such AVC-based implementations can provide multiple views in an AVC environment because the tiled views can be arranged so that, for example, a decoder knows that that the tiled pictures belong to different views (for example, top left picture in the pseudo-view is view 1, top right picture is view 2, etc).

Additionally, MVC already includes multiple views, so multiple views are not expected to be included in a single pseudo-view. Further, MVC has a limit on the number of views that can be supported, and such MVC-based implementations effectively increase the number of views that can be supported by allowing (as in the AVC-based implementations) additional views to be tiled. For example, each pseudo-view may correspond to one of the supported views of MVC, and the decoder may know that each "supported view" actually includes four views in a pre-arranged tiled order. Thus, in such an implementation, the number of possible views is four times the number of "supported views".

In the description that follows, the higher level syntax such as the SEI message is expanded in various implementations to include information about which of the plurality of spatial interleaving modes is attributable to the pictures that are tiled as the single picture. Spatial interleaving can occur in a plurality of modes such as side-by-side, top-bottom, vertical interlacing, horizontal interlacing, and checkerboard, for example. Additionally, the syntax is expanded to include relationship information about the content in the tiled pictures. Relationship information can include a designation of the left and right pictures in a stereo image, or an identification as the depth picture when 2D plus depth is used for the pictures, or even an indication that the pictures form a layered depth video (LDV) signal.

As already noted herein above, the syntax is contemplated as being implemented in any high level syntax besides the SEI message, such as syntax at the slice header level, Picture Parameter Set (PPS) level, Sequence Parameter Set (SPS) level, View Parameter Set (VPS) level, and Network Abstraction Layer (NAL) unit header level. Additionally, it is contemplated that a low level syntax can be used to signal the information. It is even contemplated that the information can be signalled out of band in various manners.

The SEI message syntax for checkerboard spatial interleaving as described in the aforementioned draft amendment of AVC is defined in Table 6 as follows:

TABLE 6

spatially_interleaved_pictures(payloadSize) {	C	Descriptor
spatially_interleaved_pictures_id	5	ue(v)
spatially_interleaved_pictures_cancel_flag	5	u(1)
if(!spatially_interleaved_pictures_cancel_flag) {		
basic_spatial_interleaving_type_id	5	u(8)
spatially_interleaved_pictures_repetition_period	5	ue(v)
}		
additional_extension_flag	5	u(1)
}		

This SEI message informs the decoder that the output decoded pictures are formed by spatial interleaving of multiple distinct pictures using an indicated spatial interleaving scheme. The information in this SEI message can be used by the decoder to appropriately de-interleave and process the picture data for display or other purposes.

The semantics for defined values of the syntax of the spatially interleaved pictures SEI message in Table 6 are as follows:

spatially_interleaved_pictures_id contains an identifying number that may be used to identify the usage of the spatially interleaved pictures SEI message.

spatially_interleaved_pictures_cancel_flag equal to 1 indicates that the spatially interleaved pictures SEI message cancels the persistence of any previous spatially interleaved pictures SEI message in output order. **spatially_interleaved_pictures_cancel_flag** equal to 0 indicates that spatially interleaved pictures information follows.

basic_spatial_interleaving_type_id indicates the type of spatial interleaving of the multiple pictures included in the single tiled picture

basic_spatial_interleaving_type_id equal to 0 indicates that each component plane of the decoded pictures contains a "checkerboard" based interleaving of corresponding planes of two pictures as illustrated in FIG. 26.

basic_spatial_interleaving_type_id equal to 1 indicates that each component plane of the decoded pictures contains a "checkerboard" based interleaving of a corresponding planes of two pictures as illustrated in FIG. 26, and additionally indicates that the two constituent pictures form the left and right views of a stereo-view scene as illustrated in FIG. 26.

spatially_interleaved_pictures_repetition_period specifies the persistence of the spatially interleaved pictures SEI message and may specify a picture order count interval within which another spatially interleaved pictures SEI message with the same value of **spatially_interleaved_pictures_id** or the end of the coded video sequence shall be present in the bitstream.

spatially_interleaved_pictures_repetition_period equal to 0 specifies that the spatially interleaved pictures SEI message applies to the current decoded picture only.

spatially_interleaved_pictures_repetition_period equal to 1 specifies that the spatially interleaved pictures SEI message persists in output order until any of the following conditions are true:

- A new coded video sequence begins.
- A picture in an access unit containing a spatially interleaved pictures SEI message with the same value of **spatially_interleaved_pictures_id** is output having $\text{PicOrderCnt}(\)$ greater than $\text{PicOrderCnt}(\text{CurrPic})$.

spatially_interleaved_pictures_repetition_period equal to 0 or equal to 1 indicates that another spatially interleaved pictures SEI message with the same value of **spatially_interleaved_pictures_id** may or may not be present.

spatially_interleaved_pictures_repetition_period greater than 1 specifies that the spatially interleaved pictures SEI message persists until any of the following conditions are true:

- A new coded video sequence begins.
- 5 – A picture in an access unit containing a spatially interleaved pictures SEI message with the same value of spatially_interleaved_pictures_id is output having PicOrderCnt() greater than PicOrderCnt(CurrPic) and less than or equal to PicOrderCnt(CurrPic) + spatially_interleaved_pictures_repetition_period.

10 **spatially_interleaved_pictures_repetition_period** greater than 1 indicates that another spatially interleaved pictures SEI message with the same value of spatially_interleaved_pictures_id shall be present for a picture in an access unit that is output having PicOrderCnt() greater than PicOrderCnt(CurrPic) and less than or equal to PicOrderCnt(CurrPic) + spatially_interleaved_pictures_repetition_period; unless the
15 bitstream ends or a new coded video sequence begins without output of such a picture.

additional_extension_flag equal to 0 indicates that no additional data follows within the spatially interleaved pictures SEI message.

Without changing the syntax shown in Table 6, an implementation of this application provides relationship information and spatial interleaving information within the exemplary
20 SEI message. The range of possible values for **basic_spatial_interleaving_type_id** is modified and expanded in this implementation to indicate a plurality of spatial interleaving methods rather than just the one checkerboard method. Moreover, the parameter **basic_spatial_interleaving_type_id** is exploited to indicate that a particular type of spatial interleaving is present in the picture and that the constituent interleaved pictures are related to
25 each other. In this implementation, for **basic_spatial_interleaving_type_id**, the semantics are as follows:

Value 2 or 3 means that the single picture contains a "side-by-side" interleaving of corresponding planes of two pictures as illustrated in FIG. 27. Value 3 means additionally that the two constituent pictures form the left and right views of a stereo-view scene. For
30 side-by-side interleaving, one picture is placed to one side of the other picture so that the composite picture includes two images side-by-side.

Value 4 or 5 means that the single picture contains a "top-bottom" interleaving of corresponding planes of two pictures as illustrated in FIG. 28. Value 5 means additionally that the two constituent pictures form the left and right views of a stereo-view scene. For

top-bottom interleaving, one picture is placed above the other picture so that the composite picture appears to have one image over the other.

Value 6 or 7 means that the single picture contains a "row-by-row" interleaving or simply row interlacing of corresponding planes of two pictures as illustrated in FIG. 29.

5 Value 7 means additionally that the two constituent pictures form the left and right views of a stereo-view scene. For row-by-row interleaving, the consecutive rows of the single picture alternate from one constituent picture to the other. Basically, the single picture is an alternation of horizontal slices of the constituent images.

10 Value 8 or 9 means that the single picture contains a "column-by-column" interleaving of corresponding planes of two pictures as illustrated in FIG. 30. Value 9 means additionally that the two constituent pictures form the left and right views of a stereo-view scene. For column-by-column interleaving, the consecutive columns of the single picture alternate from one constituent picture to the other. Basically, the single picture is an alternation of vertical slices of the constituent images.

15 In another embodiment of a syntax for use in an exemplary SEI message to convey information about an associated picture, several additional syntax elements have been included in Table 7 to indicate additional information. Such additional syntax elements are included, for example, to indicate orientation of one or more constituent pictures (for example, flip), to indicate separately whether a left-right stereo pair relationship exists for the
20 images, to indicate whether upsampling is needed for either constituent picture, and to indicate a possible degree and direction of upsampling.

TABLE 7

	C	Descriptor
spatially_interleaved_pictures(payloadSize) {		
spatially_interleaved_pictures_id	5	ue(v)
spatially_interleaved_pictures_cancel_flag	5	U(1)
if(!spatially_interleaved_pictures_cancel_flag) {		
basic_spatial_interleaving_type_id	5	U(8)
stereo_pair_flag	5	U(1)
upsample_conversion_horizontal_flag	5	U(1)
upsample_conversion_vertical_flag	5	U(1)
if(basic_spatial_interleaving_type_id == 1 or basic_spatial_interleaving_type_id == 2) {		
flip_flag	5	U(1)
}		
spatially_interleaved_pictures_repetition_period	5	ue(v)
}		
additional_extension_flag	5	U(1)
}		

Semantics are defined as follows:

basic_spatial_interleaving_type_id indicates the type of spatial interleaving of the pictures.

5 **basic_spatial_interleaving_type_id** equal to 0 indicates that each component plane of the decoded pictures contains a "checkerboard" based interleaving of corresponding planes of two pictures as in the earlier proposal (see FIG. 26).

10 **basic_spatial_interleaving_type_id** equal to 1 indicates that each component plane of the decoded pictures contains a "side-by-side" based interleaving of corresponding planes of two pictures as illustrated in FIG. 27.

basic_spatial_interleaving_type_id equal to 2 indicates that each component plane of the decoded pictures contains a "top-bottom" based interleaving of corresponding planes of two pictures as illustrated in FIG. 28.

15 **basic_spatial_interleaving_type_id** equal to 3 indicates that each component plane of the decoded pictures contains a "row-by-row" based interleaving of corresponding planes of two pictures as illustrated in FIG. 29.

basic_spatial_interleaving_type_id equal to 4 indicates that each component plane of the decoded pictures contains a "column-by-column" based interleaving of corresponding planes of two pictures as illustrated in FIG. 30.

5 **stereo_pair_flag** indicates whether the two constituent pictures have a relationship as forming the left and right views of a stereo-view scene. Value of 0 indicates that the constituent pictures do not form the left and right views. Value of 1 indicates they are related as forming the left and right views of an image.

10 **upsample_conversion_horizontal_flag** indicates whether the two constituent pictures need upsampling in the horizontal direction after they are extracted from the single picture during decoding. Value of 0 indicates that no upsampling is needed. That corresponds to a sampling factor of zero. Value of 1 indicates that upsampling by a sampling factor of two is required.

15 **upsample_conversion_vertical_flag** indicates whether the two constituent pictures need upsampling in the vertical direction after they are extracted from the single picture during decoding. Value of 0 indicates that no upsampling is needed. That corresponds to a sampling factor of zero. Value of 1 indicates that upsampling by a sampling factor of two is required.

As described in more detail below, it is contemplated that many facets of the upsampling operation can be conveyed in the SEI message so that the upsampling operation is handled properly during picture decoding. For example, an additional range of factors for upsampling can be indicated; the type of upsampling filter can also be indicated; the downsampling filter can also be indicated so that the decoder can determine an appropriate, or even optimal, filter for upsampling. It is also contemplated that filter coefficient information including the number and value of the upsampling filter coefficients can be further indicated in the SEI message so that the receiver performs the preferred upsampling operation.

20 The sampling factor indicates the ratio between the original size and the sampled size of a video picture. For example, when the sampling factor is 2, the original picture size is twice as large as the sampled picture size. The picture size is usually a measure of resolution in pixels. So a horizontally downsampled picture requires a corresponding horizontal upsampling by the same factor to restore the resolution of the original video picture. If the original pictures have a width of 1024 pixels, for example, it may be downsampled

horizontally by a sampling factor of 2 to become a downsampled picture with a width of 512 pixels. The picture size is usually a measure of resolution in pixels. A similar analysis can be shown for vertical downsampling or upsampling. The sampling factor can be applied to the downsampling or upsampling that relies on a combined horizontal and vertical approach.

5 **flip_flag** indicates whether the second constituent picture is flipped or not. Value of 0 indicates that no flip is present in that picture. Value of 1 indicates that flipping is performed. Flipping should be understood by persons skilled in the art to involve a rotation of 180° about a central axis in the plane of the picture. The direction of flipping is determined, in this embodiment, by the interleaving type. For example, when side-by-side
10 interleaving (`basic_spatial_interleaving_type_id` equals 1) is present in the constituent pictures, it is preferable to flip the right hand picture in horizontal direction (that is, about a vertical axis), if indicated by an appropriate value of `flip_flag` (see FIG. 31). When top-bottom spatial interleaving (`basic_spatial_interleaving_type_id` equals 2) is present in the constituent pictures, it is preferable to flip the bottom picture in vertical direction (that is,
15 about a central horizontal axis), if indicated by an appropriate value of `flip_flag` (see FIG. 32). While flipping the second constituent picture has been described herein, it is contemplated that other exemplary embodiments may involve flipping the first picture, such as the top picture in top-bottom interleaving or the left picture in side-by-side interleaving. Of course, in order to handle these additional degrees of freedom for the `flip_flag`, it may be necessary
20 to increase the range of values or introduce another semantic related thereto.

It is also contemplated that additional embodiments may allow flipping a picture in both the horizontal and vertical direction. For example, when four views are tiled together in a single picture with one view (that is, picture) per quadrant as shown in FIG. 2, it is possible to have the top-left quadrant include a first view that is not flipped, while the top-right
25 quadrant includes a second view that is flipped only horizontally, whereas the bottom-left quadrant includes a third view that is flipped vertically only, and whereas the bottom-right quadrant includes a fourth view that is flipped both horizontally and vertically. By tiling or interleaving in this manner, it can be seen that the borders at the interface between the views have a large likelihood of having common scene content on both sides of the borders from
30 the neighboring views in the picture. This type of flipping may provide additional efficiency in compression. Indication of this type of flipping is contemplated within the scope of this disclosure.

It is further contemplated that the syntax in Table 7 can be adopted for use in processing 2D plus depth interleaving. Commercial displays may be developed, for example, that accept such a format as input. An exemplary syntax for such an application is set forth in Table 8. Many of the semantics have already been defined above. Only the newly introduced semantics are described below.

TABLE 8

	Descriptor
spatially_interleaved_pictures(payloadSize) {	
spatially_interleaved_pictures_id	ue(v)
spatially_interleaved_pictures_cancel_flag	u(1)
if(!spatially_interleaved_pictures_cancel_flag) {	
basic_spatial_interleaving_type_id	u(8)
semantics_id	u(1)
upsample_conversion_horizontal_flag	u(1)
upsample_conversion_vertical_flag	u(1)
spatially_interleaved_pictures_repetition_period	ue(v)
if(basic_spatial_interleaving_type_id == 1 or basic_spatial_interleaving_type_id == 2) {	
flip_flag	u(1)
}	
if(semantics_id == 1) {	
camera_parameter_set()	
}	
}	
additional_extension_flag	u(1)
}	

semantics_id replaces **stereo_pair_flag** in the prior exemplary embodiment and is used to indicate what relationship is intended for the two interleaved pictures, in other words, what the two pictures physically mean. Value of 0 indicates that the two constituent pictures form the left and right views of an image. Value of 1 indicates that the second picture represents the corresponding depth map of the first constituent picture. FIG. 34 shows an exemplary single picture in which the constituent video and depth images (for example, 2D + Z) are interleaved side by side. It should be appreciated that other interleaving methods can be indicated by a particular value of the **basic_spatial_interleaving_type_id**. Value of 2

indicates that the relationship between the two interleaved pictures is not specified. Values greater than 2 can be used for indicating additional relationships.

camera_parameter_set indicates the parameters of the scene related to the camera. The parameters may appear in this same SEI message or the parameters can be conveyed in another SEI message, or they may appear within other syntaxes such as SPS, PPS, slice headers, and the like. The camera parameters generally include at least focal length, baseline distance, camera position(s), Znear (the minimum distance between the scene and cameras), and Zfar (the maximum distance between the scene and cameras). The camera parameter set may also include a full parameter set for each camera, including the 3x3 intrinsic matrix, a 3x3 rotation matrix, and a 3D translation vector. The camera parameters are used, for example, in rendering and possibly also in coding.

When using an SEI message or other syntax comparable to that shown in Table 8, it may be possible to avoid a need for synchronization at the system level between the video and depth. This results because the video and the associated depth are already tiled into one single frame.

Layer depth video (LDV) format for pictures is shown in FIG. 33. In this figure, four pictures are interleaved – side-by-side and top-bottom – to form the composite picture. The upper left quadrant picture represents the center view layer, whereas the upper right quadrant picture represents the center depth layer. Similarly, the lower left quadrant picture represents the occlusion view layer, whereas the bottom right quadrant picture represents the occlusion depth layer. The format shown in FIG. 33 may be used as an input format for a commercial auto-stereoscopic display.

The presence of LDV indicated using the syntax similar to that shown in the previous embodiments. Primarily, the semantics of **semantics_id** are extended to introduce an LDV option as follows:

TABLE 9

spatially_interleaved_pictures(payloadSize) {	C	Descriptor
spatially_interleaved_pictures_id	5	ue(v)
spatially_interleaved_pictures_cancel_flag	5	u(1)
if(!spatially_interleaved_pictures_cancel_flag) {		
basic_spatial_interleaving_type_id	5	u(8)
semantics_id	5	u(1)
upsample_conversion_horizontal_flag	5	u(1)

upsample_conversion_vertical_flag	5	u(1)
spatially_interleaved_pictures_repetition_period	5	ue(v)
if (basic_spatial_interleaving_type_id == 1 or basic_spatial_interleaving_type_id == 2) {		
flip_flag	5	u(1)
}		
if (semantics_id == 1 semantics_id == 3) {		
camera_parameter_set()		
}		
}		
additional_extension_flag	5	u(1)
}		

semantics_id indicates a relationship between the pictures, that is, what the interleaved pictures physically mean with respect to each other.

Values from 0 to 2 mean that two pictures are related as defined above. Value of 0 indicates that the two pictures form the left and right views. This value can also indicate that the views
5 are stereo views of a scene. Value of 1 indicates that the second picture stands for the corresponding depth map of the first picture, that is, a set of 2D+Z pictures. Value of 2 indicates that the relationship between the two interleaved pictures is not specified.

Value of 3 indicates that four component pictures are interleaved and they correspond to the four component pictures of the LDV representation as shown, for example, in FIG. 33.

10 Values greater than 3 can be employed for additional relationship indications.

Additional relationship indications contemplated herein include, but are not limited to: an indication that the multiple views are multiple sets of 2D+Z pictures, also known as MultiView plus Depth (MVD); an indication that the multiple views represent images in two sets of LDV pictures, which is known as Depth Enhanced Stereo (DES).

15 When the **semantics_id** is equal to 3 and the **basic_spatial_interleaving_type_id** is equal to 1 (side-by-side) or 2 (top-bottom), the four component pictures are interleaved as shown in FIG. 33.

20 When **semantics_id** is equal to 3 and **basic_spatial_interleaving_type_id** equal to 0, the four pictures are interleaved as illustrated in FIG. 22, either as Example 1 (element 2210) or Example 2 (element 2220) if the offsets of views 1, 2, and 3 relative to view 0 are additionally signaled by a particular indication such as via a new semantic in the syntax. It is

contemplated that when semantics_id is equal to 3, interleaving is performed for LDV pictures as shown in Example 1 (element 2210) of FIG. 22. It is further contemplated that when semantics_id is equal to 4, interleaving is performed for LDV pictures as shown in Example 2 (element 2220) of FIG. 22. The interleaving for LDV with and without offset has
5 already been described above with respect to FIG. 22.

As noted above, 3D/stereo video contents can be encoded using 2D video coding standards. A frame-packing method allows spatially or even temporally downsampled views to be packed into one frame for encoding. A decoded frame is unpacked into the constituent multiple views, which are then generally upsampled to the original resolution.

10 Upsampling and, more particularly, the upsampling filter play an important role in the quality of the reconstruction of the constituent views or pictures. Selection of the upsample filter depends typically on the downsample filter used in the encoder. The quality of each upsampled decoded picture can be improved if information about either the downsampling or upsampling filter is conveyed to the decoder either in the bitstream or by other non-
15 standardized means. Although the characteristics of upsample filter in the decoder are not necessarily bound to the characteristics of the downsample filter in the encoder such as solely by an inverse or reciprocal relationship, a matched upsample filter can be expected to provide optimal signal reconstruction. In at least one implementation from experimental practice, it has been determined that an SEI message can be used to signal the upsample parameters to
20 the decoder. In other implementations, it is contemplated that upsample and downsample filter parameters are indicated within an SEI message to the video decoder. Typical H.264/MPEG-4 AVC video encoders and decoders are depicted in FIGs. 3 and 4, respectively, as described above. These devices are suitable for practice of the aspects of various implementations disclosed herein. Additionally, other devices not tailored to or even
25 operative with, H.264/MPEG-4 AVC are suitable for practice of various implementations described in this application.

As described above, a frame packing upsample SEI message is defined below within the framework of H.264/MPEG-4 AVC frame packing for signaling various upsample filters. This syntax is not intended to be limited to SEI message alone, since it is contemplated that
30 the message information can also be provided by other means, such as, for example, being used in other high level syntaxes, such as, for example, SPS, PPS, VPS, slice header, and NAL header.

In an exemplary embodiment employing upsampling information, the upsample filter is selected as a 2-dimensional symmetrical FIR filter with an odd number of taps defined by k non-null coefficients represented in the following form:

$$c_k, 0, c_{k-1}, 0, \dots, c_1, 1, c_1, 0, \dots, 0, c_{k-1}, 0, c_k.$$

5 The filter coefficients c_i are normalized such that the sum of all the coefficients c_i is 0.5 for i in the range $\{1, \dots, k\}$.

The exemplary horizontal upsample process using the filter parameters shown above is described through a series of steps as presented in more detail below. The original image is depicted as an $m \times n$ matrix X . The matrix A is defined as an $n \times (2k+n)$ matrix having the
 10 following attributes:

$$A = \begin{bmatrix} D & | & I_{n,n} & | & 0_{(n-k),k} \\ 0_{(n-k),k} & | & & | & D \end{bmatrix},$$

where

$$D = \text{adiag} \left(\overbrace{1, \dots, 1}^k \right)$$

is a $k \times k$ anti-diagonal identity matrix. The original image X is converted into a new image
 15 matrix X' using the tensor product operation as follows:

$$X' = (XA) \otimes [1 \ 0].$$

The matrix H is defined as an $n \times (4k+n-1)$ matrix including the k non-null tap coefficients as shown above (that is, $c_k, 0, c_{k-1}, 0, \dots, c_1, 1, c_1, 0, \dots, 0, c_{k-1}, 0, c_k$) shifted along each successive row and padded with leading and/or trailing zeros as follows:

20

$$H = \begin{bmatrix} c_k & 0 & \dots & c_1 & 1 & c_1 & \dots & c_k & 0 & 0 & \dots \\ 0 & c_k & 0 & \dots & c_1 & 1 & c_1 & \dots & c_k & 0 & \dots \\ \dots & & & & & & & & & & \\ \dots & 0 & c_k & 0 & \dots & c_1 & 1 & c_1 & \dots & c_k & 0 \end{bmatrix}.$$

The output image matrix Y after the horizontal upsampling is then represented as:

$$Y^T = HX'^T.$$

In a similar manner, an exemplary vertical upsampling process is described as the follows. A new matrix A' is defined similar to matrix as:

25

$$A' = \begin{bmatrix} D & | & I_{m,m} & | & 0_{(m-k),k} \\ 0_{(m-k),k} & | & & | & D \end{bmatrix}.$$

The image Y is then converted to a new image matrix Y' via the tensor product operation as shown below:

$$Y' = (YA') \otimes [1 \ 0]$$

The output image matrix after vertical upsampling is then represented as:

$$F^T = HY'^T .$$

The matrix F is the final image matrix after horizontal and vertical upsampling conversion.

5 Upsampling in the example above has been depicted as both a horizontal and a vertical operation. It is contemplated that the order of the operations can be reversed so that vertical upsampling is performed prior to horizontal upsampling. Moreover, it is contemplated that only one type of upsampling may be performed in certain cases.

10 In order to employ the SEI message for conveying upsampling information to the decoder, it is contemplated that certain semantics must be included within the syntaxes presented above. The syntax shown below in Table 10 includes information useful for upsampling at the video decoder. The syntax below has been abridged to show only the parameters necessary for conveying the upsampling information. It will be appreciated by persons skilled in the art that the syntax shown below can be combined with any one or more
 15 of the prior syntaxes to convey a substantial amount of frame packing information relating to relationship indication, orientation indication, upsample indication, and spatial interleaving indication, for example.

TABLE 10

Frame_packing_filter(payloadSize) {	Descriptor
for (c = 0; c<3; c++)	
{	
number_of_horizontal_filter_parameters[c]	ue(v)
for (i=0; i< number_of_horizontal_filter_parameters[c] ; i++)	
{	
h[c][i]	s(16)
}	
number_of_vertical_filter_parameters[c]	ue(v)
for (i=0; i< number_of_horizontal_filter_parameters[c] ; i++)	
{	
v[c][i]	s(16)
}	

}	
}	

Semantics of the syntax elements are defined below as follows:

number_of_horizontal_filter_parameters[c] indicate order of the (4k-1)-tap horizontal filter for color component c, where c is one of three different color components.

5 **h[c][i]** specifies the i^{th} coefficient of the horizontal filter for color component c in 16-bit precision which is in the range of -2^{15} to $2^{15}-1$. The norm of the filter is 2^{16} .

number_of_vertical_filter_parameters indicate order of the (4k-1)-tap vertical filter for color component c.

10 **v[c][i]** specifies the i^{th} coefficient of the vertical filter for color component c in 16-bit precision which is in the range of -2^{15} to $2^{15}-1$. The norm of the filter is 2^{16} .

For this syntax, it is assumed that a known filter type is present in the decoder. It is contemplated, however, that various different types of upsample filters may be used in addition to, or in lieu of, the above-described 2-dimensional symmetrical FIR filter. Such other upsample filters include, but are not limited to, any interpolation filter designs such as a
 15 bilinear filter, a cubic filter, a spline filter, a Wiener filter, or a least squares filter. When other filter types are available for use, it is contemplated that the syntax above should be expanded to include one or more semantics to indicate information about the filter type for upsampling.

It will be appreciated by persons skilled in the art that the upsample filter parameters
 20 may be derived at the encoder for inclusion in the SEI message based on various inputs such as downsampling filter information or the like. In addition, it will also be appreciated that such information may simply be provided to the encoder for inclusion in the SEI message.

It is further contemplated that the syntax shown in Table 10 could be modified to transmit the downsample filter parameters instead of the upsample filter parameters. In this
 25 embodiment, indication of the downsample filter parameters to the decoder affords the decoding application the ability to determine the optimal upsample filter parameters based on the downsample filter parameters. Similar syntax and semantics can be used to signal the downsample filter. Semantics may be included to indicate that the parameters are representative of either a downsample filter or an upsample filter.

30 If an optimal upsample filter is to be determined by the decoder, the decoder may perform such a determination in any manner known in the art. Additionally, a decoder may

determine an upsample filter that is not optimal, or that is not necessarily optimal. The determination at the decoder may be based on processing limitations or display parameter considerations or the like. As a further example, a decoder could initially select multiple different types of upsampling filters and then ultimately select the one filter that produces a result determined by some decoder criteria to be optimal for the decoder.

A flowchart showing an exemplary encoding technique related to the indication of upsample filter information is shown in FIG. 35 and described briefly below. Initially, the encoder configuration is determined and the high level syntax is created. Pictures (for example, views) are downsampled by a downsample filter for packing into frames. The downsampled pictures are spatially interleaved into the frames. Upsample filter parameters are determined by the encoder or are supplied directly to the encoder. In one embodiment, the upsample filter parameters such as type, size, and coefficient values are derived from the downsample filter. For each color component (for example, YUV, RGB), the horizontal and vertical upsample filter parameters are determined in number and value. This is shown in FIG. 35 with three loops: a first loop for the components, a second loop (nested in the first loop) for the horizontal filter parameters for a component, and a third loop (nested in the first loop) for the vertical filter parameters for a component. The upsample filter parameters are then written into an SEI message. The SEI message is then sent to the decoder either separately (out-of-band) from the video image bitstream or together (in-band) with the video image bitstream. The encoder combines the message with the sequence of packed frames, when needed.

In contrast to the encoding method depicted in FIG. 35, the alternative embodiment shown in FIG. 37 shows that the downsample filter parameters are included in the SEI message instead of the upsample filter parameters. In this embodiment, upsample filter parameters are not derived by the encoder and are not conveyed to the decoder with the packed frames.

A flowchart showing an exemplary decoding technique related to the indication of upsample filter information is shown in FIG. 36 and described briefly below. Initially, the SEI message and other related messages are received by the decoder and the syntaxes are read by the decoder. The SEI message is parsed to determine the information included therein including sampling information, interleaving information, and the like. In this example, it is determined that upsampling filter information for each color component is included in the SEI message. Vertical and horizontal upsampling filter information is extracted to obtain the number and value of each horizontal and vertical upsampling filter

coefficient. This extraction is shown in FIG. 36 with three loops: a first loop for the components, a second loop (nested in the first loop) for the horizontal filter parameters for a component, and a third loop (nested in the first loop) for the vertical filter parameters for a component. The SEI message is then stored and the sequence of packed frames is decoded to
5 obtain the pictures packed therein. The pictures are then upsampled using the recovered upsampling filter parameters to restore the pictures to their original full resolution.

In contrast to the decoding method depicted in FIG. 36, the alternative embodiment shown in FIG. 38 shows that the downsample filter parameters are included in the SEI message instead of the upsample filter parameters. In this embodiment, upsample filter
10 parameters are not derived by the encoder and are not conveyed to the decoder with the packed frames. Hence, the decoder employs the received downsample filter parameters to derive upsample filter parameters to be used for restoring the full original resolution to the pictures extracted from interleaving in the packed video frames.

FIG. 39 shows an exemplary video transmission system 2500, to which the present
15 principles may be applied, in accordance with an implementation of the present principles. The video transmission system 2500 may be, for example, a head-end or transmission system for transmitting a signal using any of a variety of media, such as, for example, satellite, cable, telephone-line, or terrestrial broadcast. The transmission may be provided over the Internet or some other network.

20 The video transmission system 2500 is capable of generating and delivering compressed video with depth. This is achieved by generating an encoded signal(s) including depth information or information capable of being used to synthesize the depth information at a receiver end that may, for example, have a decoder.

The video transmission system 2500 includes an encoder 2510 and a transmitter 2520
25 capable of transmitting the encoded signal. The encoder 2510 receives video information and generates an encoded signal(s) with depth. The encoder 2510 may include sub-modules, including for example an assembly unit for receiving and assembling various pieces of information into a structured format for storage or transmission. The various pieces of information may include, for example, coded or uncoded video, coded or uncoded depth
30 information, and coded or uncoded elements such as, for example, motion vectors, coding mode indicators, and syntax elements.

The transmitter 2520 may be, for example, adapted to transmit a program signal having one or more bitstreams representing encoded pictures and/or information related thereto. Typical transmitters perform functions such as, for example, one or more of

providing error-correction coding, interleaving the data in the signal, randomizing the energy in the signal, and/or modulating the signal onto one or more carriers. The transmitter may include, or interface with, an antenna (not shown). Accordingly, implementations of the transmitter 2520 may include, or be limited to, a modulator.

5 FIG. 40 shows an exemplary video receiving system 2600 to which the present principles may be applied, in accordance with an embodiment of the present principles. The video receiving system 2600 may be configured to receive signals over a variety of media, such as, for example, satellite, cable, telephone-line, or terrestrial broadcast. The signals may be received over the Internet or some other network.

10 The video receiving system 2600 may be, for example, a cell-phone, a computer, a set-top box, a television, or other device that receives encoded video and provides, for example, decoded video for display to a user or for storage. Thus, the video receiving system 2600 may provide its output to, for example, a screen of a television, a computer monitor, a computer (for storage, processing, or display), or some other storage, processing, or display
15 device.

The video receiving system 2600 is capable of receiving and processing video content including video information. The video receiving system 2600 includes a receiver 2610 capable of receiving an encoded signal, such as for example the signals described in the implementations of this application, and a decoder 2620 capable of decoding the received
20 signal.

The receiver 2610 may be, for example, adapted to receive a program signal having a plurality of bitstreams representing encoded pictures. Typical receivers perform functions such as, for example, one or more of receiving a modulated and encoded data signal, demodulating the data signal from one or more carriers, de-randomizing the energy in the
25 signal, de-interleaving the data in the signal, and/or error-correction decoding the signal. The receiver 2610 may include, or interface with, an antenna (not shown). Implementations of the receiver 2610 may include, or be limited to, a demodulator. The decoder 2620 outputs video signals including video information and depth information.

FIG. 41 shows an exemplary video processing device 2700 to which the present
30 principles may be applied, in accordance with an embodiment of the present principles. The video processing device 2700 may be, for example, a set top box or other device that receives encoded video and provides, for example, decoded video for display to a user or for storage. Thus, the video processing device 2700 may provide its output to a television, computer monitor, or a computer or other processing device.

The video processing device 2700 includes a front-end (FE) device 2705 and a decoder 2710. The front-end device 2705 may be, for example, a receiver adapted to receive a program signal having a plurality of bitstreams representing encoded pictures, and to select one or more bitstreams for decoding from the plurality of bitstreams. Typical receivers
5 perform functions such as, for example, one or more of receiving a modulated and encoded data signal, demodulating the data signal, decoding one or more encodings (for example, channel coding and/or source coding) of the data signal, and/or error-correcting the data signal. The front-end device 2705 may receive the program signal from, for example, an antenna (not shown). The front-end device 2705 provides a received data signal to the
10 decoder 2710.

The decoder 2710 receives a data signal 2720. The data signal 2720 may include, for example, one or more Advanced Video Coding (AVC), Scalable Video Coding (SVC), or Multi-view Video Coding (MVC) compatible streams.

The decoder 2710 decodes all or part of the received signal 2720 and provides as
15 output a decoded video signal 2730. The decoded video 2730 is provided to a selector 2750. The device 2700 also includes a user interface 2760 that receives a user input 2770. The user interface 2760 provides a picture selection signal 2780, based on the user input 2770, to the selector 2750. The picture selection signal 2780 and the user input 2770 indicate which of multiple pictures, sequences, scalable versions, views, or other selections of the available
20 decoded data a user desires to have displayed. The selector 2750 provides the selected picture(s) as an output 2790. The selector 2750 uses the picture selection information 2780 to select which of the pictures in the decoded video 2730 to provide as the output 2790.

In various implementations, the selector 2750 includes the user interface 2760, and in other implementations no user interface 2760 is needed because the selector 2750 receives
25 the user input 2770 directly without a separate interface function being performed. The selector 2750 may be implemented in software or as an integrated circuit, for example. In one implementation, the selector 2750 is incorporated with the decoder 2710, and in another implementation, the decoder 2710, the selector 2750, and the user interface 2760 are all integrated.

30 In one application, front-end 2705 receives a broadcast of various television shows and selects one for processing. The selection of one show is based on user input of a desired channel to watch. Although the user input to front-end device 2705 is not shown in Figure 41, front-end device 2705 receives the user input 2770. The front-end 2705 receives the broadcast and processes the desired show by demodulating the relevant part of the broadcast

spectrum, and decoding any outer encoding of the demodulated show. The front-end 2705 provides the decoded show to the decoder 2710. The decoder 2710 is an integrated unit that includes devices 2760 and 2750. The decoder 2710 thus receives the user input, which is a user-supplied indication of a desired view to watch in the show. The decoder 2710 decodes the selected view, as well as any required reference pictures from other views, and provides the decoded view 2790 for display on a television (not shown).

Continuing the above application, the user may desire to switch the view that is displayed and may then provide a new input to the decoder 2710. After receiving a "view change" from the user, the decoder 2710 decodes both the old view and the new view, as well as any views that are in between the old view and the new view. That is, the decoder 2710 decodes any views that are taken from cameras that are physically located in between the camera taking the old view and the camera taking the new view. The front-end device 2705 also receives the information identifying the old view, the new view, and the views in between. Such information may be provided, for example, by a controller (not shown in Figure 41) having information about the locations of the views, or the decoder 2710. Other implementations may use a front-end device that has a controller integrated with the front-end device.

The decoder 2710 provides all of these decoded views as output 2790. A post-processor (not shown in Figure 41) interpolates between the views to provide a smooth transition from the old view to the new view, and displays this transition to the user. After transitioning to the new view, the post-processor informs (through one or more communication links not shown) the decoder 2710 and the front-end device 2705 that only the new view is needed. Thereafter, the decoder 2710 only provides as output 2790 the new view.

The system 2700 may be used to receive multiple views of a sequence of images, and to present a single view for display, and to switch between the various views in a smooth manner. The smooth manner may involve interpolating between views to move to another view. Additionally, the system 2700 may allow a user to rotate an object or scene, or otherwise to see a three-dimensional representation of an object or a scene. The rotation of the object, for example, may correspond to moving from view to view, and interpolating between the views to obtain a smooth transition between the views or simply to obtain a three-dimensional representation. That is, the user may "select" an interpolated view as the "view" that is to be displayed.

As presented herein, the implementations and features described in this application may be used in the context of coding video, coding depth, and/or coding other types of data. Additionally, these implementations and features may be used in the context of, or adapted for use in the context of, the H.264/MPEG-4 AVC (AVC) Standard, the AVC standard with the MVC extension, the AVC standard with the SVC extension, a 3DV standard, and/or with another standard (existing or future), or in a context that does not involve a standard. Accordingly, it is to be understood that the specific implementations described in this application that operate in accordance with AVC are not intended to be restricted to AVC, and may be adapted for use outside of AVC.

Also as noted above, implementations may signal or indicate information using a variety of techniques including, but not limited to, SEI messages, slice headers, other high level syntax, non-high-level syntax, out-of-band information, datastream data, and implicit signaling. Although implementations described herein may be described in a particular context, such descriptions should in no way be taken as limiting the features and concepts to such implementations or contexts.

Various implementations involve decoding. "Decoding", as used in this application, may encompass all or part of the processes performed, for example, on a received encoded sequence in order to produce a final output suitable for display. Such processes may include processes typically performed by a decoder such as, for example, entropy decoding, inverse transformation, inverse quantization, differential decoding. Such processes may also, or alternatively, include processes performed by a decoder of various implementations described in this application, such as, for example, extracting a picture from a tiled (packed) picture, determining an upsample filter to use and then upsampling a picture, and flipping a picture back to its intended orientation.

Various implementations described in this application combine multiple pictures into a single picture. The process of combining pictures may involve, for example, selecting pictures based on for example a relationship between pictures, spatial interleaving, sampling, and changing the orientation of a picture. Accordingly, information describing the combining process may include, for example, relationship information, spatial interleaving information, sampling information, and orientation information.

Spatial interleaving is referred to in describing various implementations. "Interleaving", as used in this application, is also referred to as "tiling" and "packing", for example. Interleaving includes a variety of types of interleaving, including, for example, side-by-side juxtaposition of pictures, top-to-bottom juxtaposition of pictures, side-by-side

and top-to-bottom combined (for example, to combine 4 pictures), row-by-row alternating, column-by-column alternating, and various pixel-level (also referred to as pixel-wise or pixel-based) schemes.

Also, as used herein, the words “picture” and “image” are used interchangeably and refer, for example, to all or part of a still image or all or part of a picture from a video sequence. As is known, a picture may be a frame or a field. Additionally, as used herein, a picture may also be a subset of a frame such as, for example, a top half of a frame, a single macroblock, alternating columns, alternating rows, or periodic pixels. As another example, a depth picture may be, for example, a complete depth map or a partial depth map that only includes depth information for, for example, a single macroblock of a corresponding video frame.

Reference in the specification to “one embodiment” or “an embodiment” or “one implementation” or “an implementation” of the present principles, as well as other variations thereof, mean that a particular feature, structure, characteristic, and so forth described in connection with the embodiment is included in at least one embodiment of the present principles. Thus, the appearances of the phrase “in one embodiment” or “in an embodiment” or “in one implementation” or “in an implementation”, as well any other variations, appearing in various places throughout the specification are not necessarily all referring to the same embodiment.

It is to be appreciated that the use of any of the following “/”, “and/or”, and “at least one of”, for example, in the cases of “A/B”, “A and/or B” and “at least one of A and B”, is intended to encompass the selection of the first listed option (A) only, or the selection of the second listed option (B) only, or the selection of both options (A and B). As a further example, in the cases of “A, B, and/or C” and “at least one of A, B, and C” and “at least one of A, B, or C”, such phrasing is intended to encompass the selection of the first listed option (A) only, or the selection of the second listed option (B) only, or the selection of the third listed option (C) only, or the selection of the first and the second listed options (A and B) only, or the selection of the first and third listed options (A and C) only, or the selection of the second and third listed options (B and C) only, or the selection of all three options (A and B and C). This may be extended, as readily apparent by one of ordinary skill in this and related arts, for as many items listed.

Additionally, this application or its claims may refer to “determining” various pieces of information. Determining the information may include one or more of, for example,

estimating the information, calculating the information, predicting the information, or retrieving the information from memory.

Similarly, "accessing" is intended to be a broad term. Accessing a piece of information may include any operation that, for example, uses, stores, sends, transmits, receives, retrieves, modifies, or provides the information.

It should be understood that the elements shown in the figures may be implemented in various forms of hardware, software, or combinations thereof. Preferably, these elements are implemented in a combination of hardware and software on one or more appropriately programmed general-purpose devices, which may include a processor, memory and input/output interfaces. Moreover, the implementations described herein may be implemented as, for example, a method or process, an apparatus, or a software program. Even if only discussed in the context of a single form of implementation (for example, discussed only as a method), the implementation of features discussed may also be implemented in other forms (for example, an apparatus or program). An apparatus may be implemented as mentioned above. The methods may be implemented in, for example, an apparatus such as, for example, a processor, which refers to processing devices in general, including, for example, a computer, a microprocessor, an integrated circuit, or a programmable logic device. Processing devices also include communication devices, such as, for example, computers, cell phones, portable/personal digital assistants ("PDAs"), and other devices that facilitate communication of information between end-users.

Implementations of the various processes and features described herein may be embodied in a variety of different equipment or applications, particularly, for example, equipment or applications associated with data encoding and decoding. Examples of equipment include video coders, video decoders, video codecs, web servers, set-top boxes, laptops, personal computers, cell phones, PDAs, other communication devices, personal recording devices (for example, PVRs, computers running recording software, VHS recording devices), camcorders, streaming of data over the Internet or other communication links, and video-on-demand. As should be clear, the equipment may be mobile and even installed in a mobile vehicle.

Additionally, the methods may be implemented by instructions being performed by a processor, and such instructions may be stored on a processor-readable medium such as, for example, an integrated circuit, a software carrier or other storage device such as, for example, a hard disk, a compact diskette, a random access memory ("RAM"), or a read-only memory ("ROM"). The instructions may form an application program tangibly embodied on a

processor-readable medium. As should be clear, a processor may include a processor-readable medium having, for example, instructions for carrying out a process. Such application programs may be uploaded to, and executed by, a machine comprising any suitable architecture. Preferably, the machine is implemented on a computer platform having
5 hardware such as one or more central processing units (“CPU”), a random access memory (“RAM”), and input/output (“I/O”) interfaces. The computer platform may also include an operating system and microinstruction code. The various processes and functions described herein may be either part of the microinstruction code or part of the application program, or any combination thereof, which may be executed by a CPU. In addition, various other
10 peripheral units may be connected to the computer platform such as an additional data storage unit and a printing unit.

As should be evident to one of skill in the art, implementations may also produce a signal formatted to carry information that may be, for example, stored or transmitted. The information may include, for example, instructions for performing a method, or data
15 produced by one of the described implementations. Such a signal may be formatted, for example, as an electromagnetic wave (for example, using a radio frequency portion of spectrum) or as a baseband signal. The formatting may include, for example, encoding a data stream, producing syntax, and modulating a carrier with the encoded data stream and the syntax. The information that the signal carries may be, for example, analog or digital
20 information. The signal may be transmitted over a variety of different wired or wireless links, as is known.

It is to be further understood that, because some of the constituent system components and methods depicted in the accompanying drawings are preferably implemented in software, the actual connections between the system components or the process function blocks may
25 differ depending upon the manner in which the present principles are programmed. Given the teachings herein, one of ordinary skill in the pertinent art will be able to contemplate these and similar implementations or configurations of the present principles.

A number of implementations have been described. Nevertheless, it will be understood that various modifications may be made. For example, elements of different
30 implementations may be combined, supplemented, modified, or removed to produce other implementations. Additionally, one of ordinary skill will understand that other structures and processes may be substituted for those disclosed and the resulting implementations will perform at least substantially the same function(s), in at least substantially the same way(s), to achieve at least substantially the same result(s) as the implementations disclosed. In

particular, although illustrative embodiments have been described herein with reference to the accompanying drawings, it is to be understood that the present principles is not limited to those precise embodiments, and that various changes and modifications may be effected therein by one of ordinary skill in the pertinent art without departing from the scope or spirit
5 of the present principles. Accordingly, these and other implementations are contemplated by this application and are within the scope of the following claims.

CLAIMS:

1. A method comprising:

5 accessing a video picture that includes multiple pictures combined into a single picture, the video picture being part of a received video stream;

 accessing information that is part of the received video stream, the accessed information indicating how the multiple pictures in the accessed video picture are combined, wherein the accessed information includes spatial interleaving information and sampling
10 information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and

15 decoding the video picture to provide a decoded representation of at least one of the multiple pictures.

2. The method of claim 1, wherein the decoding is performed based on the spatial interleaving information and the sampling information.

20

3. The method of claim 1, wherein the spatial interleaving information indicates spatial interleaving of a type from a group including side-by-side interleaving, top-bottom interleaving, row-by-row interleaving, column-by-column interleaving, and checkerboard interleaving.

25

4. The method of claim 1, wherein the spatial interleaving information further includes relationship information, the relationship information indicating whether and a type of relationship that exists between the multiple pictures.

30 5. The method of claim 4, wherein the decoding is further based on the relationship information.

6. The method of claim 4, wherein the relationship information indicates a relationship from the group consisting of: the multiple pictures are left and right views of an image, the multiple pictures are stereo views of an image, the multiple views are not related, the multiple views are a 2D image and its related depth map (2D+Z), the multiple views are multiple sets of a 2D+Z (MVD), the multiple views represent images in a layer depth video format (LDV), and the multiple views represent images in two sets of LDV (DES), and the multiple views are formatted according to a depth related format.
- 10 7. The method of claim 6, wherein the accessed information further includes camera parameter information, the camera parameter information including at least one of a focal length of a camera, a baseline distance, at least one camera position, a Znear parameter to indicate a minimum distance between an image in the multiple pictures and the camera, a Zfar parameter to indicate a maximum distance between the image in the multiple pictures and the camera, a 3x3 intrinsic matrix, a 3x3 rotation matrix, and a 3D translation vector.
- 15 8. The method of claim 7, wherein the decoding is further based on the camera parameter information.
- 20 9. The method of claim 1, wherein the accessed information further includes orientation information, the orientation information indicating an orientation of at least one of the multiple pictures.
- 25 10. The method of claim 9, wherein the decoding is further based on the orientation information.
- 30 11. The method of claim 9, wherein the orientation information is from a group consisting of: no picture has a flipped orientation, one picture of the multiple pictures has its orientation flipped vertically, and one picture of the multiple pictures has its orientation flipped horizontally.
12. The method of claim 11, wherein the orientation information also indicates the one or more pictures of the multiple pictures to which the orientation information pertains.

13. The method of claim 1, wherein the sampling information includes information from a group consisting of sampling indication, sampling factor, sampling filter type, a number of filter coefficients, and one or more sampling filter coefficient values.
- 5 14. The method of claim 13, wherein the sampling indication indicates whether the sampling information relates to upsampling or downsampling.
15. The method of claim 13, wherein the sampling factor indicates that sampling is performed by a factor selected from a group consisting of the values 0, 2, and 4.
- 10 16. The method of claim 13, wherein the sampling filter type indicates a type of sampling filter from the group consisting of 2-dimensional symmetrical FIR filter, an interpolation filter, a bilinear filter, a cubic filter, a spline filter, a Wiener filter, a least squares filter, a 6-tap AVC filter, a 4-tap SVC filter, a custom filter including a 2D customized filter, a 1D filter, and a nonlinear filter.
- 15 17. The method of claim 13, wherein the one or more sampling filter coefficient values are from a group consisting of upsampling filter coefficient values and downsampling filter coefficient values.
- 20 18. The method of claim 13, wherein the decoding further includes deriving a set of one or more upsampling filter coefficient values from downsampling filter coefficient values included in the accessed information.
- 25 19. The method of claim 1, wherein the information is included in at least one of a slice header, a sequence parameter set, a picture parameter set, a view parameter set, a network abstraction layer unit header, or a supplemental enhancement information message (804).
- 30 20. The method of claim 1, wherein the indication of filtering direction indicates that upsampling is to be performed in the horizontal direction.
21. The method of claim 1, wherein the indication of filtering direction indicates that upsampling is to be performed in the vertical direction.

22. The method of claim 1, wherein the indication of filtering direction indicates that upsampling is to be performed in the horizontal direction and the vertical direction.

23. The method of claim 1, further comprising providing the decoded representation and the
5 accessed information as output for further processing of the decoded representation based on the accessed information.

24. The method of claim 1, wherein the accessed information is included in a supplemental
10 enhancement information message of a standard such as H.264/AVC.

25. The method of claim 1, further comprising providing the decoded representation for
display.

26. The method of claim 1, wherein the one or more parameters related to the upsampling
15 filter include an indication of filter parameters used in downsampling one or more of the multiple pictures to combine the multiple pictures into the single picture.

27. The method of claim 26, wherein the indication of filter parameters used in
20 downsampling is usable in determining upsample filter parameters for use in upsampling to provide the decoded representation.

28. An apparatus comprising

means for accessing a video picture that includes multiple pictures combined into a
single picture, the video picture being part of a received video stream;

25 means for accessing information that is part of the received video stream, the accessed information indicating how the multiple pictures in the accessed video picture are combined, wherein the accessed information includes spatial interleaving information and
sampling information, wherein the spatial interleaving information indicates spatial
interleaving applied to the multiple pictures in forming the single picture and wherein the
30 sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and

means for decoding the video picture to provide a decoded representation of at least
one of the multiple pictures.

29. An apparatus, comprising a processor configured to perform at least the following:

accessing a video picture that includes multiple pictures combined into a single picture, the video picture being part of a received video stream;

5 accessing information that is part of the received video stream, the accessed information indicating how the multiple pictures in the accessed video picture are combined, wherein the accessed information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling
10 information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction;

decoding the video picture to provide a decoded representation of at least one of the multiple pictures.

15

30. A processor readable medium having stored therein instructions for causing a processor to perform at least the following:

accessing a video picture that includes multiple pictures combined into a single picture, the video picture being part of a received video stream;

20 accessing information that is part of the received video stream, the accessed information indicating how the multiple pictures in the accessed video picture are combined, wherein the accessed information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling
25 information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction;

decoding the video picture to provide a decoded representation of at least one of the multiple pictures.

30

31. An apparatus comprising:

a decoder configured to access a video picture that includes multiple pictures combined into a single picture, the video picture being part of a received video stream; and
5 configured to access information that is part of the received video stream, the accessed information indicating how the multiple pictures in the accessed video picture are combined, wherein the accessed information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling
10 information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and configured to decode the video picture to provide a decoded representation of at least one of the multiple pictures.

15 32. An apparatus comprising:

a demodulator for demodulating a signal that includes a received video stream; and
a decoder configured to access a video picture that includes multiple pictures combined into a single picture, the video picture being part of a received video stream; and
configured to access information that is part of the received video stream, the accessed
20 information indicating how the multiple pictures in the accessed video picture are combined, wherein the accessed information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for restoring
25 each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and configured to decode the video picture to provide a decoded representation of at least one of the multiple pictures.

33. A video signal formatted to include information, the video signal comprising:

an encoded picture section including an encoding of a video picture, the video picture including multiple pictures combined into a single picture; and

5 a signaling section including an encoding of generated information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more
10 parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction, the generated information providing information for use in decoding the encoded video picture.

15 34. A video signal structure comprising:

an encoded picture section including an encoding of a video picture, the video picture including multiple pictures combined into a single picture; and

a signaling section including an encoding of generated information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated
20 information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an
25 indication of filtering direction, the generated information providing information for use in decoding the encoded video picture.

35. A processor readable medium having stored thereon a video signal structure, the video signal structure comprising:

an encoded picture section including an encoding of a video picture, the video picture
5 including multiple pictures combined into a single picture; and

a signaling section including an encoding of generated information indicating how the
multiple pictures in the accessed video picture are combined, wherein the generated
information includes spatial interleaving information and sampling information, wherein the
spatial interleaving information indicates spatial interleaving applied to the multiple pictures
10 in forming the single picture and wherein the sampling information indicates one or more
parameters related to an upsampling filter for restoring each of the multiple pictures to a
desired resolution, the one or more parameters related to the upsampling filter including an
indication of filtering direction, the generated information providing information for use in
decoding the encoded video picture.

15

36. A method comprising:

encoding a video picture that includes multiple pictures combined into a single
picture (826);

generating information indicating how the multiple pictures in the accessed video
20 picture are combined, wherein the generated information includes spatial interleaving
information and sampling information, wherein the spatial interleaving information indicates
spatial interleaving applied to the multiple pictures in forming the single picture and wherein
the sampling information indicates one or more parameters related to an upsampling filter for
restoring each of the multiple pictures to a desired resolution, the one or more parameters
25 related to the upsampling filter including an indication of filtering direction; and

forming a bitstream that includes the encoded video picture and the generated
information, the generated information providing information for use in processing the
encoded video picture.

30 37. The method of claim 36, wherein the spatial interleaving information indicates spatial
interleaving of a type from a group including side-by-side interleaving, top-bottom
interleaving, row-by-row interleaving, column-by-column interleaving, and checkerboard
interleaving.

38. The method of claim 36, wherein the spatial interleaving information further include relationship information, the relationship information indicating whether and a type of relationship that exists between the multiple pictures; and

5 wherein the relationship information is also for use in processing the encoded video picture.

39. The method of claim 36, wherein the generated information further includes orientation information, the orientation information indicating an orientation of at least one of the multiple pictures; and

10 wherein the orientation information is also for use in processing the encoded video picture.

40. The method of claim 39, wherein the orientation information is from a group consisting of: no picture has a flipped orientation, one picture of the multiple pictures has its orientation flipped vertically, and one picture of the multiple pictures has its orientation flipped horizontally.

41. The method of claim 1, wherein the indication of filtering direction indicates that upsampling is to be performed in the horizontal direction.

20

42. The method of claim 1, wherein the indication of filtering direction indicates that upsampling is to be performed in the vertical direction.

43. The method of claim 1, wherein the indication of filtering direction indicates that upsampling is to be performed in the horizontal direction and the vertical direction.

25

44. An apparatus comprising

means for encoding a video picture that includes multiple pictures combined into a single picture;

30 means for generating information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for

restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and

means for forming a bitstream that includes the encoded video picture and the generated information, the generated information providing information for use in processing
5 the encoded video picture.

45. An apparatus, comprising a processor configured to perform at least the following:

encoding a video picture that includes multiple pictures combined into a single picture (826);

10 generating information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for
15 restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and

forming a bitstream that includes the encoded video picture and the generated information, the generated information providing information for use in processing the encoded video picture.

20

46. A processor readable medium having stored therein instructions for causing a processor to perform at least the following:

encoding a video picture that includes multiple pictures combined into a single picture (826);

25 generating information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for
30 restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and

forming a bitstream that includes the encoded video picture and the generated information, the generated information providing information for use in processing the encoded video picture.

47. An apparatus comprising:

an encoder configured to encode a video picture that includes multiple pictures combined into a single picture; and to generate information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and to form a bitstream that includes the encoded video picture and the generated information, the generated information providing information for use in processing the encoded video picture.

48. An apparatus comprising:

an encoder configured to encode a video picture that includes multiple pictures combined into a single picture; and to generate information indicating how the multiple pictures in the accessed video picture are combined, wherein the generated information includes spatial interleaving information and sampling information, wherein the spatial interleaving information indicates spatial interleaving applied to the multiple pictures in forming the single picture and wherein the sampling information indicates one or more parameters related to an upsampling filter for restoring each of the multiple pictures to a desired resolution, the one or more parameters related to the upsampling filter including an indication of filtering direction; and to form a bitstream that includes the encoded video picture and the generated information, the generated information providing information for use in processing the encoded video picture; and

a modulator for modulating a signal that includes the encoded video picture.

1/40

100



FIG. 1

200

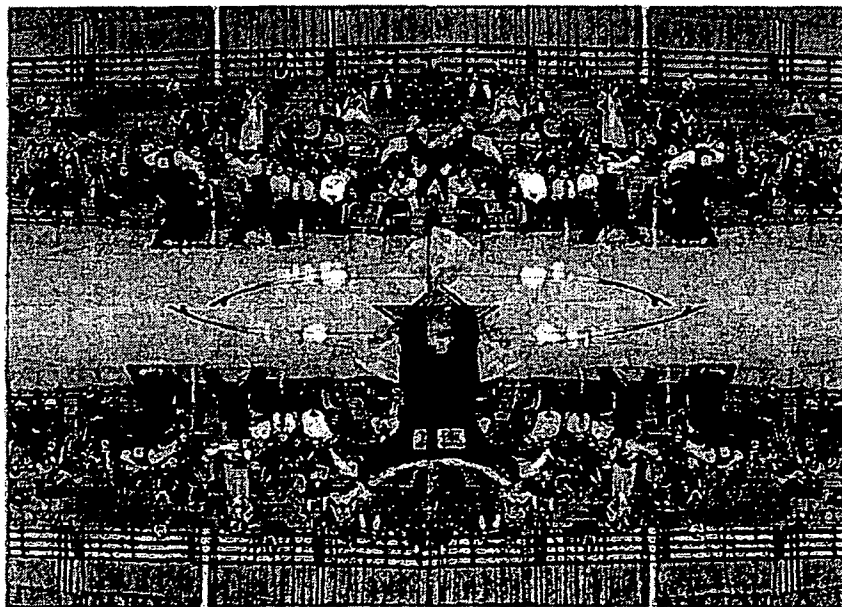


FIG. 2

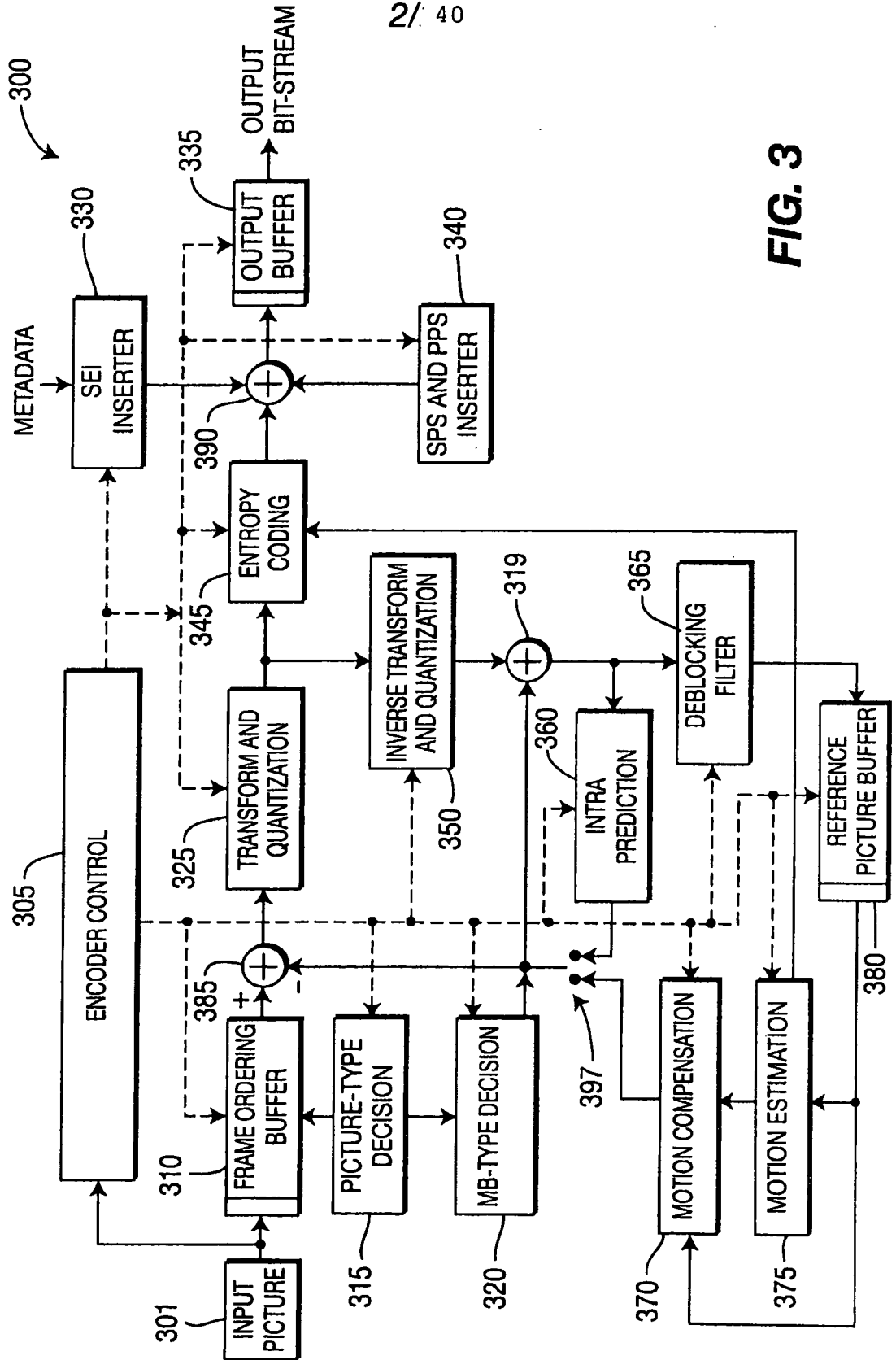


FIG. 3

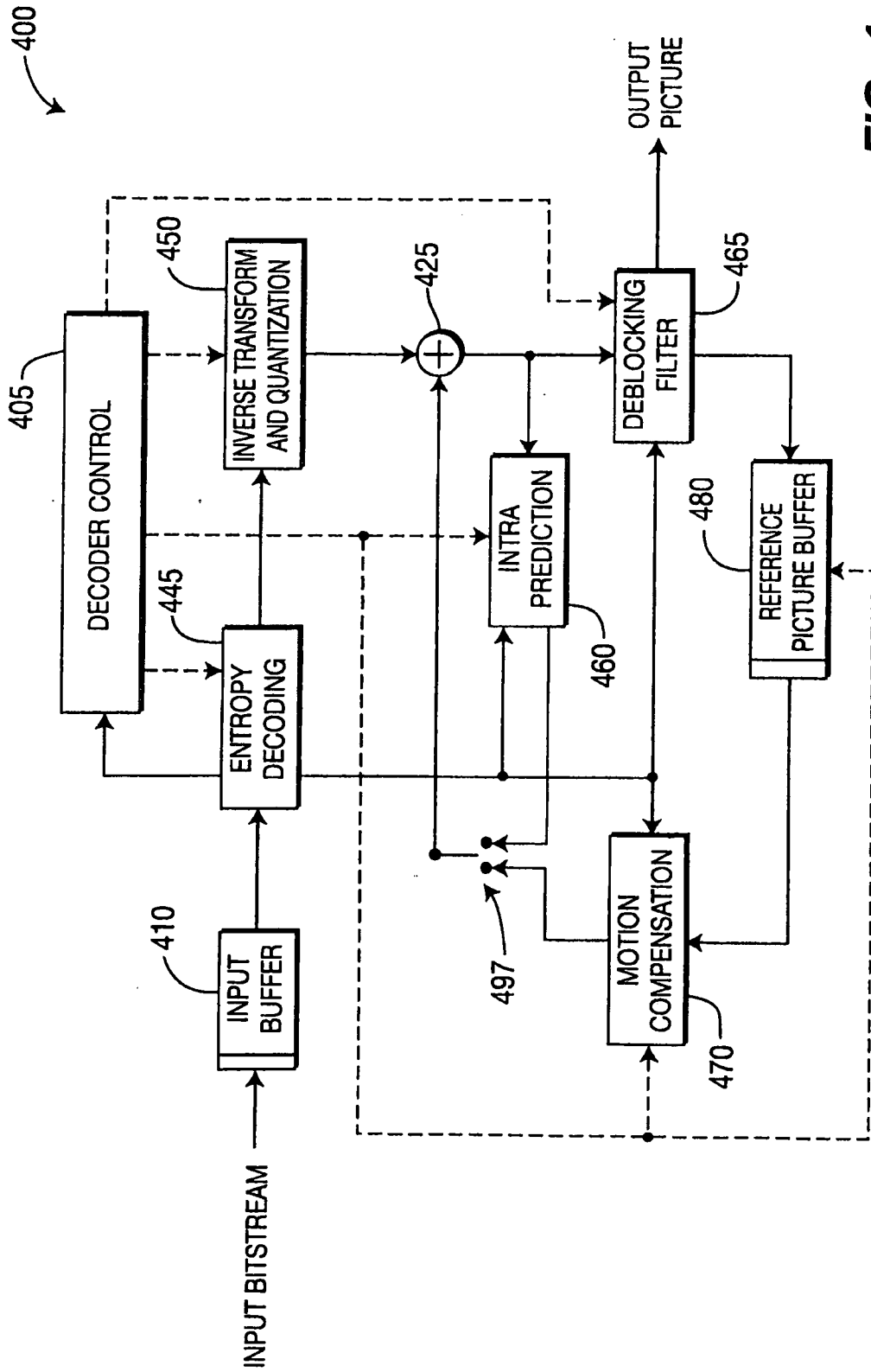
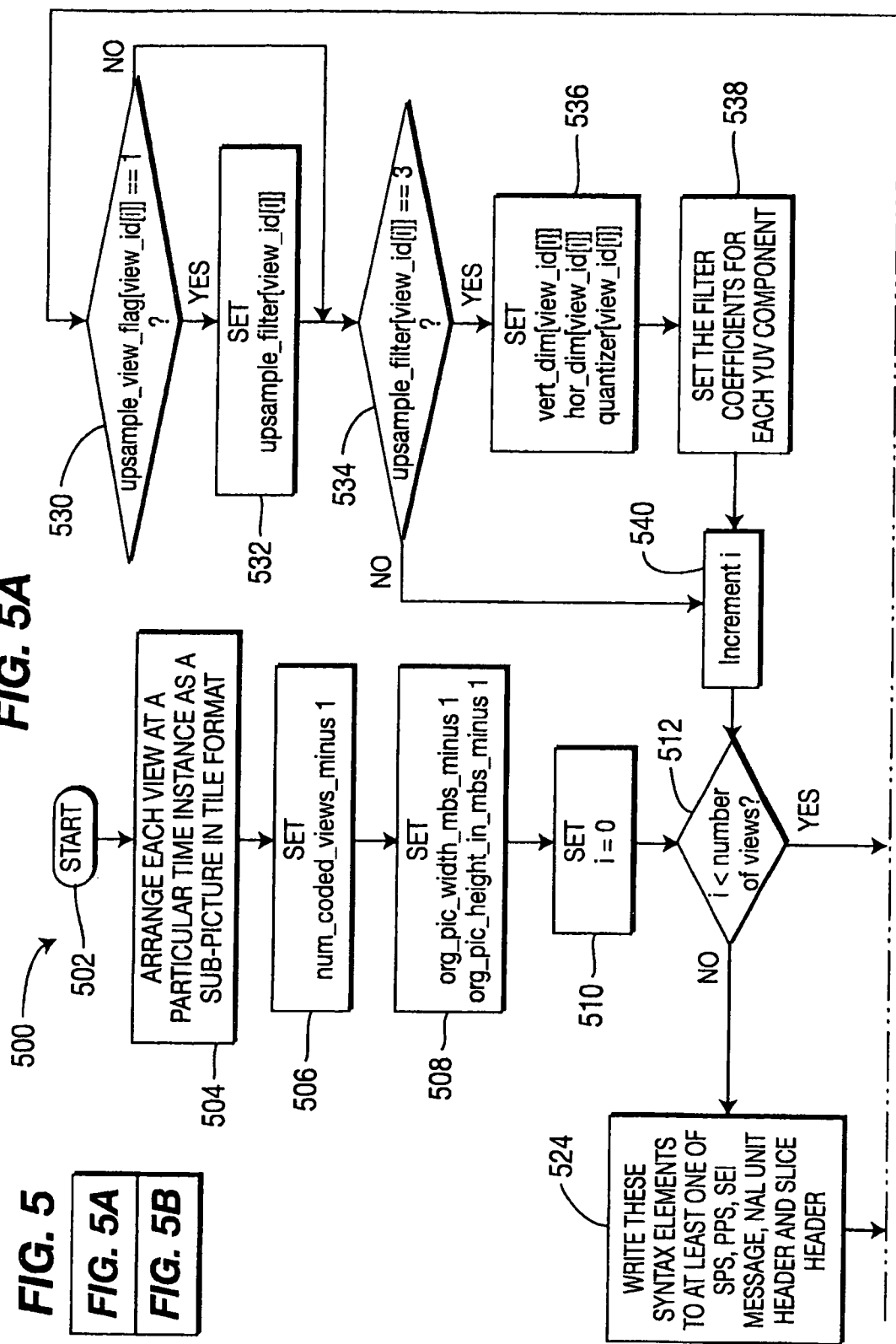


FIG. 4

FIG. 5

FIG. 5A
FIG. 5B

FIG. 5A



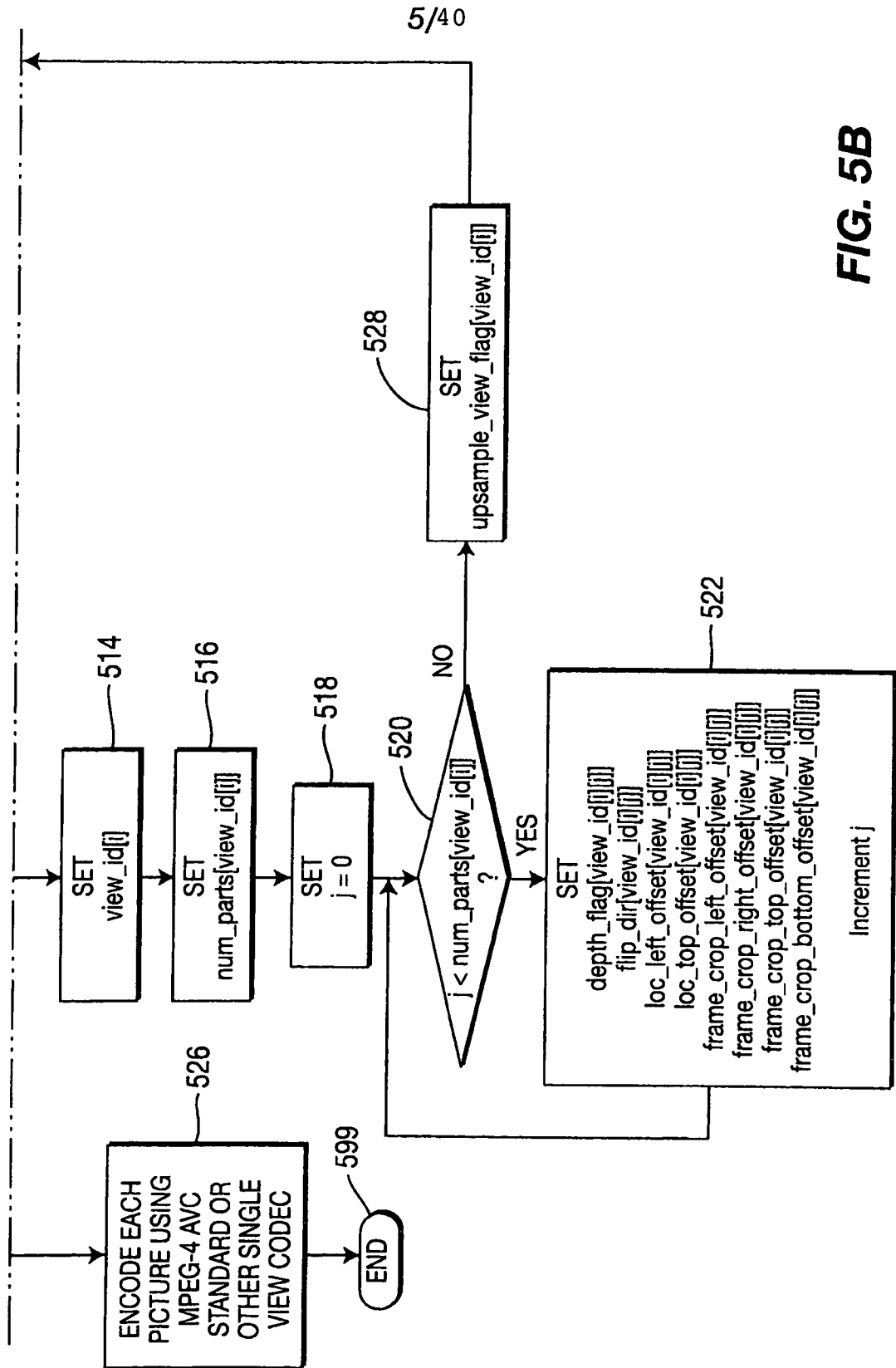


FIG. 5B

FIG. 6A

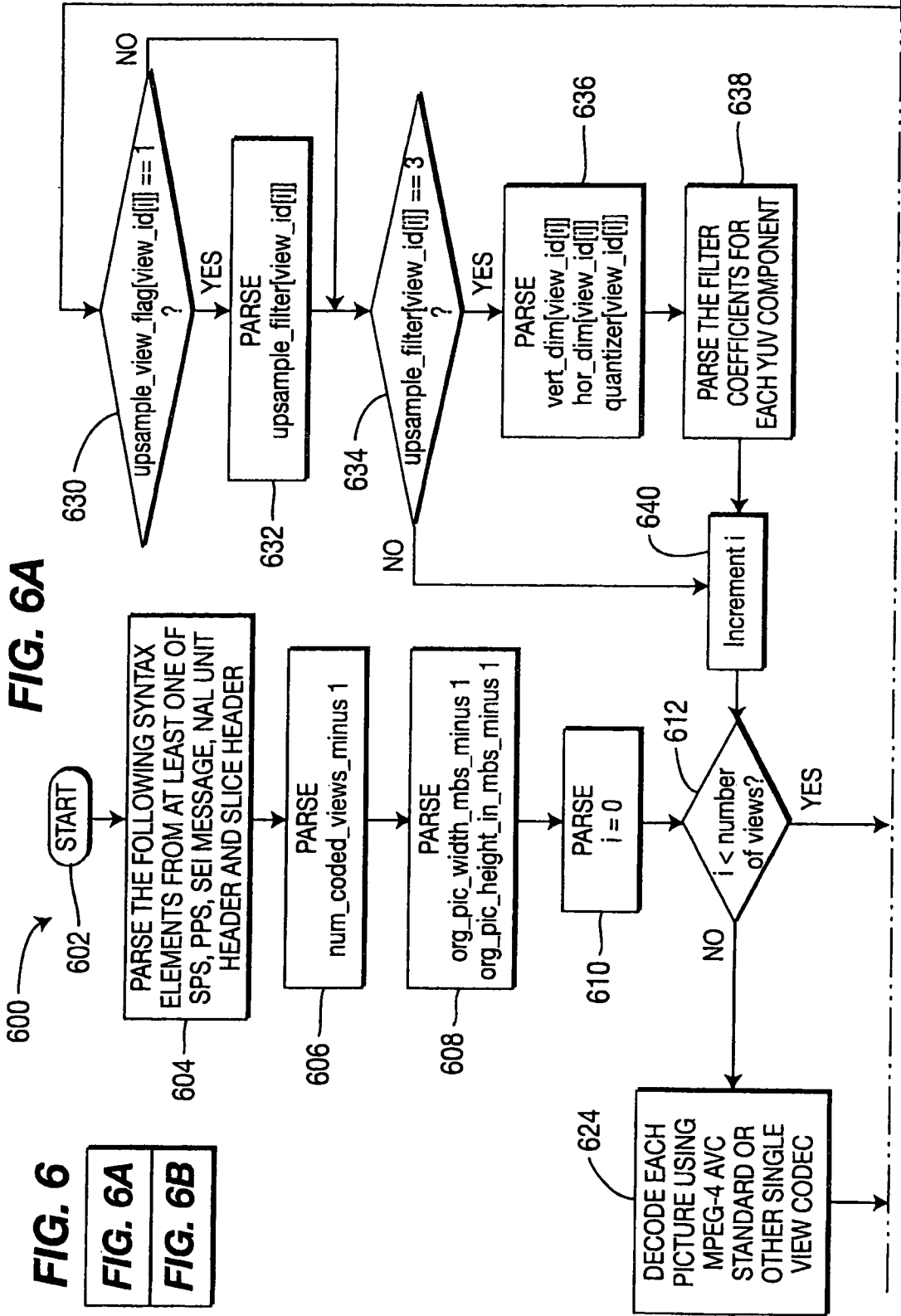


FIG. 6

FIG. 6A

FIG. 6B

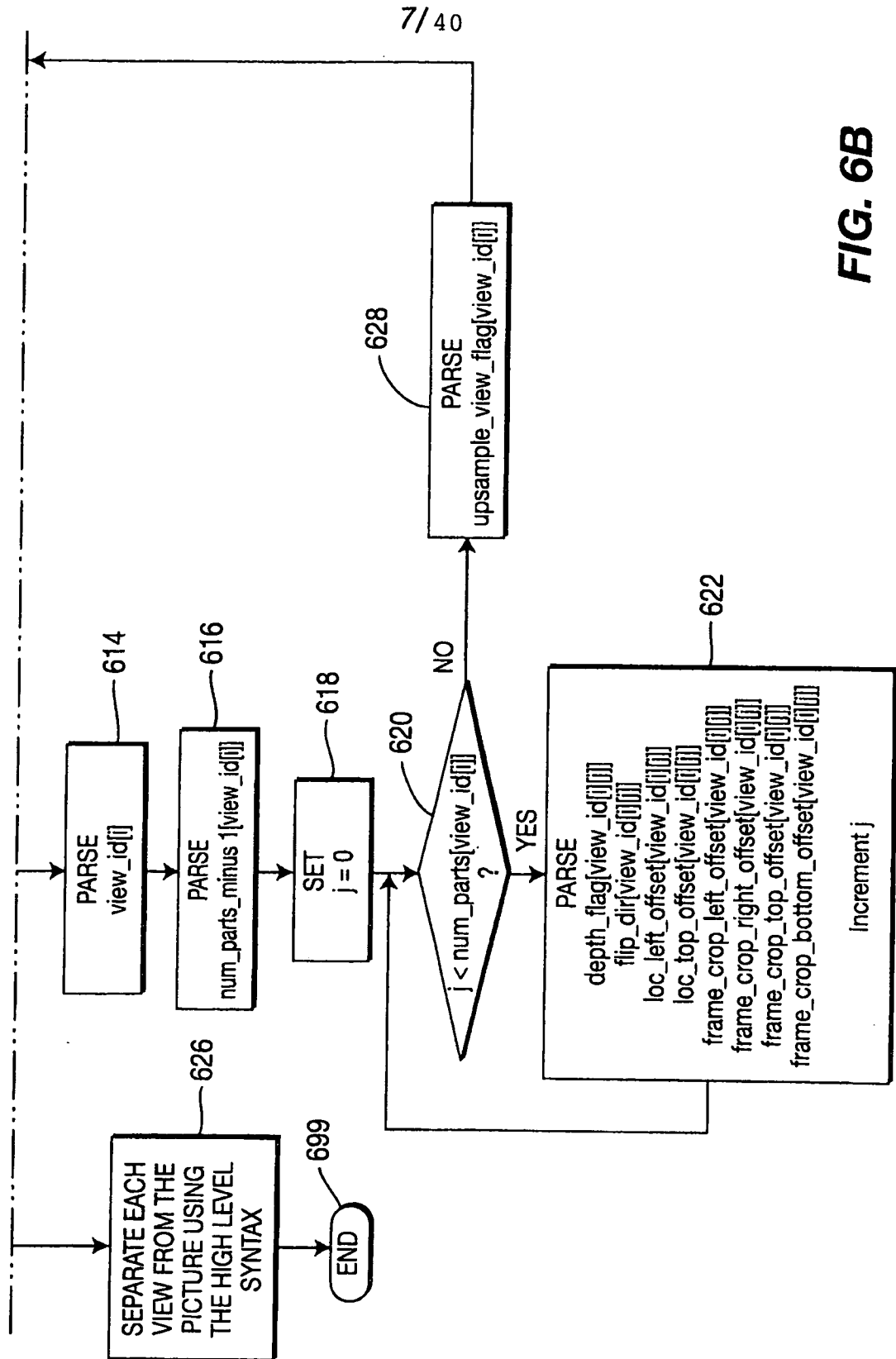


FIG. 6B

FIG. 7A

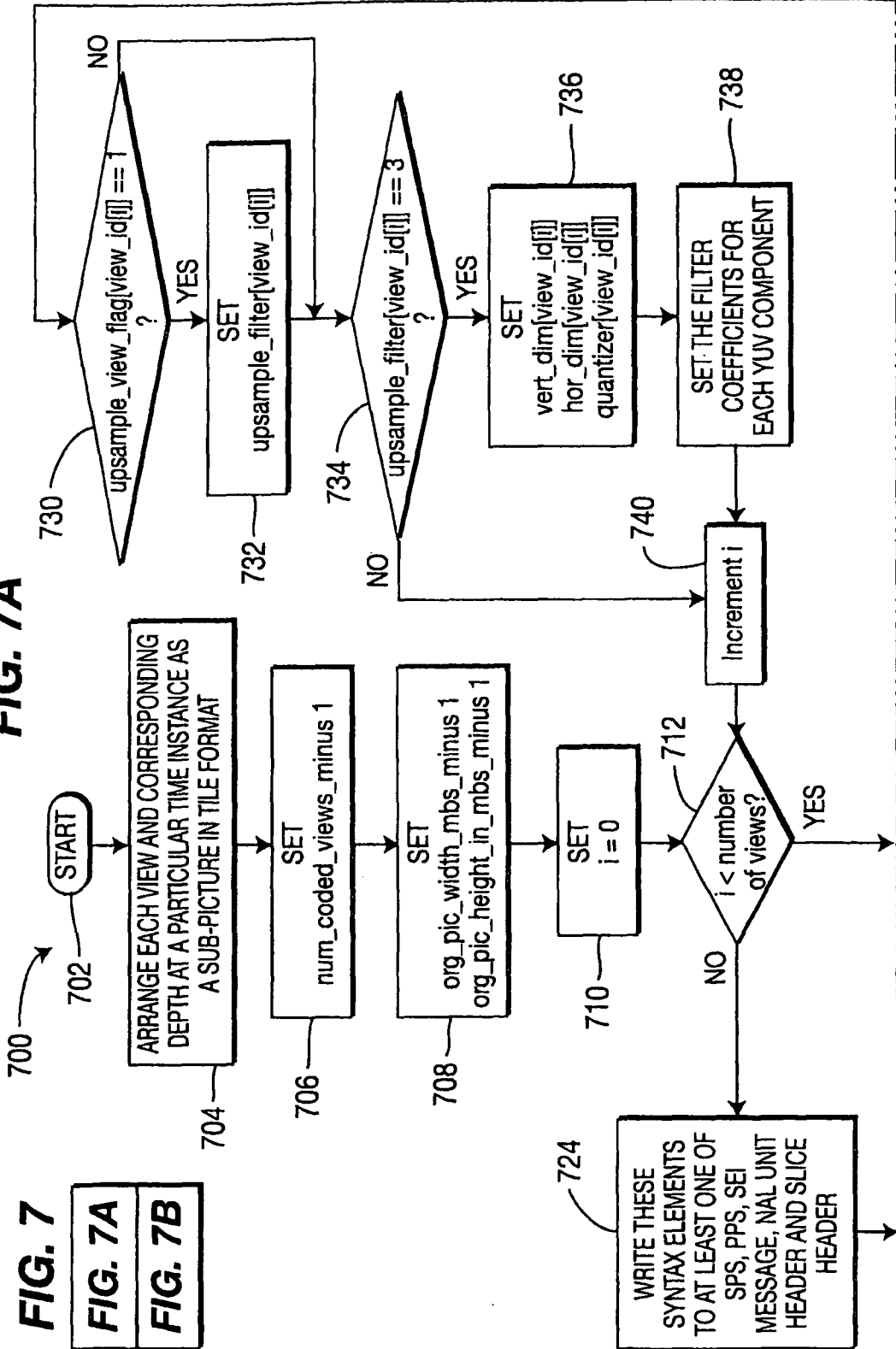
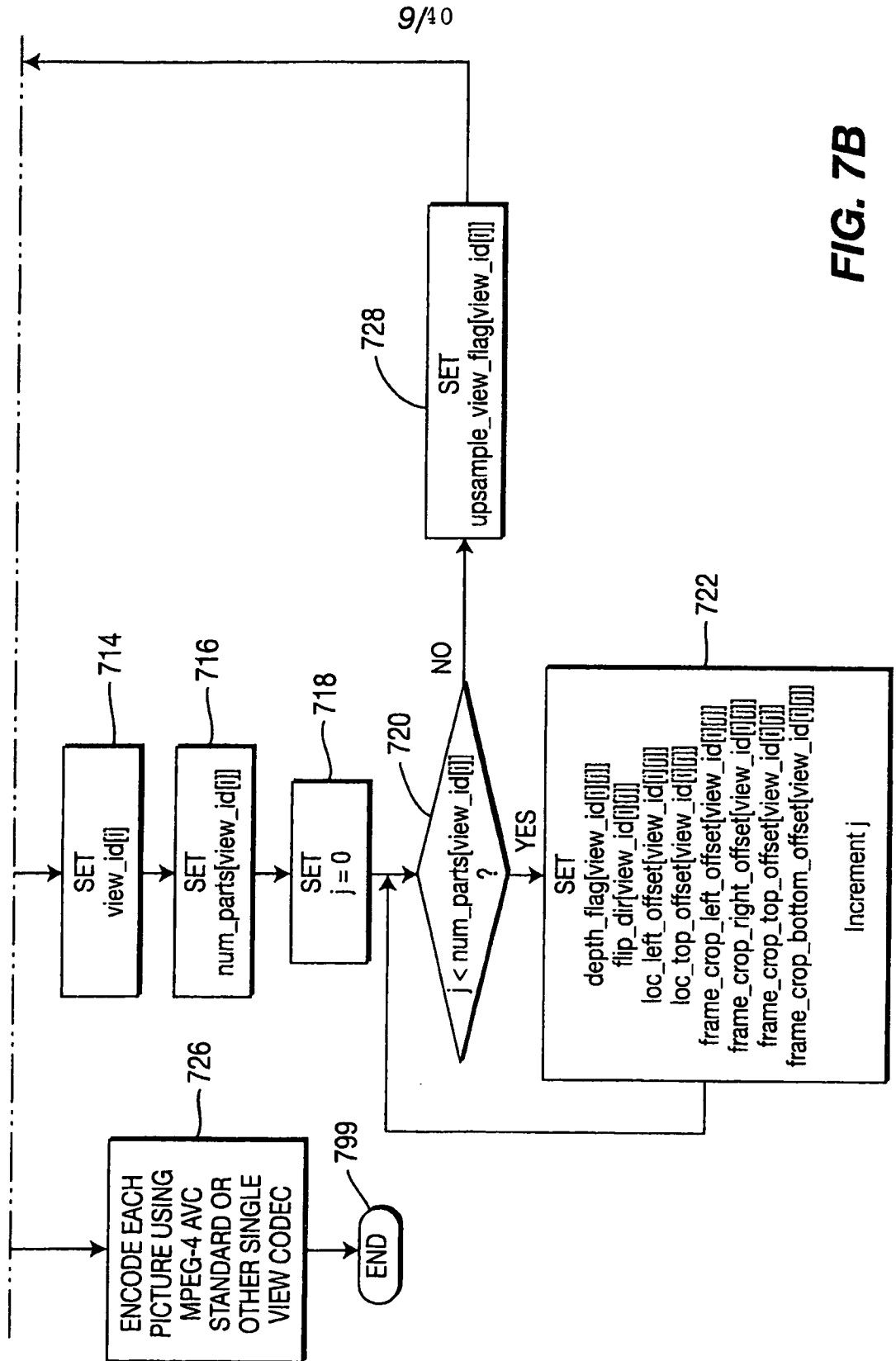


FIG. 7

FIG. 7A

FIG. 7B



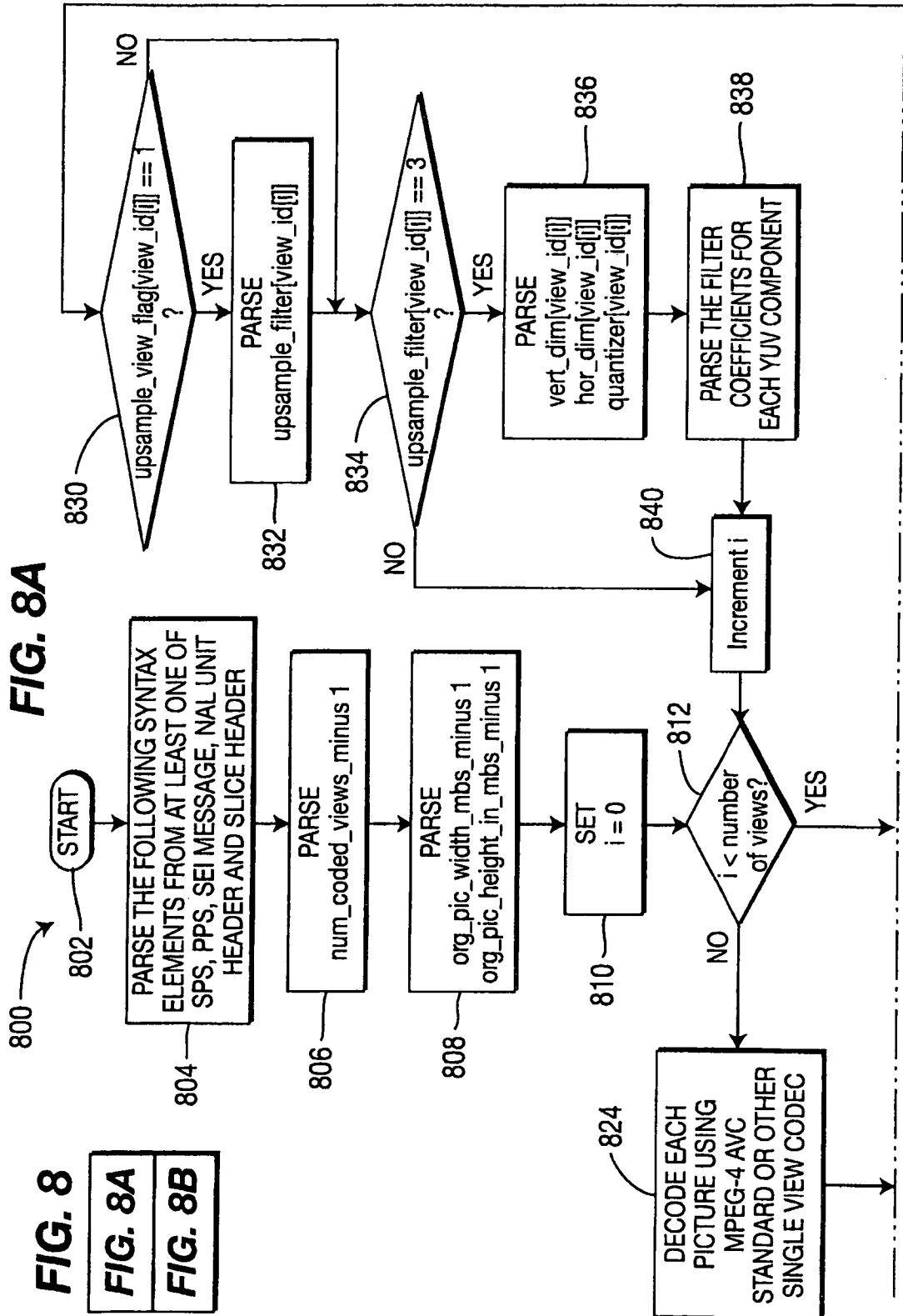
9/40

FIG. 7B

FIG. 8A

FIG. 8

FIG. 8A
FIG. 8B



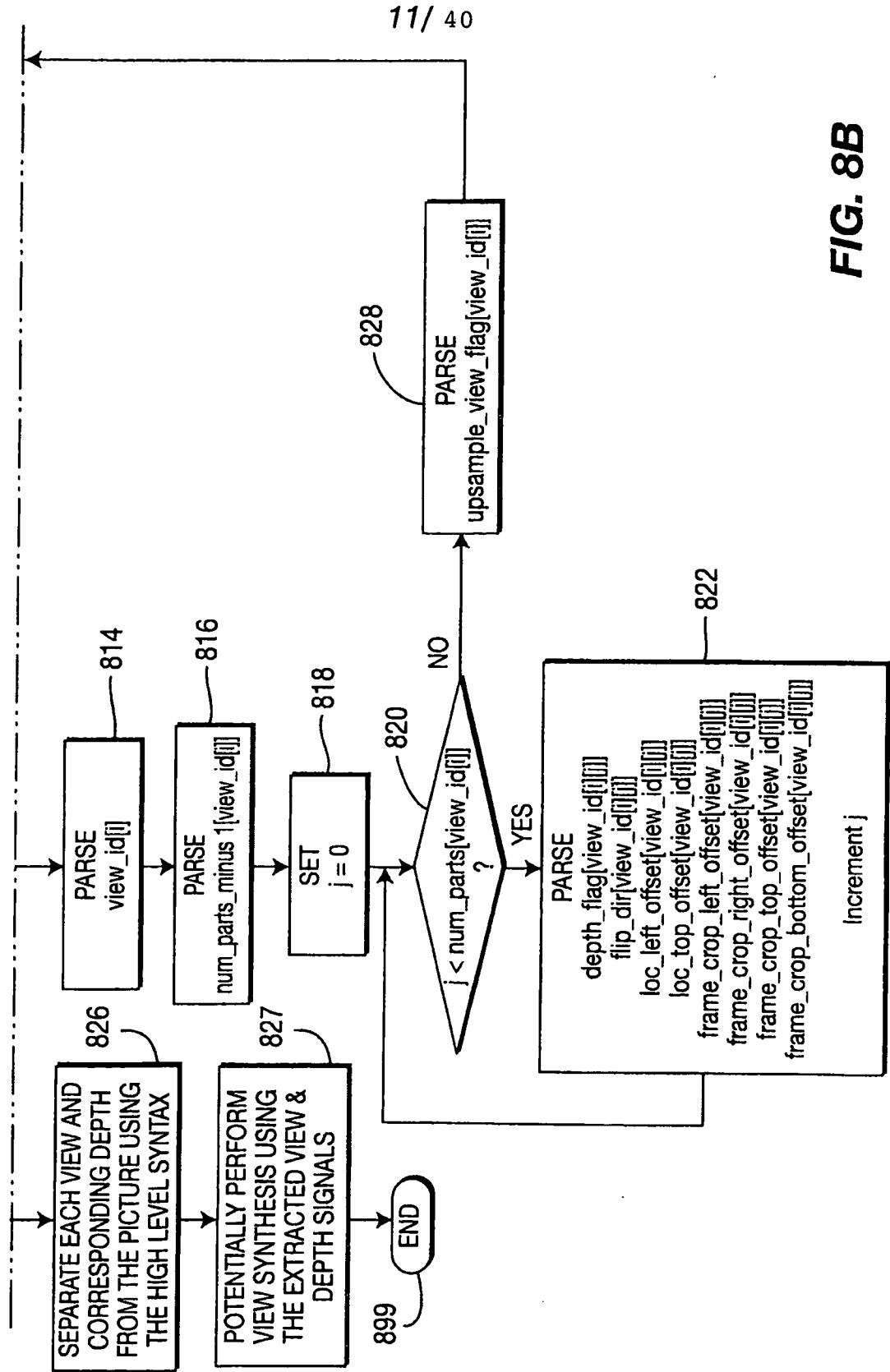


FIG. 8B

12/40

900



FIG. 9

1000



FIG. 10

1100

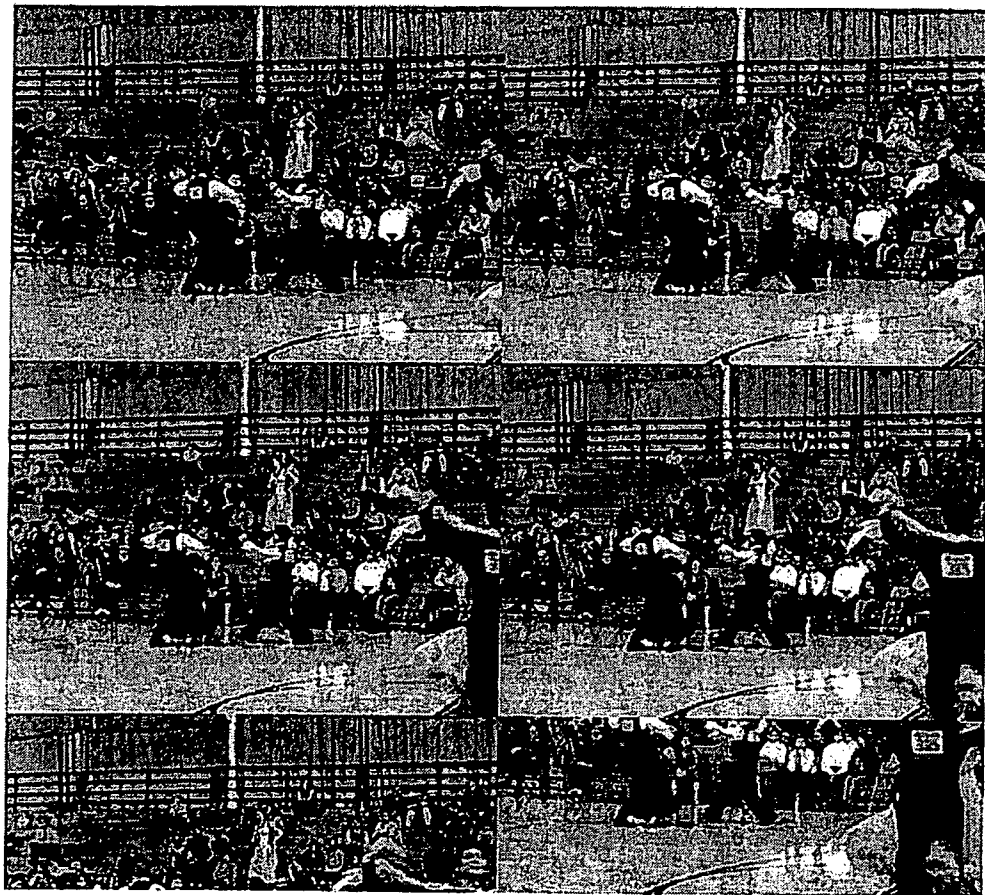


FIG. 11

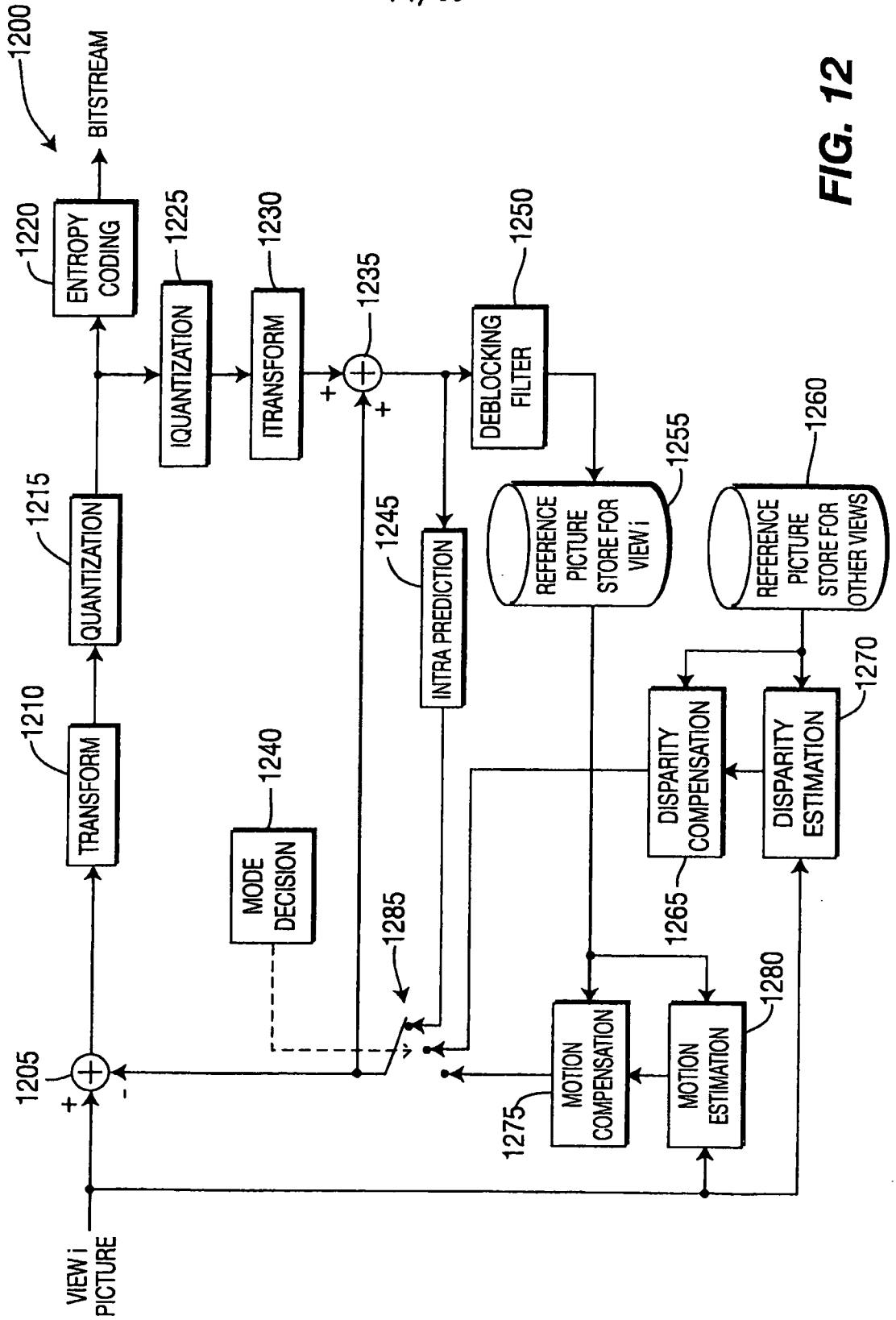


FIG. 12

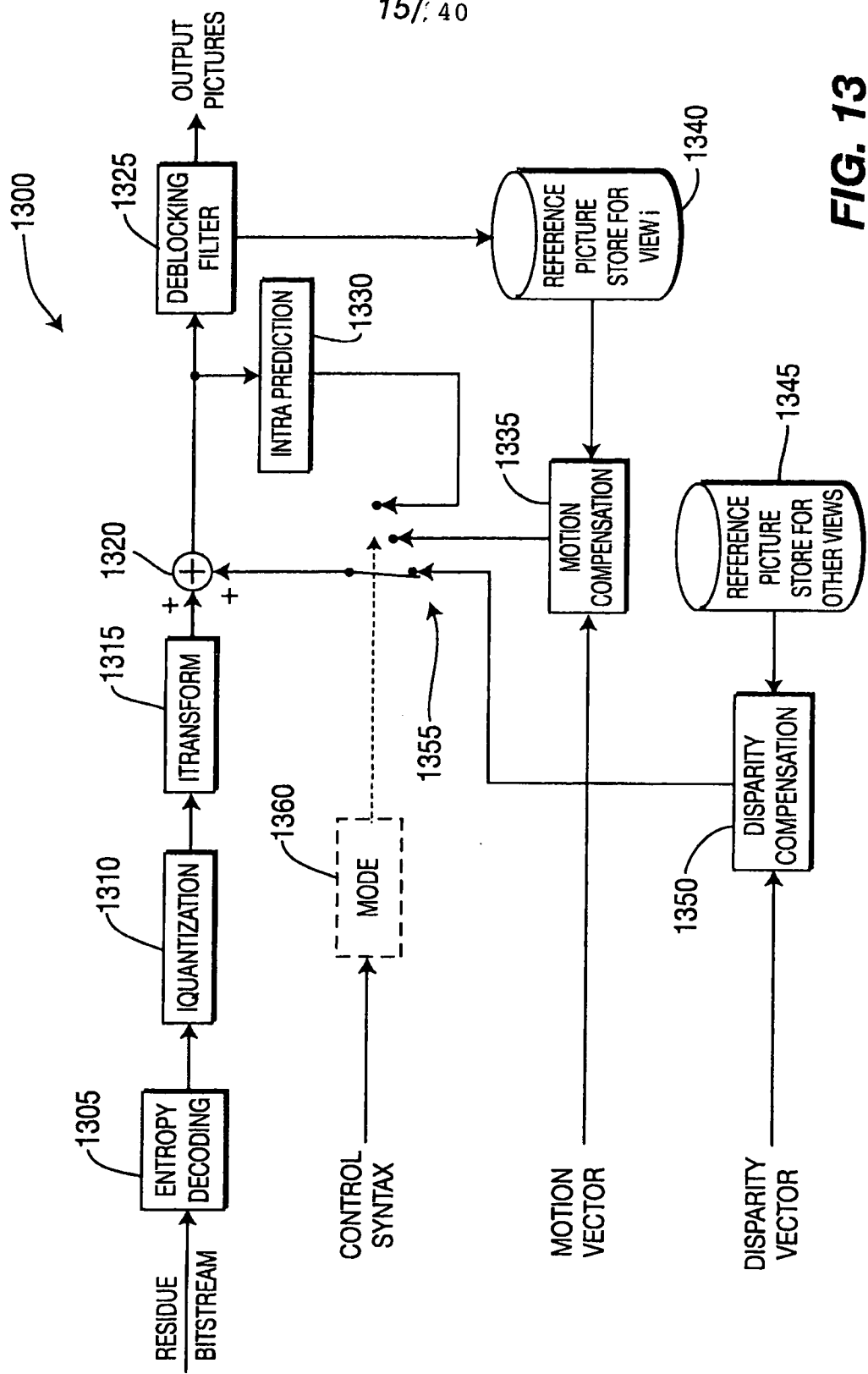


FIG. 13

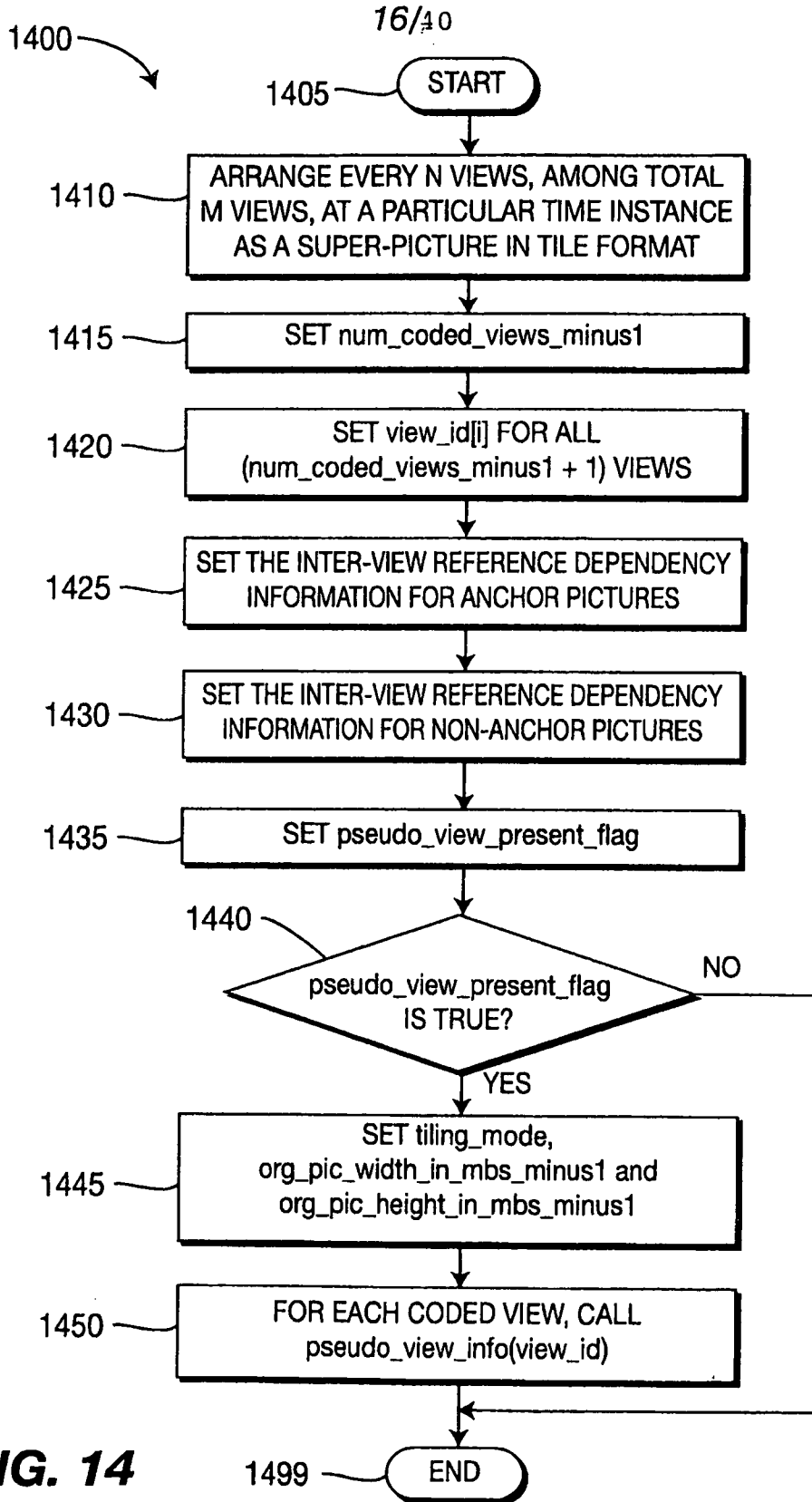
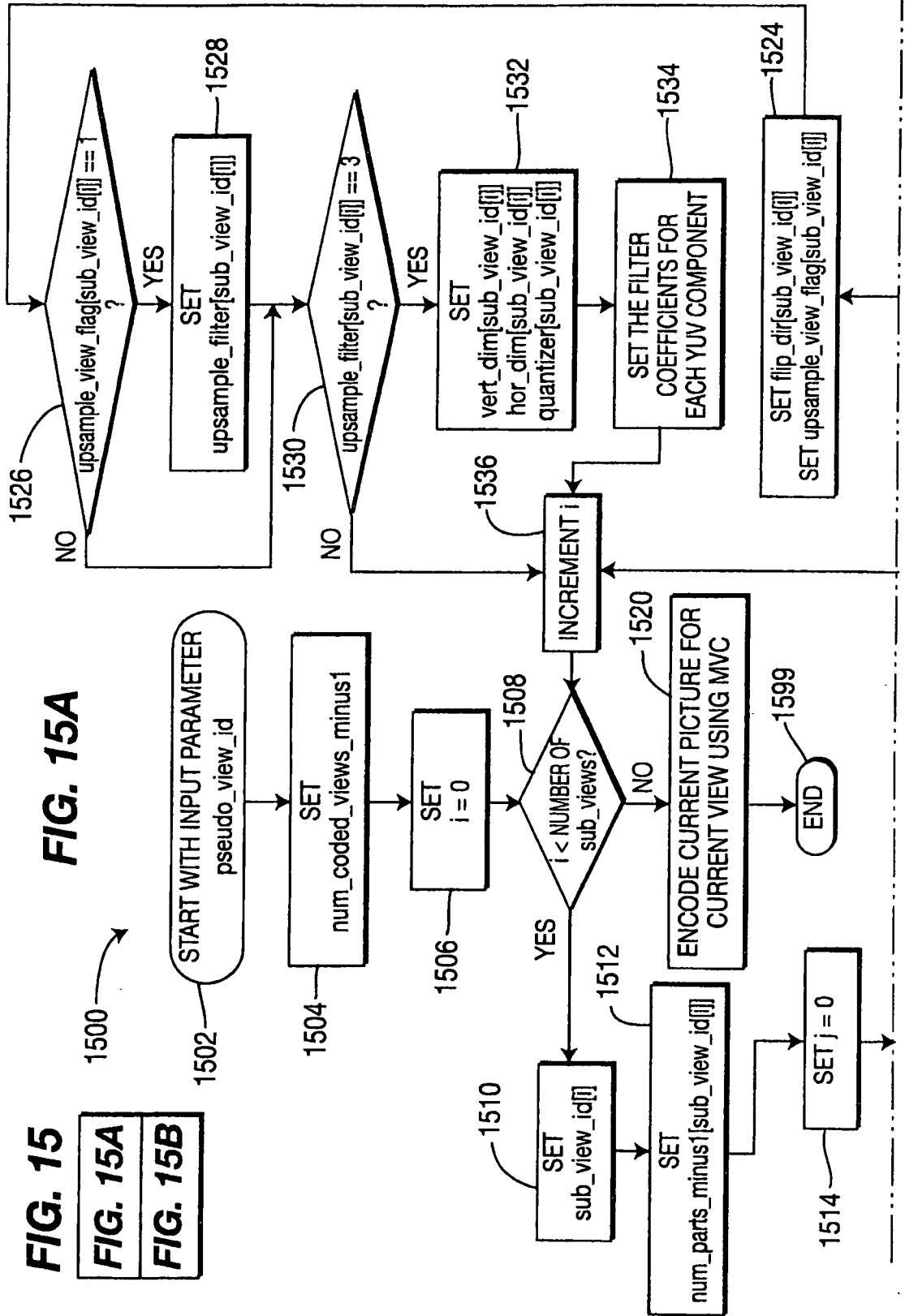


FIG. 14

FIG. 15
FIG. 15A
FIG. 15B



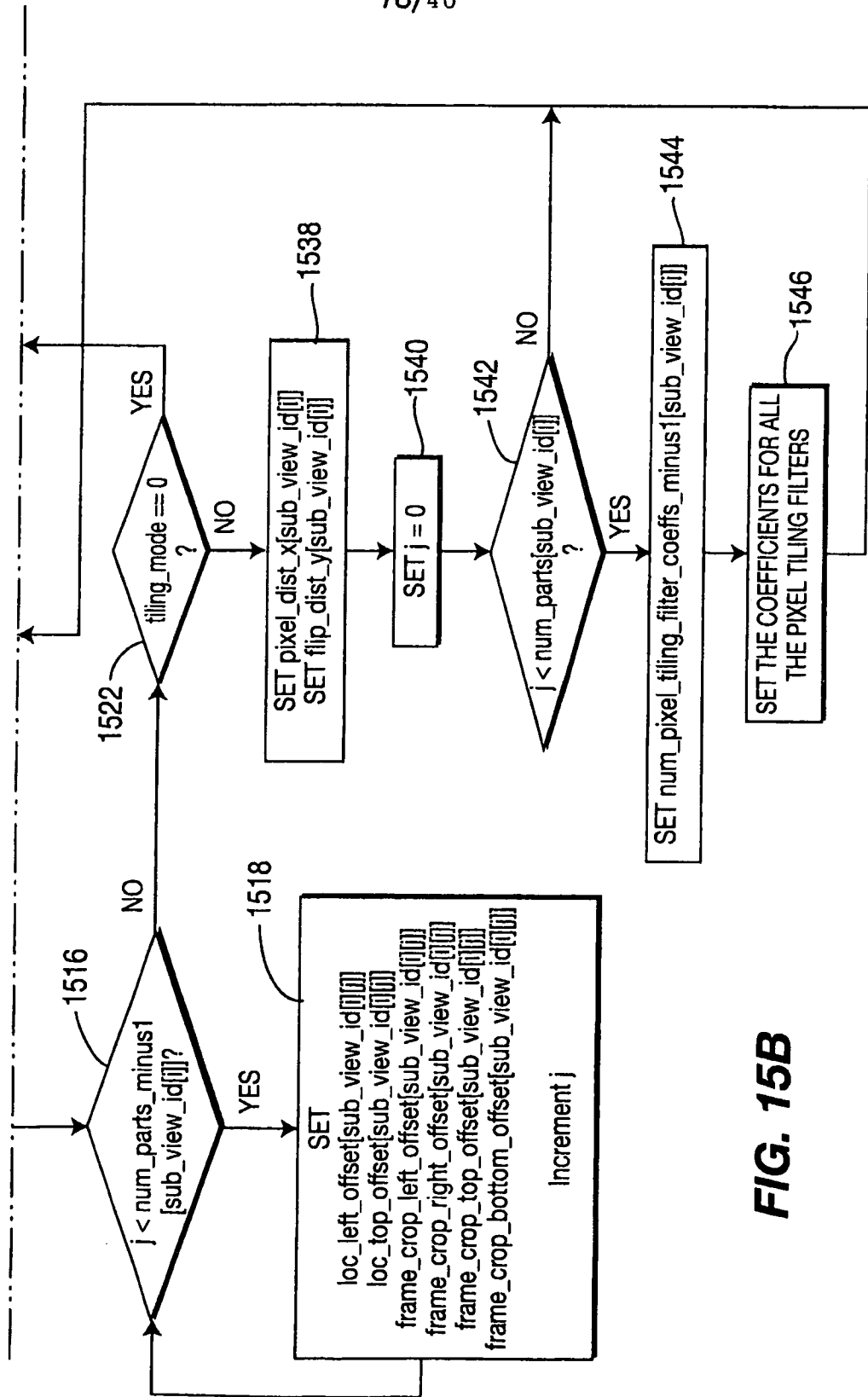


FIG. 15B

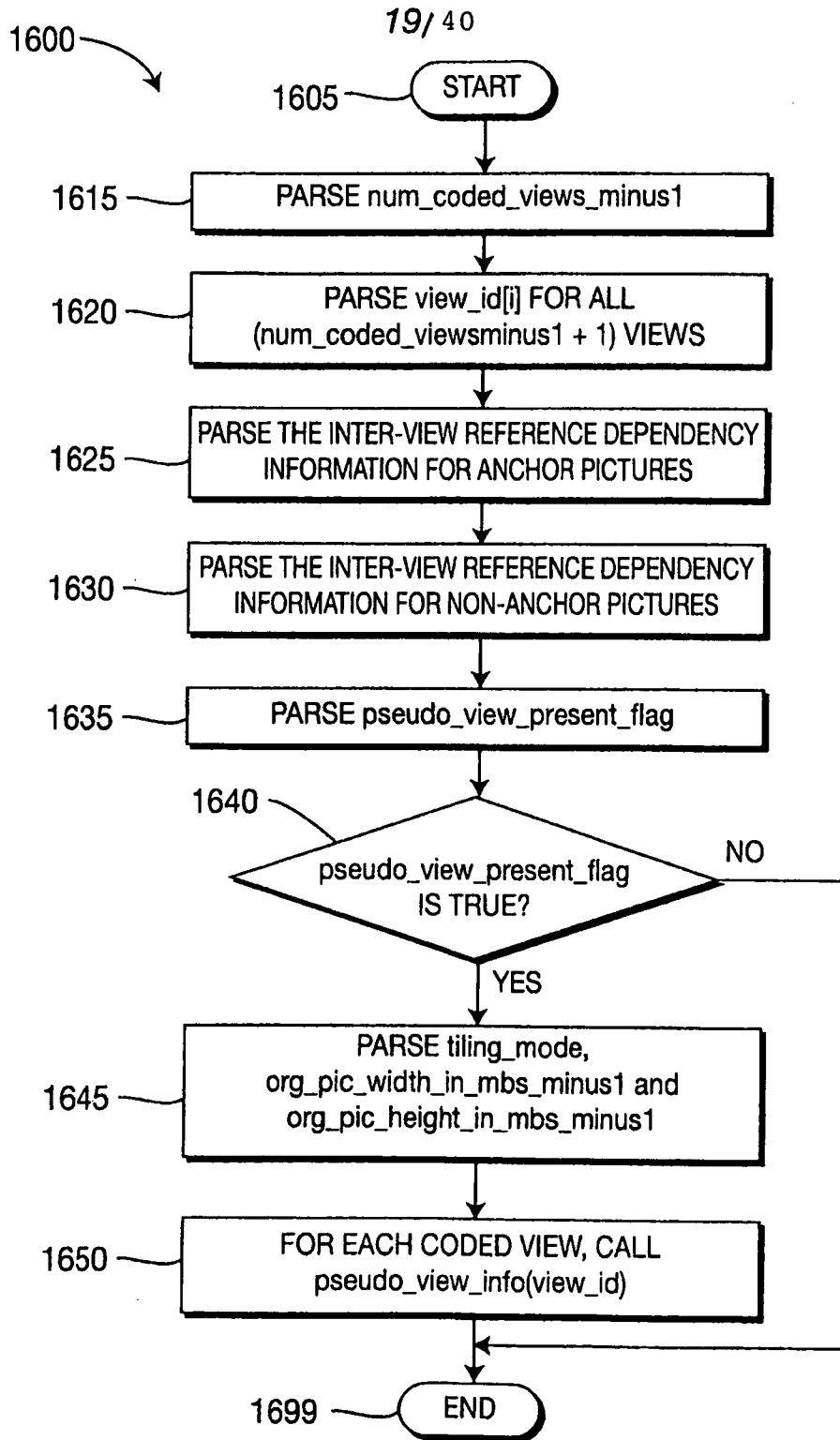


FIG. 16

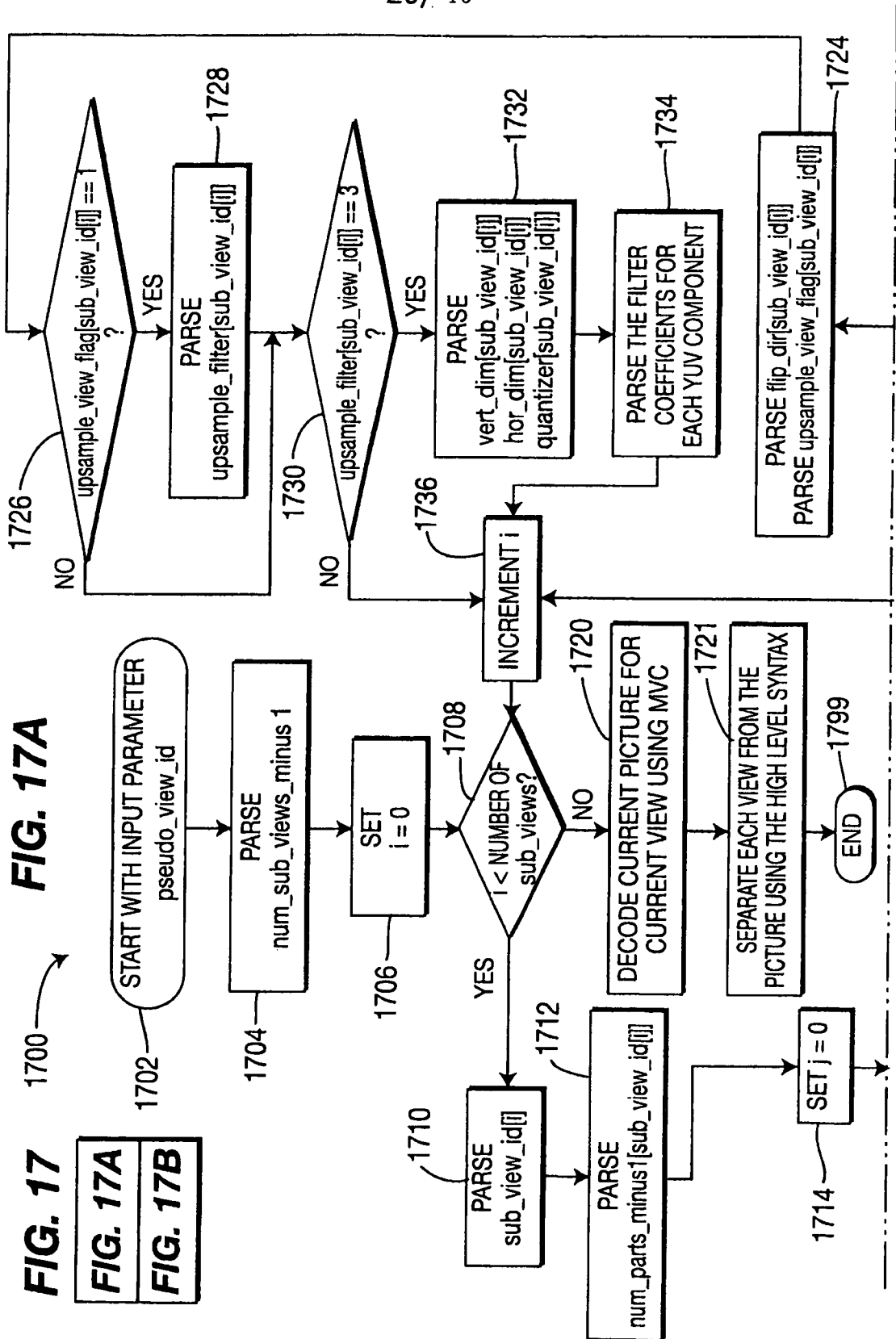


FIG. 17A

FIG. 17

FIG. 17A

FIG. 17B

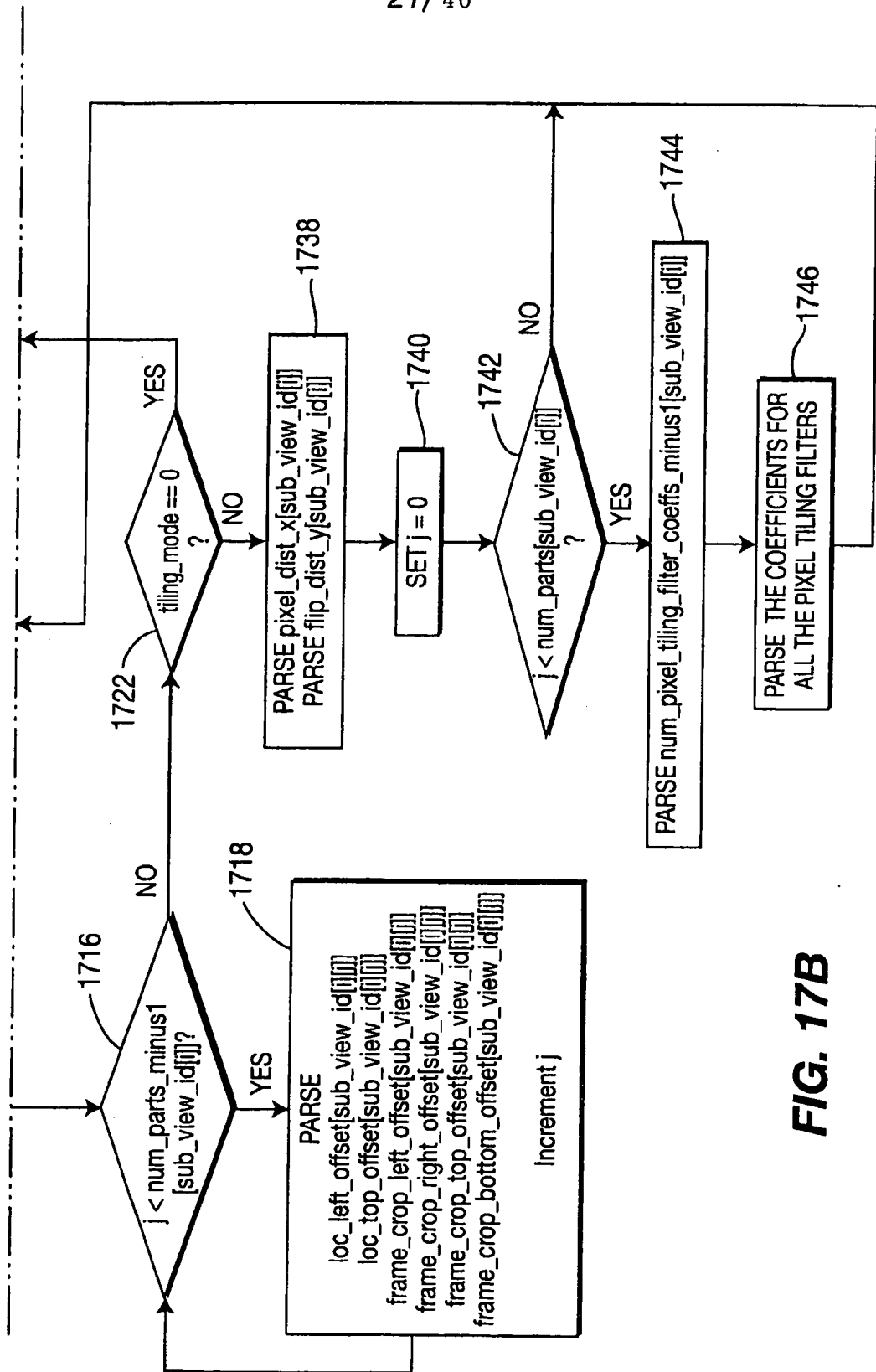


FIG. 17B

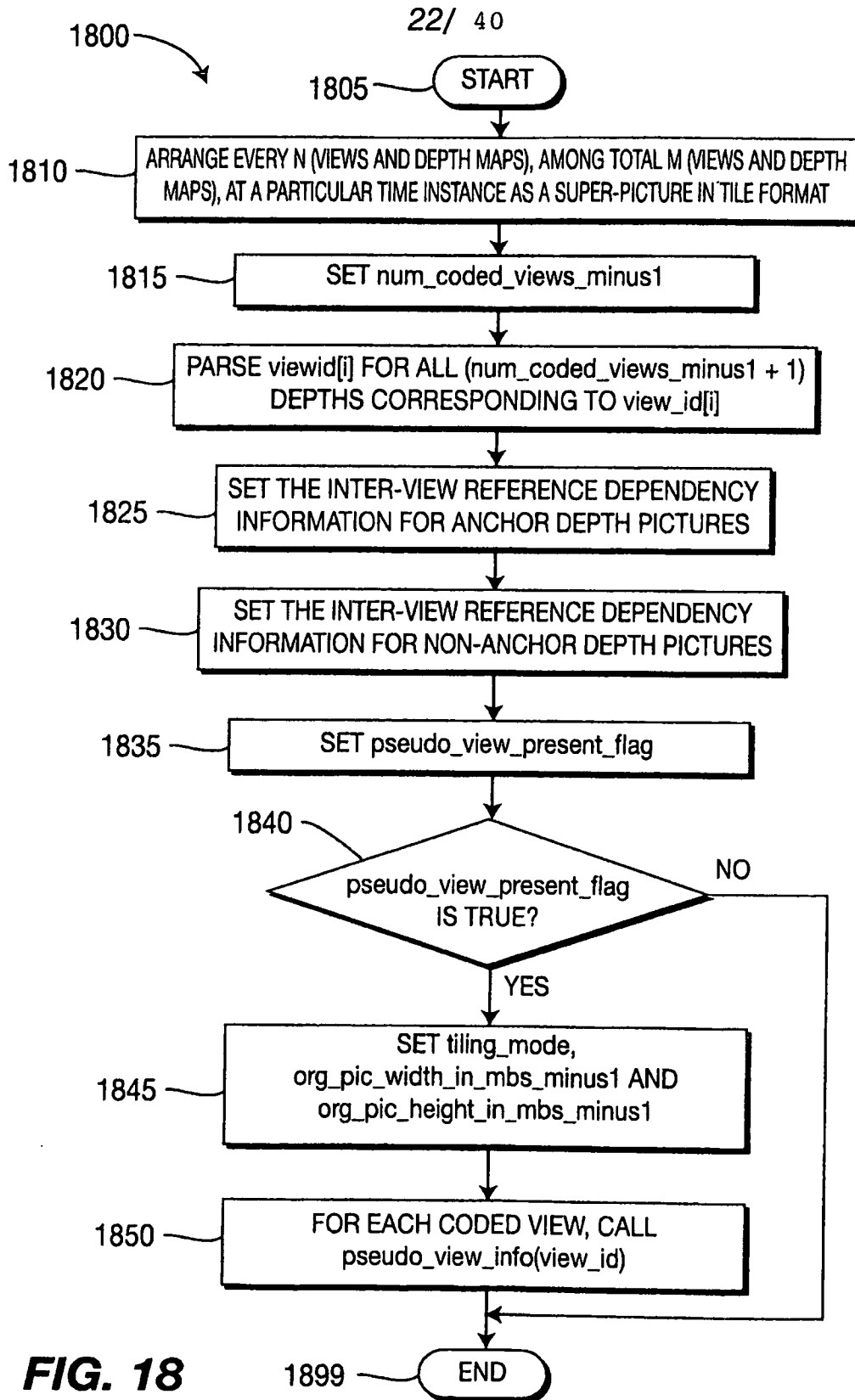
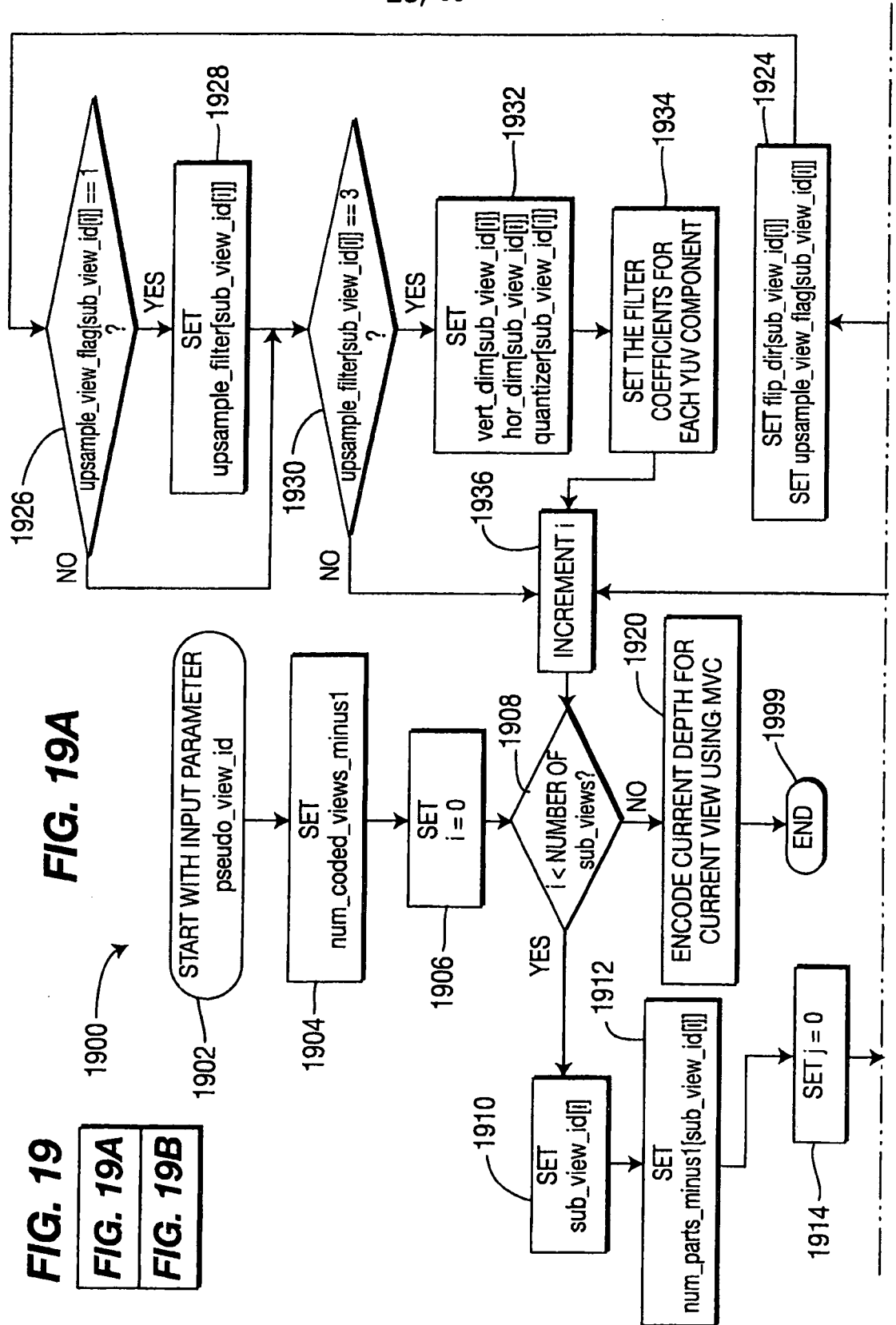


FIG. 18

FIG. 19
FIG. 19A
FIG. 19B



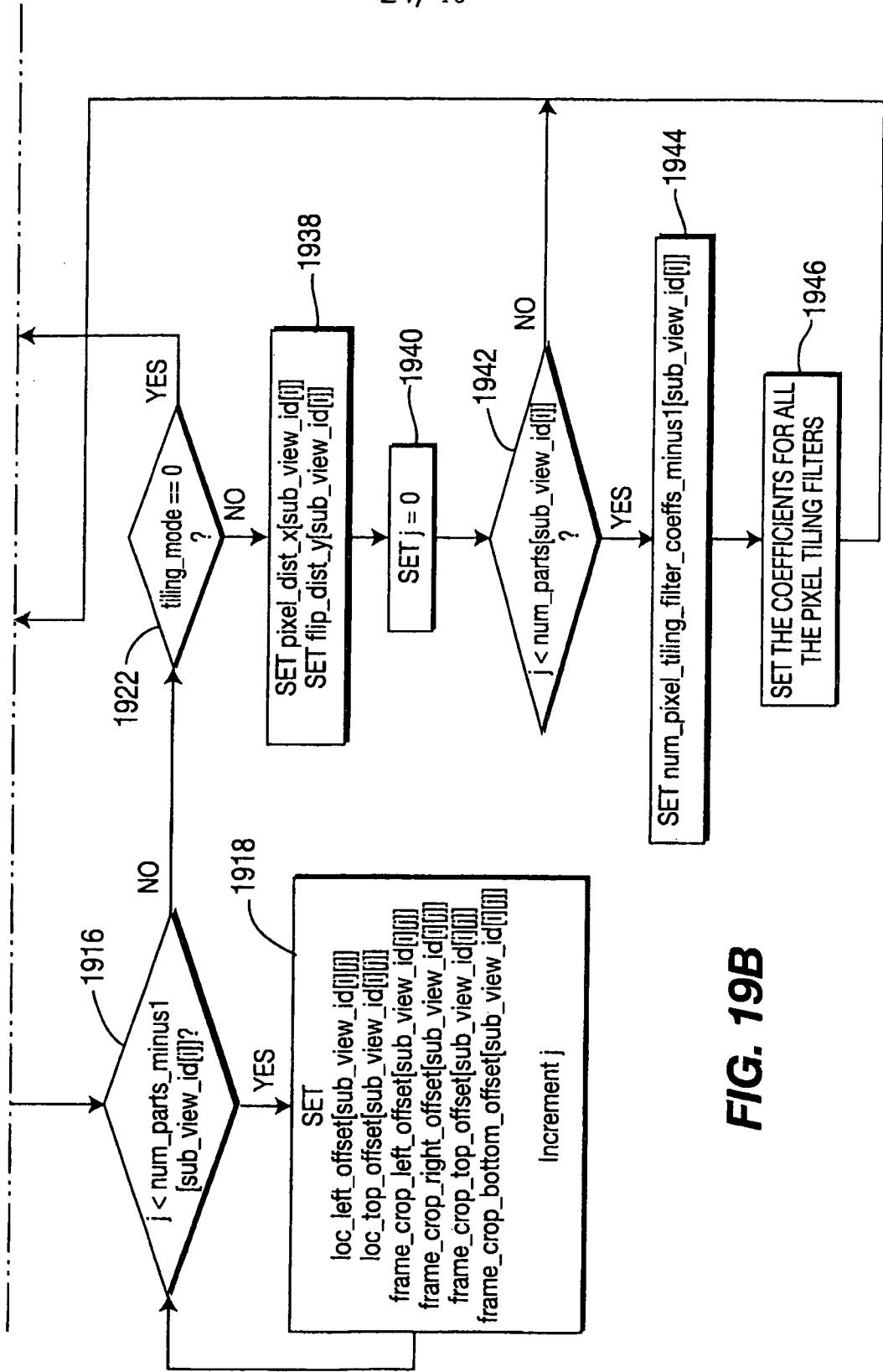


FIG. 19B

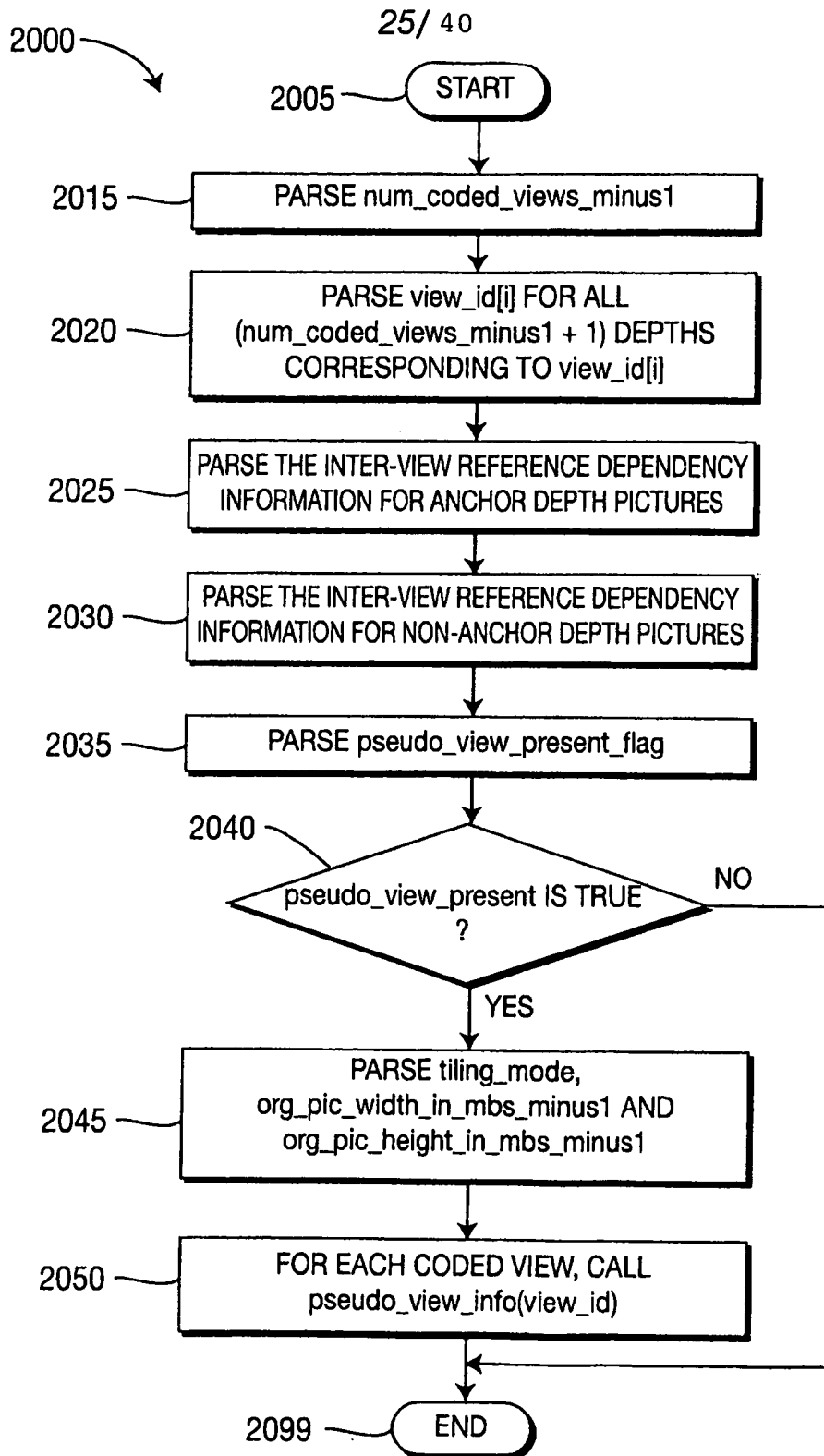
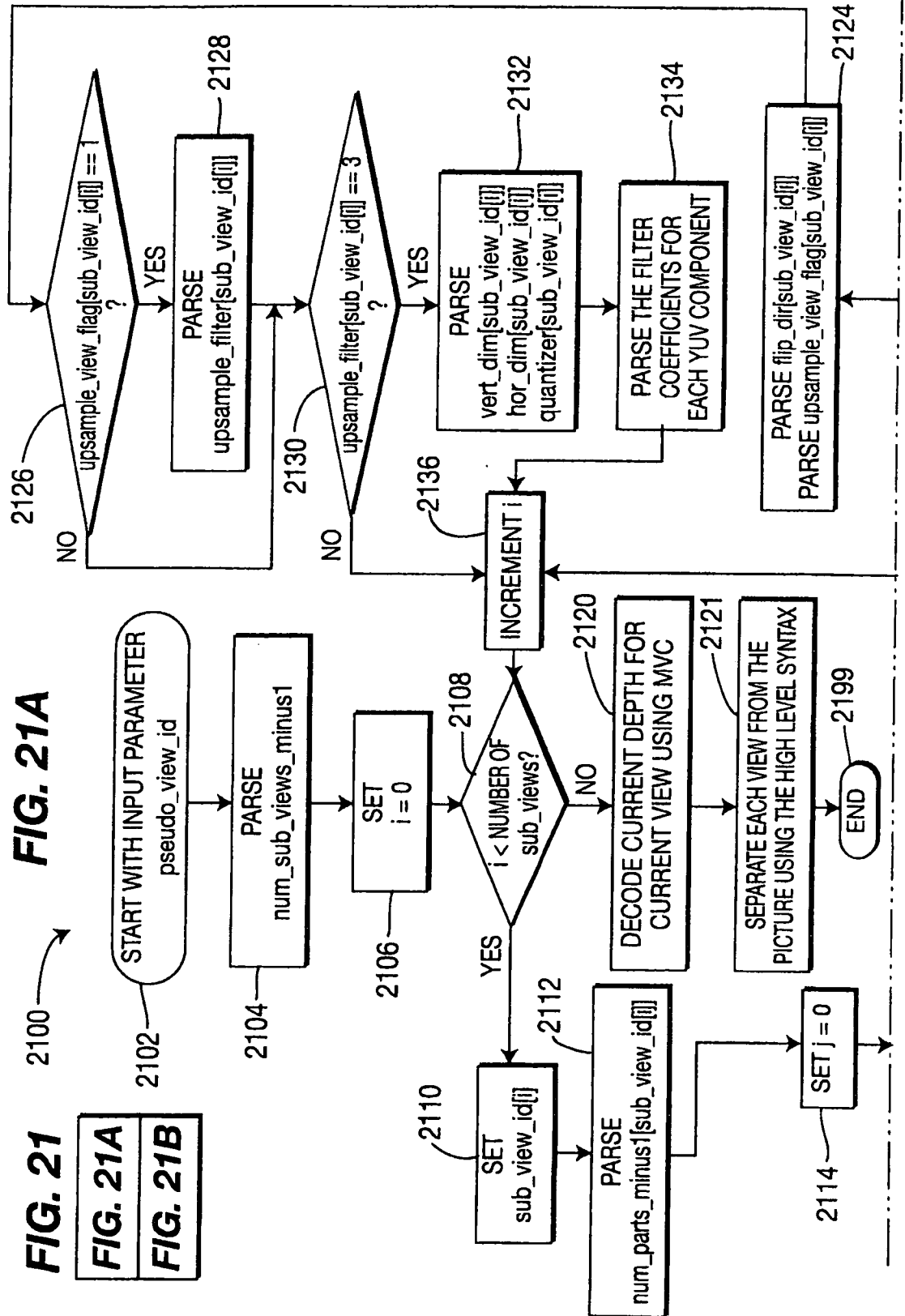


FIG. 20

FIG. 21A

FIG. 21

FIG. 21A
FIG. 21B



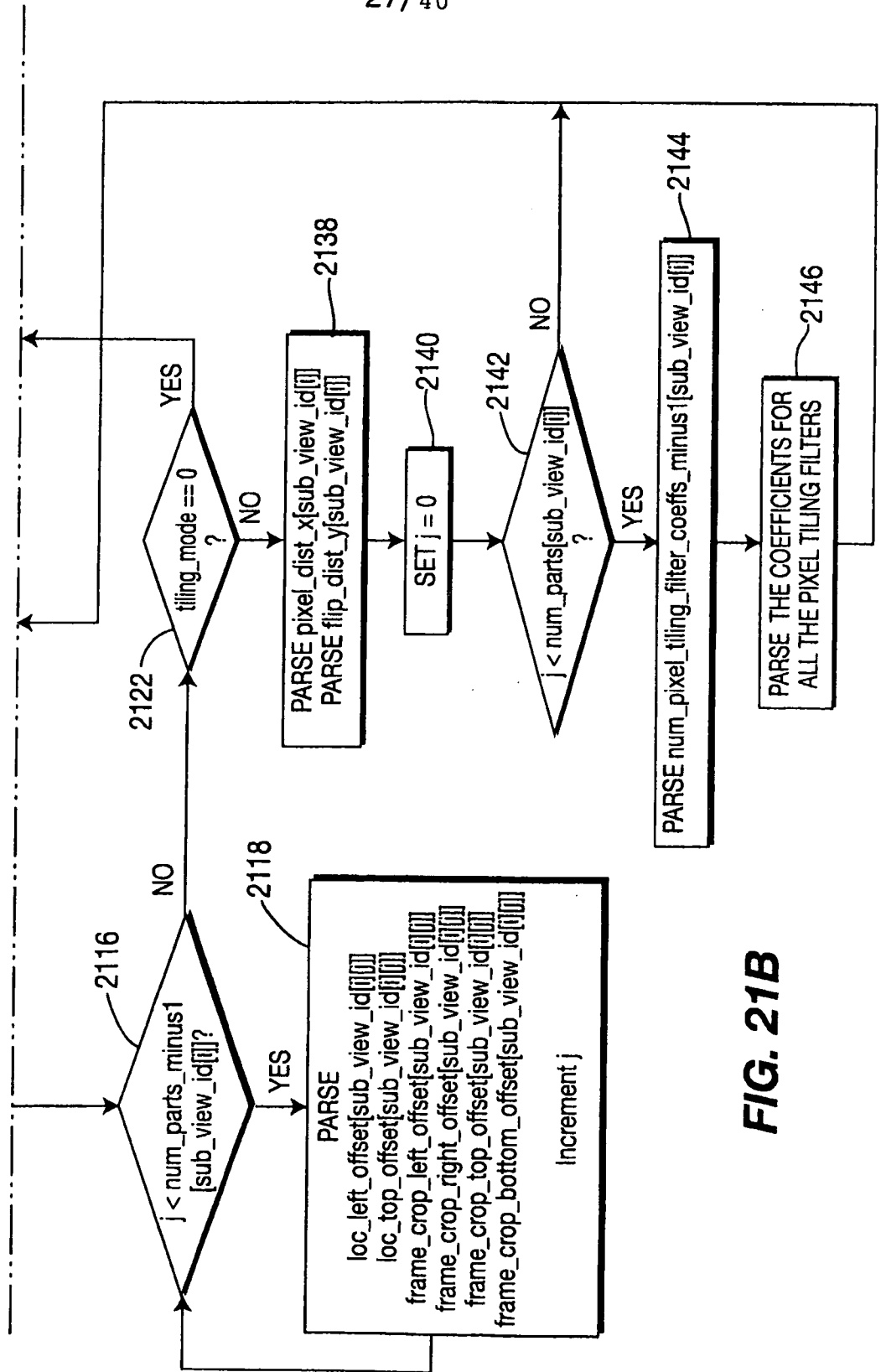


FIG. 21B

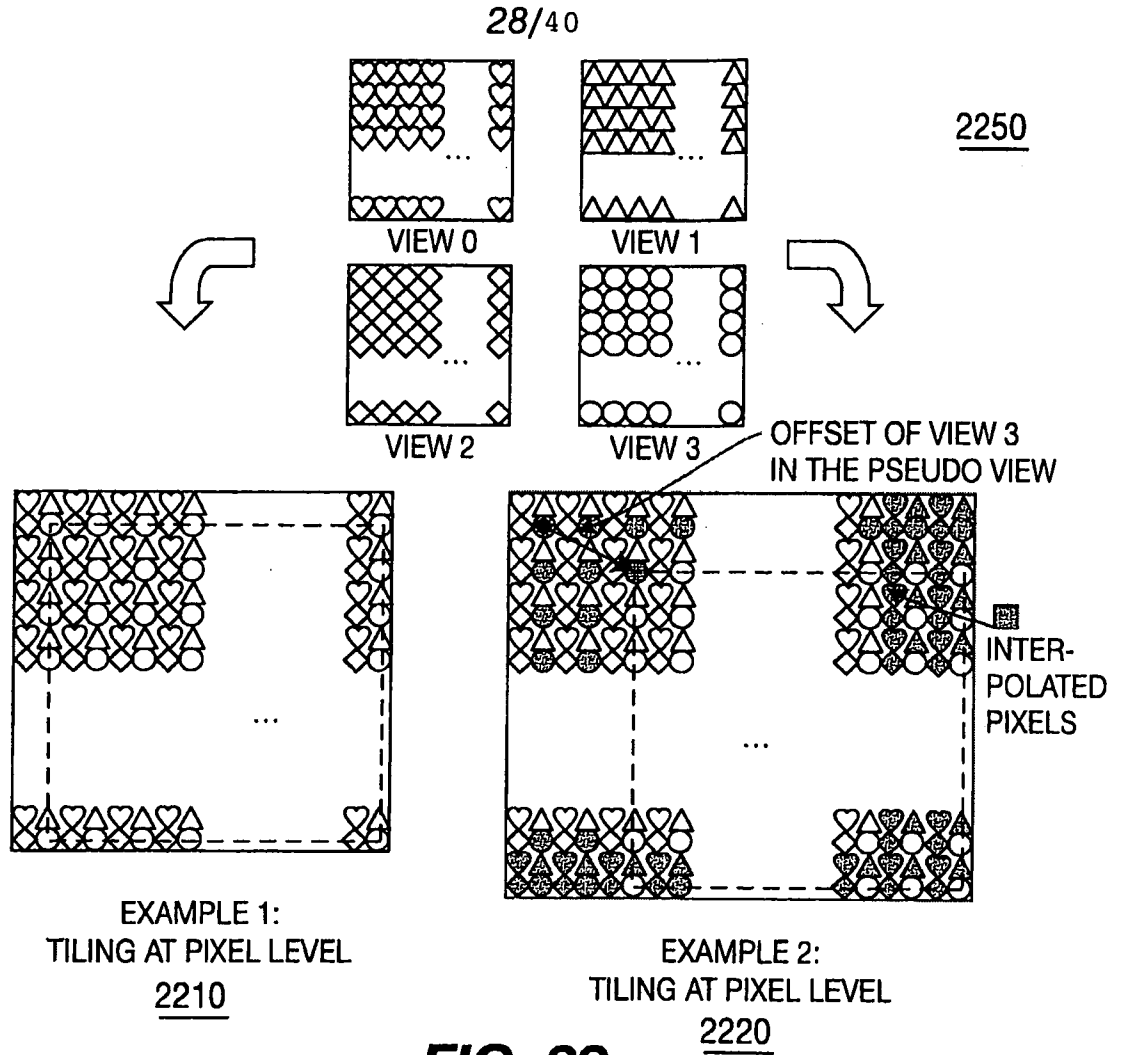


FIG. 22

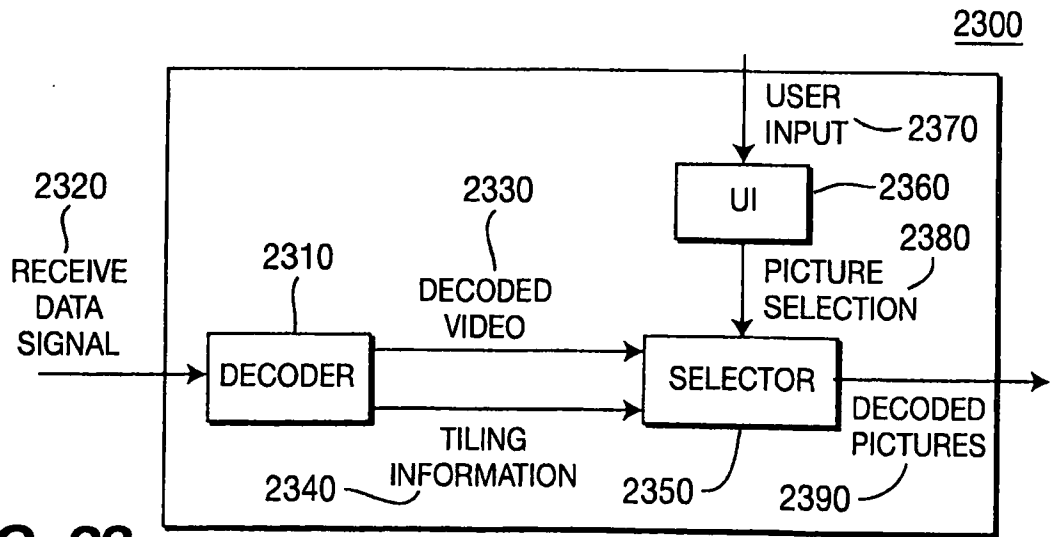


FIG. 23

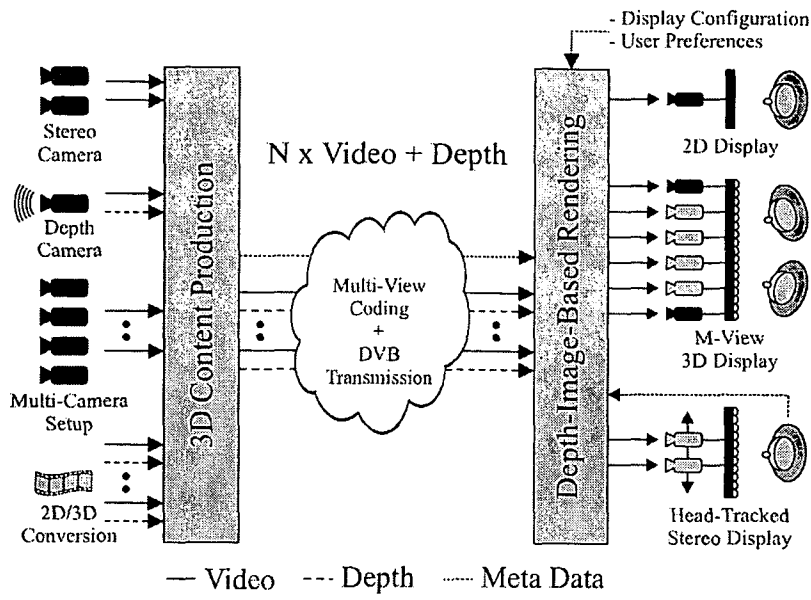
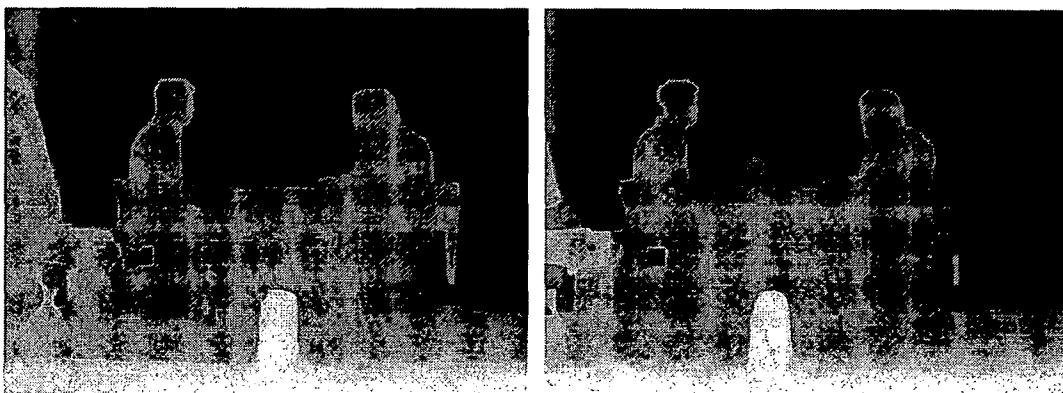


Figure 24



(a) Left depth map

(b) Right depth map

Figure 25

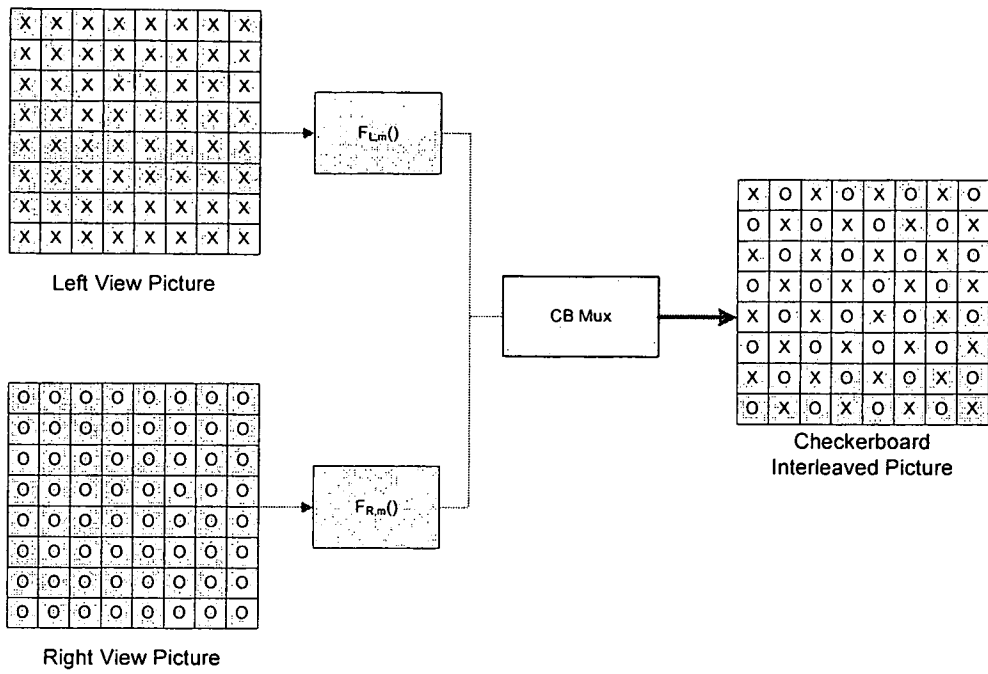


Figure 26



Figure 27

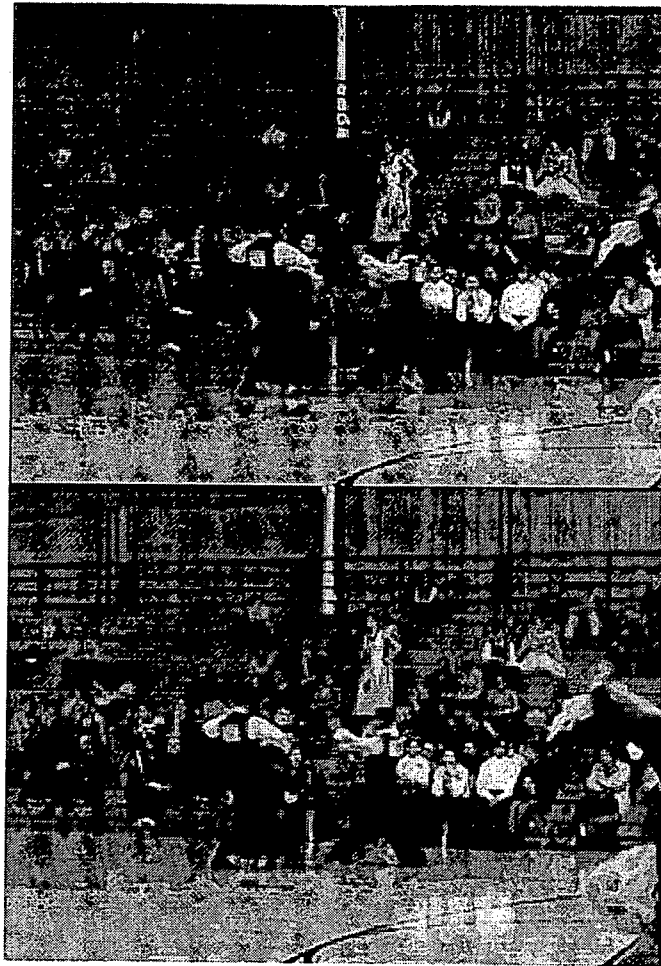


Figure 28

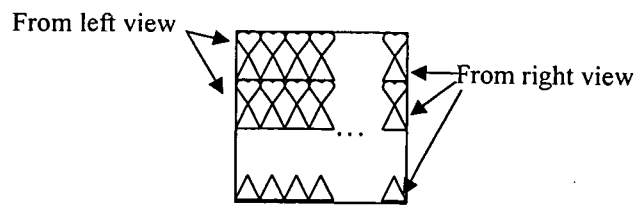
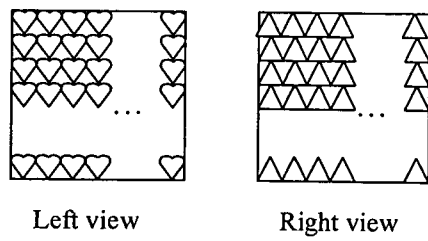


Figure 29

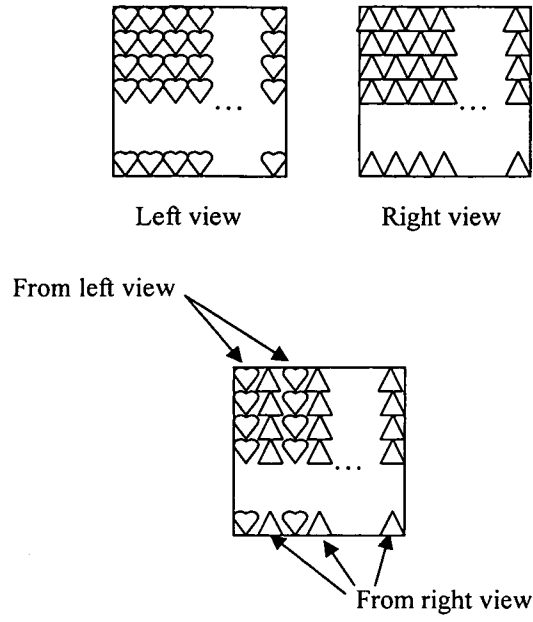


Figure 30

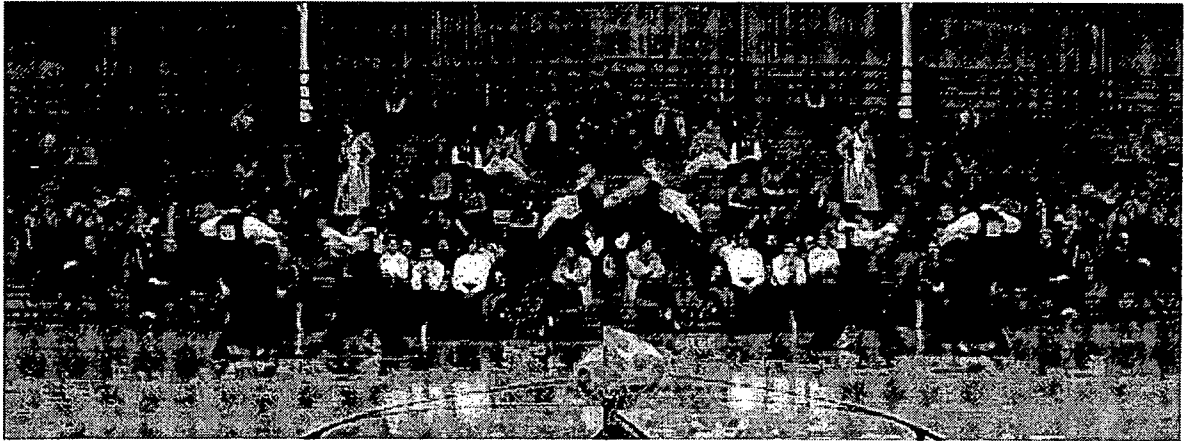


Figure 31

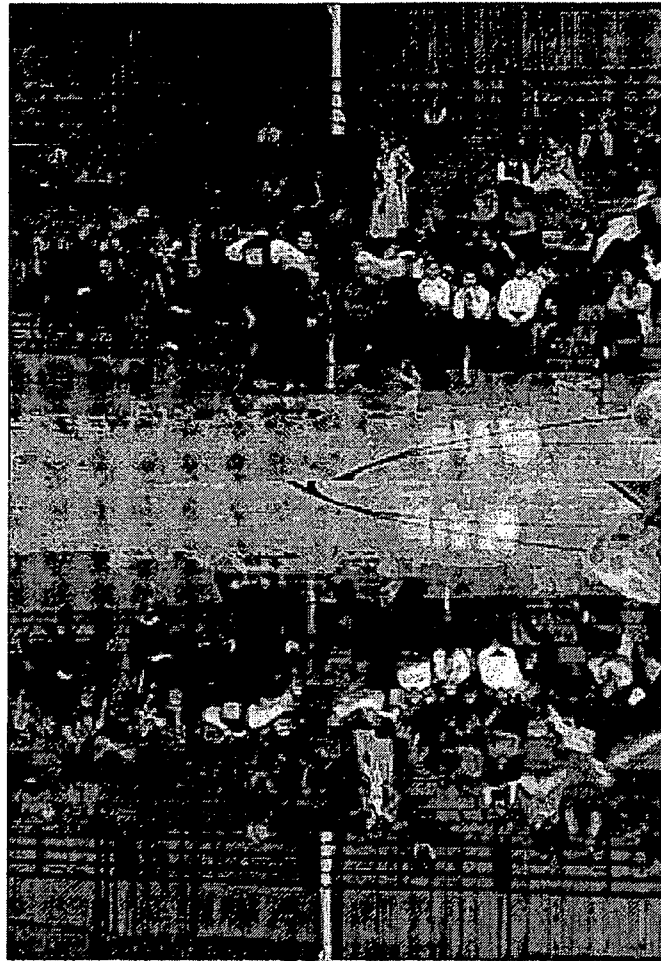


Figure 32



Figure 33

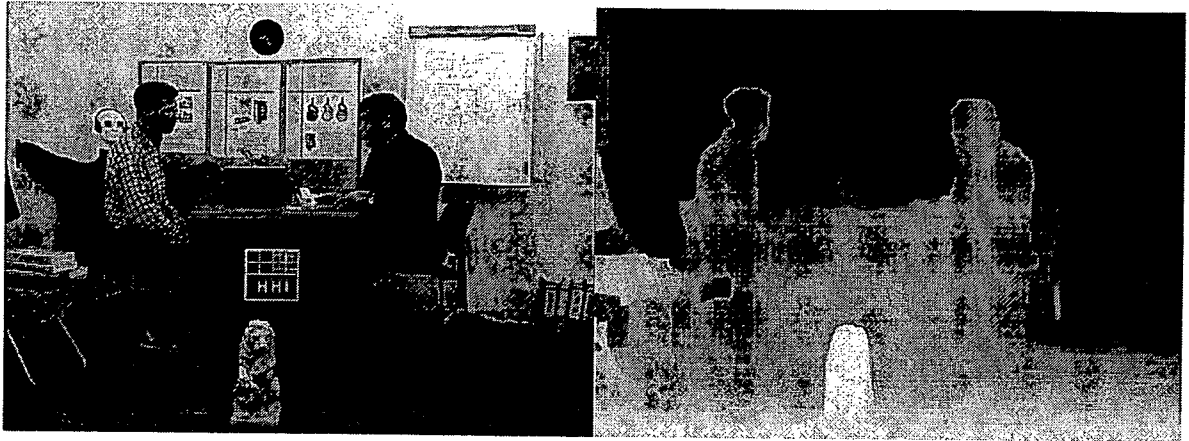


Figure 34

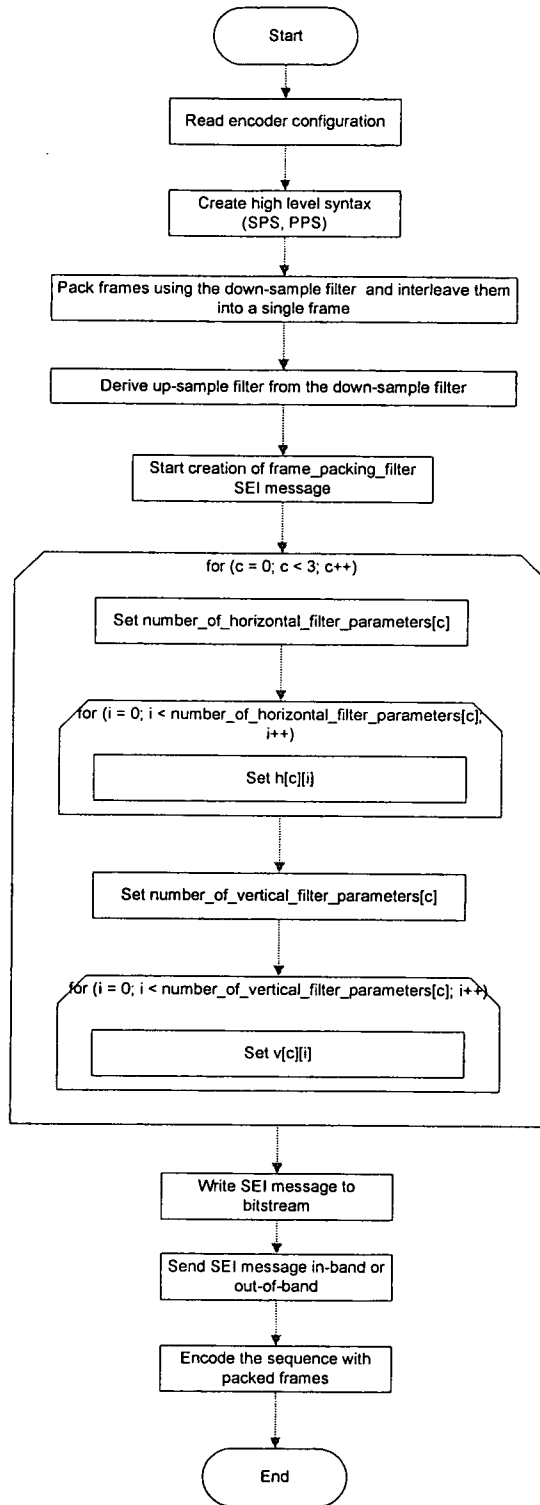


Figure 35. Encoding flowchart (embodiment 1)

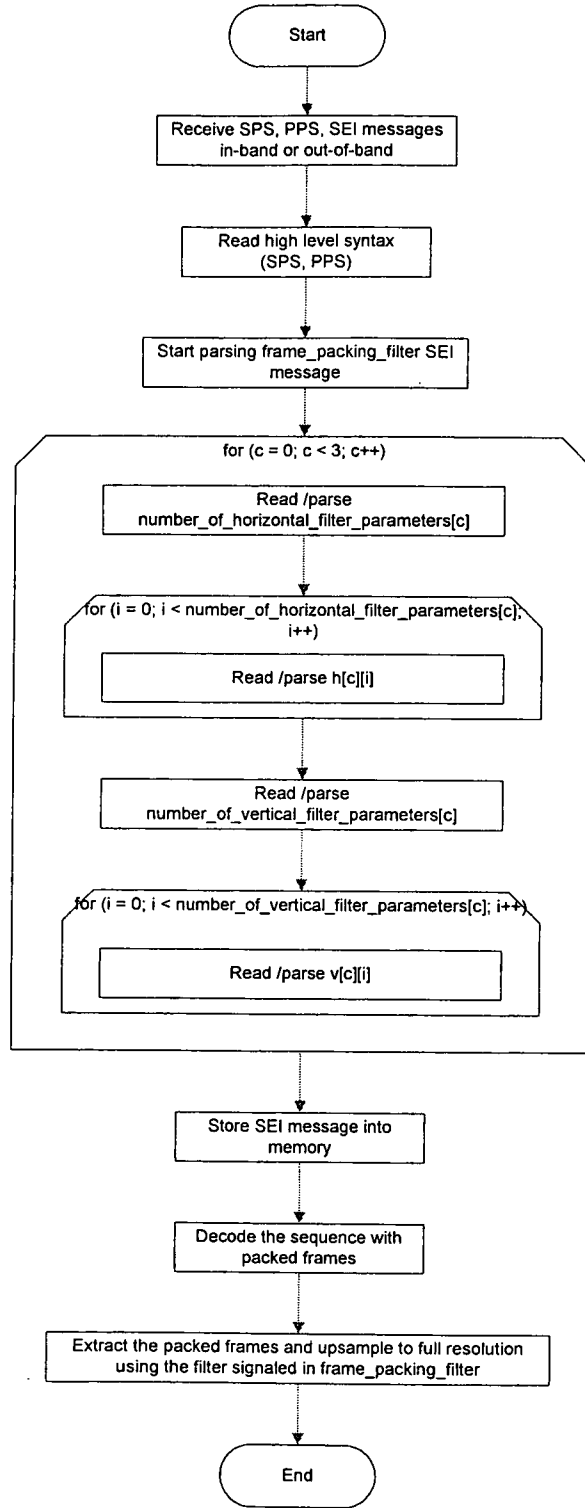


Figure 36. Decoding flowchart (embodiment 1)

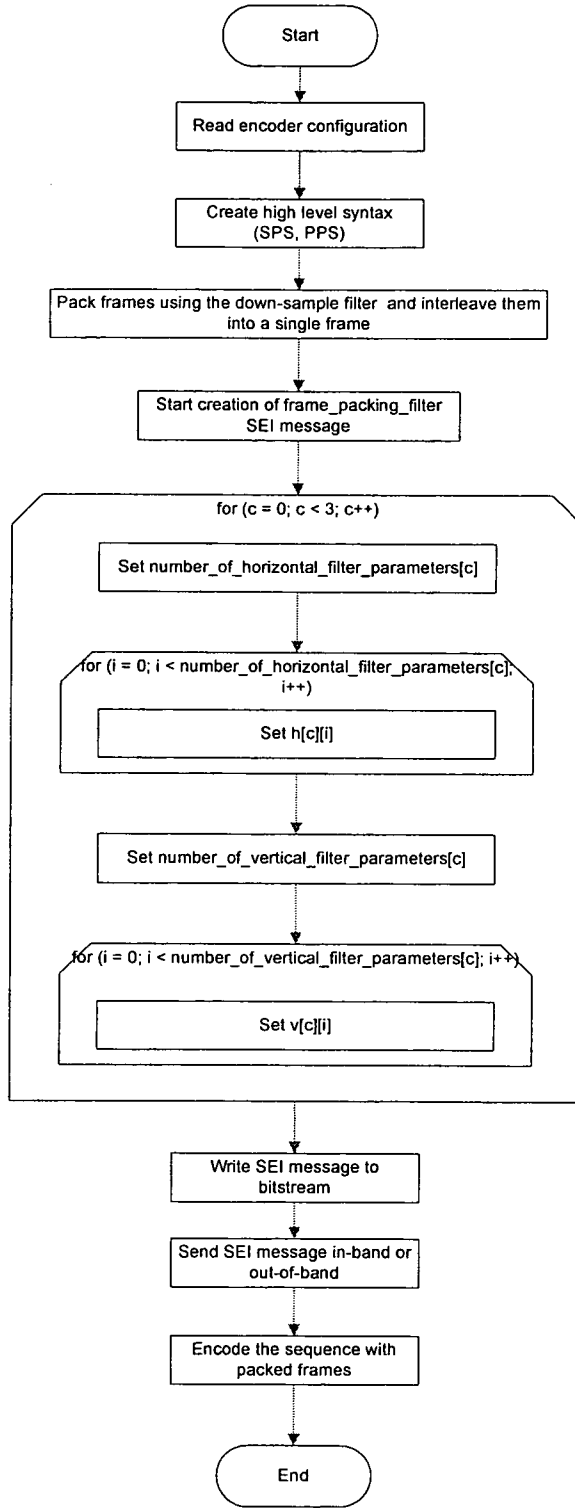


Figure 37. Encoding flowchart (embodiment 2)

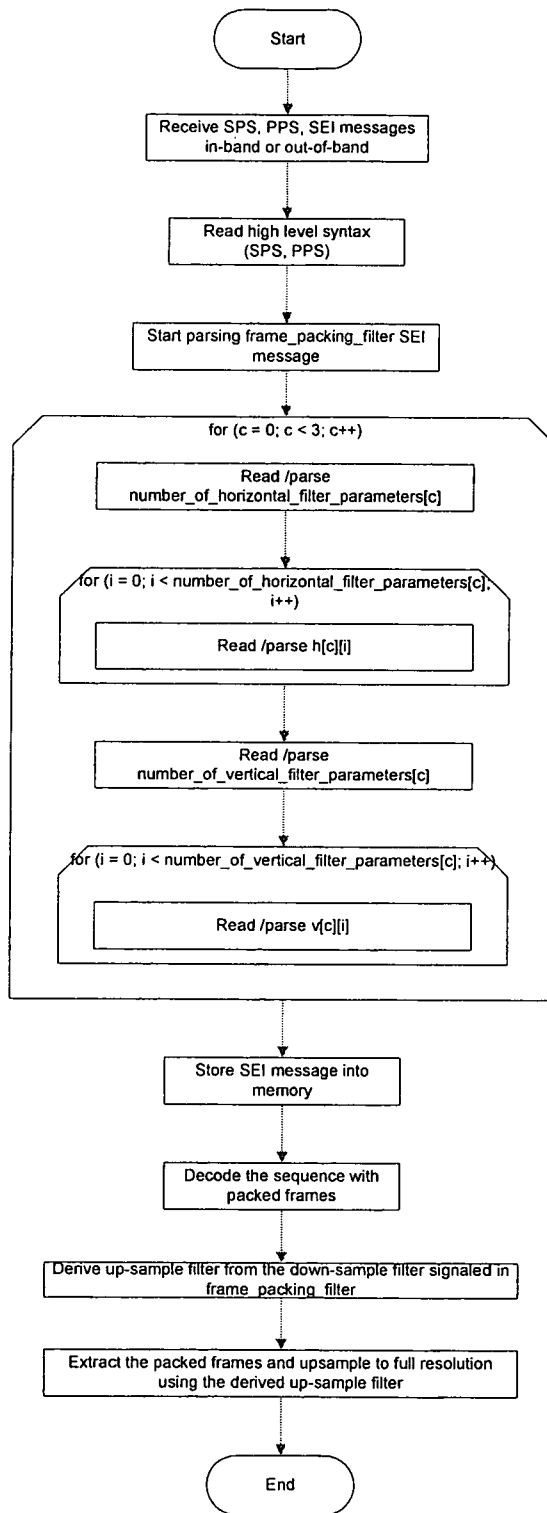


Figure 38. Decoding flowchart (embodiment 2)

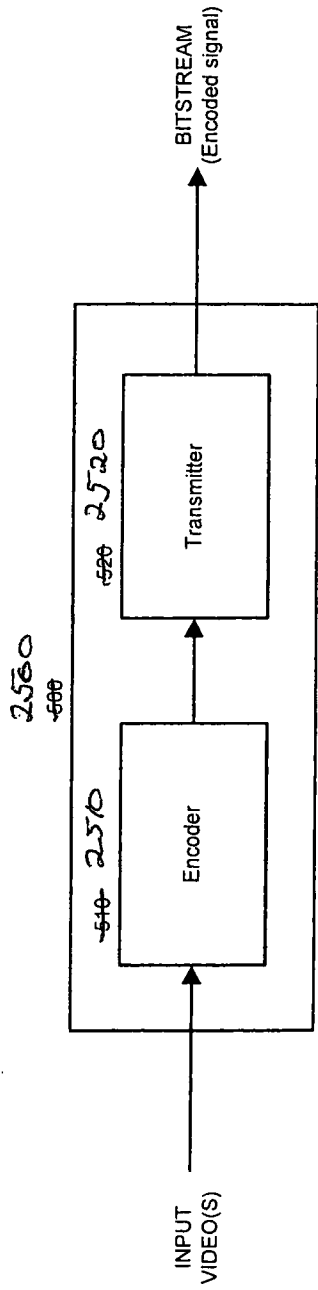


FIG. 37

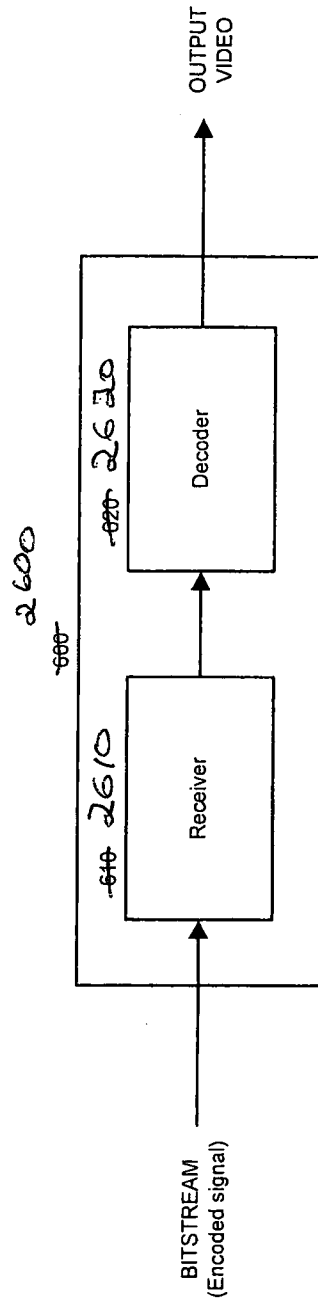


FIG. 40

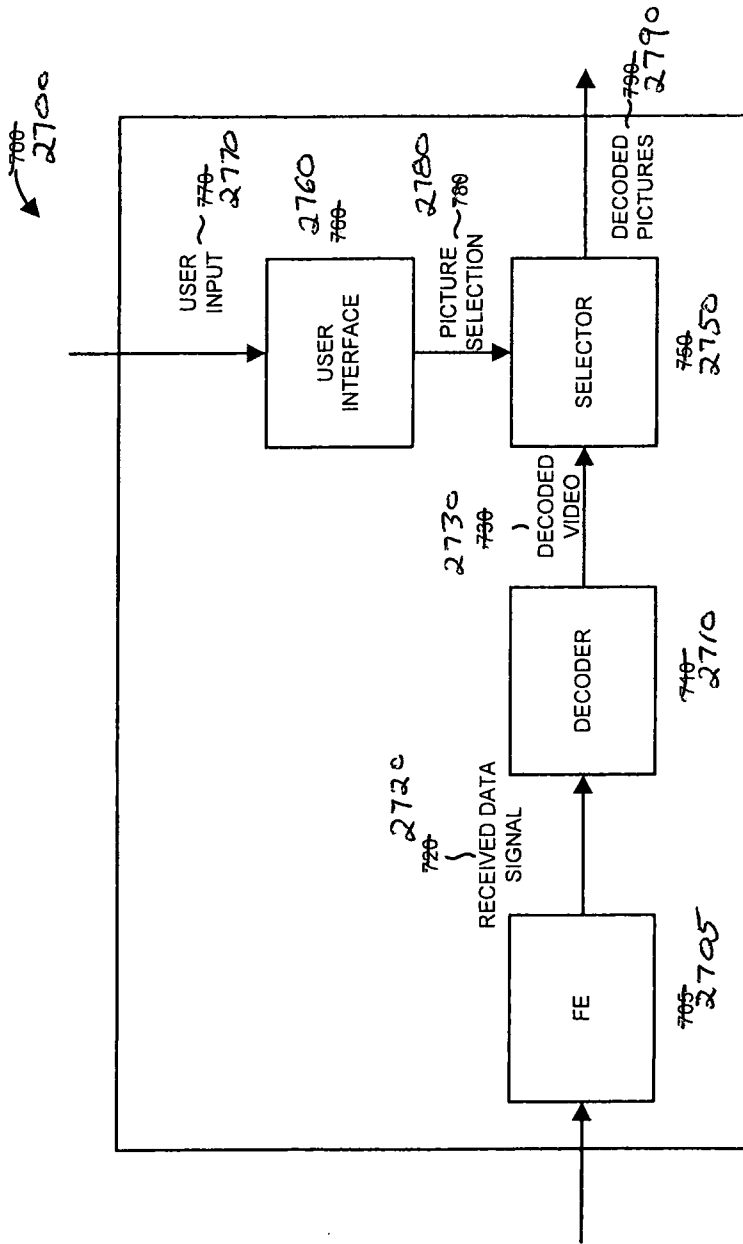


FIG. 41