



(51) International Patent Classification:

H04N 19/30 (2014.01) H04N 19/46 (2014.01)

H04N 19/70 (2014.01)

(21) International Application Number:

PCT/EP2020/065681

(22) International Filing Date:

05 June 2020 (05.06.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

19179158.1 07 June 2019 (07.06.2019) EP

(71) Applicant: **FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.** [DE/DE]; Hansastraße 27c, 80686 München (DE).

(72) Inventors: **SÁNCHEZ DE LA FUENTE, Yago**; c/o Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI, Einsteinufer 37, 10587 Berlin (DE). **SÜHRING, Karsten**; c/o Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI, Einsteinufer 37, 10587 Berlin (DE). **HELLGE, Cornelius**; c/o Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI, Einsteinufer 37, 10587 Berlin (DE). **SCHIERL, Thomas**; c/o Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI, Einsteinufer 37, 10587 Berlin (DE). **SKUPIN, Robert**; c/o Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI, Einsteinufer 37, 10587 Berlin (DE). **WIEGAND, Thomas**; c/o Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI, Einsteinufer 37, 10587 Berlin (DE).

(54) Title: CONCEPTS USING MULTI-LAYERED CODING

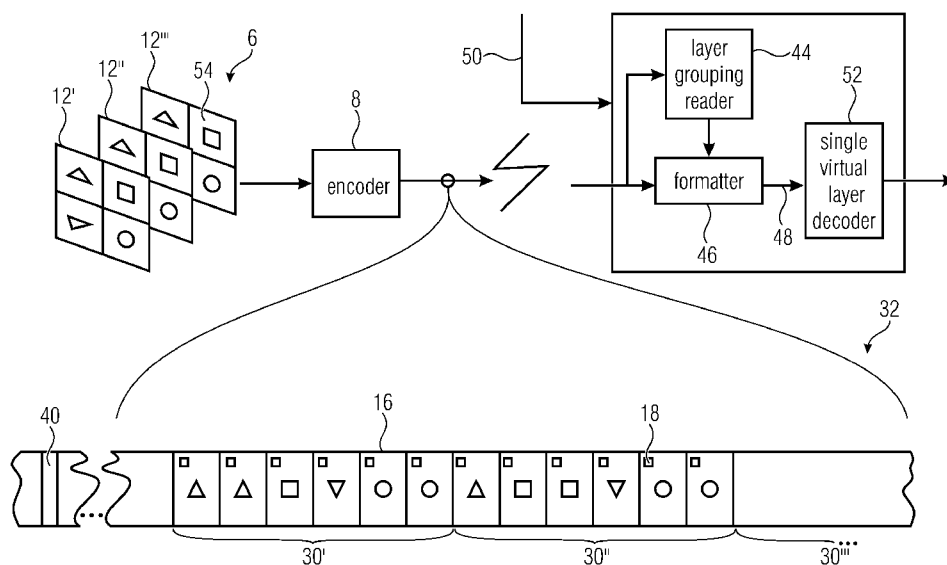


Fig. 1

(57) **Abstract:** Decoder (28) for decoding a multi-layered video data stream (32) which is partitioned into portions (16) each of which comprises a layer indication (18) indicating a layer (70) the respective portion (16) belongs to, configured to read (44), from the multi-layered video data stream (32), a layer grouping information (40) which indicates a grouping of layers (70) which the portions (16) of the multi-layered video data stream (32) belong to, into one or more groups (72) of layers, a virtual layer (42) being associated with each group (72) of layers. Additionally the decoder is configured to form (46) a video data stream (48) out of portions (16) of the multilayered video data stream (32), which belong to the group (72) of layers associated with a predetermined virtual layer (50), by taking over the portions (16) from the multi-layered video data stream (32) into the video data stream (48) with leaving the layer indication (18) of the portions (16) unchanged, and decode (52) the video data stream (48) as a single layer data stream, the group of

(74) **Agent: STÖCKELER, Ferdinand** et al.; Schoppe, Zimmermann, Stöckeler, Zinkler, Schenk & Partner mbB, Radlkofenstr. 2, 81373 München (DE).

(81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

Concepts using Multi-Layered Coding

Description

5

Embodiments according to the invention are related to concepts using multi-layered coding which might be used to render a layered codec such as a multi-layer extension of the merging VVC standard, more effective and/or powerful in various applications and scenarios. From a different perspective, the application is concerned with applications such as 360 video, volumetric video or adaptive streaming, and enables to implement such applications efficiently using multi-layered coding.

10

15

There exist several video applications in which multi-layer video coding is beneficial, e.g., adaptive streaming, video conferencing. Multi-layer extensions have been typically developed for previous video coding standards that specify scalable extensions (e.g., SNR, spatial resolution) or 3D extensions.

20

With the emergence of new applications, such as 360 video streaming, 6DoF -volumetric video, new needs for multi-layer coding have appeared.

25

With this respect, a 360 video could be split into several rectangular regions that are independently encoded, and these could be offered as layers of a same bitstream. Also, for applications such as volumetric video, several components (e.g., depth, occupancy, texture patches) need to be present for the video bitstream. Therefore, a multi-layer extension is the most straight-forward solution where each of the components is stored as a different and independent layer in the bitstream. The described examples constitute applications for which the encoding of the layers is carried out independently. I.e., contrary to typical use-cases considered for previous video coding standards, there is no inter-layer dependency among layers, but the bitstream with several layers is merely for multiplexing purposes. A further example thereof is adaptive streaming where several single-layer encoded versions of the same content are available so that one of them, with either different bitrate, different resolution, different bit-depth... is selected that matches the device or network capabilities. In such a scenario, the different single-layer encoded versions can be multiplexed/"encapsulated" into a single bitstream with different layers.

30

35

Contrary to the examples mentioned above, scalability can exploit inter-layer dependency, as has been done in the past in several video coding standards such as SVC or SHVC. SVC was

a single-loop decoding process where only the highest layer was decoded and reconstructed (with some exceptions when Medium Grain Scalability -MGS – was used). SHVC was a multi-loop decoding process where all picture of all layers where decoded and reconstructed. SVC allowed some kind of “error resilient” decoding where if the highest layer for a picture was not
5 decoded still a picture was reconstructed and could be used as reference for further picture with some kind of drift. In SHVC such a feature was removed and only complete layers could be decoded where the reference pictures at the same layer were available. A multi-loop decoding process has the benefits that the decoding process is essentially similar to single-layer and is easy to implement. Still, it misses some capabilities for “error resiliency” that could
10 be beneficial. Furthermore, the multi-loop approach to scalability has typically a lower coding efficiency than the single-loop approach.

Pros and Cons

1) Multi-layer for 360 video regions

The benefit of using the multi-layer extension for the regions in which a 360 video is split is that no rectangular form needs to be achieved when decoding the regions necessary for displaying the content as every region could potentially be output independently.

20 On the contrary, when the MCTS encoded variant is used in which several regions are stitched together into a single bitstream, the stitched bitstream needs to be rectangular so that a regular decoder can handle the decoding of the video bitstream. This, can lead to requiring the decoder to decode more samples than the ones required if the layered version was used.

25 On the other side, the layered version, where in principle several single layer decoders could be used to decode each of the selected layers, would require to apply frame-synchronization at the output which could become a cumbersome operation and could be impossible to be handled by some implementations. The MCTS approach with bitstream merging/stitching outputs the pictures already synchronized, which is an advantage over the layered case.

2) Volumetric-Video/6DoF

Volumetric video is gaining a lot of attention in the last years. Currently a standard is being developed for point cloud, called Point Cloud Coding (PCC). Typically, there are several components that need to be encoded for such an application such as texture, depth and
35 patches. Having all of the components within a single bitstream is beneficial as, for instance, it provides directly synchronization among different component, it simplifies relationship of potential metadata that might refer to more than one component simultaneously.

However, such components could have different framerates and might not be easy to integrate following multi-layer coding principles as used in previous standards. Previous standards are based on Access Units that contain one or more pictures of several layers. All pictures within
5 a single AU contain pictures that have the same Picture Order Count (POC) value. Different framerates might be present at different layers but usually, the different framerates are multiples, e.g., a base layer having 30Hz and an enhancement layer having 60Hz. In such an example, every second AU contains pictures of both layers and the rest only pictures of the enhancement layer.

10 However, there are use cases for volumetric video with several components that have different framerates that are not multiple of each other, which makes it difficult to fit pictures into AUs as done previously.

15 3) Adaptive Streaming

Adaptive streaming is typically carried out with single-layer video coding where several versions of the same content are offered and fractions of them are decoded by a single decoder based on the current network characteristics. If they change, a different version is selected
20 and a fraction of the new matching bitstream is sent to the decoder. Typically, switching happens at so-called Random Access Points (RAPs).

A similar procedure can be also carried out with layers where additional layers are downloaded to a base layer (lowest quality) if the network characteristics allow for it. These additional layers
25 are typically encoded using inter-layer prediction.

An approach in between referred to Adaptive Resolution Coding (ARC) or Reference Picture Resampling (RPR) consist of allowing having different resolution at every picture. Thus, when the resolution changes, a resampling process is carried out when temporal prediction is
30 required to match the resolution of the current picture. This changes the fact that a single-layer bitstream has a constant resolution, but does not require more than a simple re-sampling process when the resolution changes.

At least two valuable use-cases can be envisioned for such a technique:

- 35 - Point to point communication: There is a single bitstream generated and the resolution of the bitstream can change at every picture as required
 - o Single layer

- Point to multi-point: several single-layer versions are offered and switching is allowed at concrete pictures.
 - o Multi layer

5 Therefore, it is desired to provide concepts for rendering multi-layer coding more efficient to support various applications such as 360 video, volumetric video and/or adaptive streaming.

This is achieved by the subject matter of the independent claims of the present application.

10 Further embodiments according to the invention are defined by the subject matter of the dependent claims of the present application.

Summary of the Invention

A first aspect of the present invention is based on the idea that layers of a multi-layered data stream subject to grouping so as to result into one or more layer groups each which is treated as one virtual layer. The grouping and the association to virtual layer(s) is coded in the multi-layered data stream as layer grouping information. A decoder is, thus, able to form a video data stream out of portions of the multi-layered video data stream, which belong to the group of layers associated with a predetermined – namely wanted or targeted - virtual layer, by taking over those portions from the multi-layered video data stream which belong to that group into the video data stream with leaving the layer indication of the portions unchanged. The resulting video data stream may be decoded as a single layer data stream, quasi a data stream of only the virtual layer, wherein the group of layers are coded independent from each other and relate to mutually different picture regions of a video represented by the single layer data stream.

25 Remarkably, several virtual layers may be defined and their layer groups may overlap. Although the portions for the reconstruction of the output pictures are still associated with different layers, namely the layers belonging to the predetermined group of layers correspond to the predetermined virtual layer, the decoder can disregard the different layer association as the decoder expects the indicated portions to belong to the one predetermined virtual layer only and to, thus, form a single-layer-like data stream.

30

The group of layers associated with a virtual layer can comprise one, some or all layers present in the multi-layered video data stream. The layers of the multi-layered video data stream can be grouped into one or more groups of layers with each group of layers being associated with a different virtual layer. Portions with a different layer indication can belong to the same virtual layer according to the layer grouping information. Different virtual layers can lead to output pictures with different details, i.e. picture regions, of input pictures. In other words, output

35

pictures can have a different size than corresponding input pictures, depending on a predetermined virtual layer chosen for the respective pictures of a video. In even other words, the output pictures' picture size may be different for different virtual layers. The predetermined virtual layer defines, for example, one virtual layer out of two or more virtual layers, signaled in the multi-layered video data stream by way of the layer grouping information. For example, the predetermined virtual layer can be chosen based on a positioning or view direction of a viewer of the video, since the predetermined virtual layer may group layers having picture regions of input pictures encoded therein, which are relevant for this view. The predetermined virtual layer may change from one picture to another or from one group of pictures to another. Portions of a multi-layered video data stream comprising the same layer indication are, for example, predicted, encoded and/or decoded independent from other portions of the multi-layered video data stream comprising, or relating to, other layer indications. Only portions related to the predetermined virtual layer are, for example, decoded. Picture regions of interest can be defined with the layers of the group of layers associated with the predetermined virtual layer.

Accordingly, in accordance with a first aspect of the present application, a decoder for decoding a multi-layered video data stream which is partitioned into portions each of which comprises a layer indication indicating a layer the respective portion belongs to, is configured to read, from the multi-layered video data stream, a layer grouping information which indicates a grouping of layers which the portions of the multi-layered video data stream belong to, into one or more groups of layers, a virtual layer being associated with each group of layers. The portions are, for example, associated with picture regions into which pictures of the video are partitioned into. Portions associated with the same picture region, e.g., associated with different slices of the same picture region, i.e., associated with different prediction/encoding cycles of one picture region, or portions associated with picture regions out of a collection of mutually co-located picture regions of all pictures comprise, for example, the same layer indication. A group of layers comprises, for example, at least one layer or two or more layer. Thus portions, e.g., picture regions, comprising different layer indications can belong to the same virtual layer. Additionally, the decoder is configured to form a video data stream out of portions of the multi-layered video data stream, which belong to the group of layers associated with a predetermined virtual layer, by taking over the portions from the multi-layered video data stream into the video data stream with leaving the layer indication of the portions unchanged. Thus, for example, only portions related to the virtual layer are incorporated into the video data stream to be decoded by the decoder. By leaving the layer indication of each portion taken over into the video data stream unchanged, the decoder is, for example, configured to read, from the video data stream, parameter sets associated with each layer of the group of layers.

Furthermore, the decoder is configured to decode the video data stream as a single layer data stream, the group of layers being coded independent from each other and relating to mutually different picture regions of a video represented by the single layer data stream. In other words, each layer of the group of layers is decoded individually. The decoder is, for example, configured to decode portions associated with a first layer indication independent of portions associated with a second layer indication. Coding dependencies merely exist, for example, between portions belonging to the same layer indication. According to a different interpretation, the formation and decoding of the video data stream can be understood such that the decoder is configured to only read information from the multi-layered video data stream and decode portions related to the layers of the group of layers associated with the predetermined virtual layer and ignoring the rest of the multi-layered video data stream.

Accordingly, in accordance with a first aspect of the present application, an encoder for encoding a video, comprises independently encoding mutually different picture regions of the video into video data stream portions and embedding the video data stream portions into a multi-layered video data stream by providing each video data stream portion with a layer indication indicating a layer the respective video data stream portion belongs to, so that video data stream portions having one picture region of the different picture regions encoded therein belong to a different layer than video data stream portions having another picture region of the different picture regions encoded therein. Additionally, the encoder comprises providing the multi-layered video data stream with a layer grouping information which indicates a grouping of the layers which the video data stream portions belong to, into one or more groups of layers, a virtual layer being associated with each group of layers.

Accordingly, in accordance with a first aspect of the present application, a multi-layered video data stream comprises video data stream portions having mutually different picture regions of the video independently encoded therein, wherein each video data stream portion comprises a layer indication indicating a layer the respective video data stream portion belongs to, wherein video data stream portions having one picture region of the different picture regions encoded therein belong to a different layer than video data stream portions having another picture region of the different picture regions encoded therein. Additionally, the multi-layered video data stream comprises a layer grouping information which indicates a grouping of the layers which the video data stream portions belong to, into one or more groups of layers, a virtual layer being associated with each group of layers.

A second aspect of the present invention is based on the idea that the layer coded into the multi-layered data stream, or the layer up-to which, in a hierarchical sense, layers of the multi-

layered data stream are coded into, or are present in, the multi-layered data stream, varies in time such as per picture or at a picture basis, or per picture sequence or the like. The variation may be caused by adaptive streaming approaches, may be made in an on demand encoding sense, may be made adaptively to varying transmission bandwidth, or using some other approach. The hierarchical nature or mutual dependency among the layers between which the variation or switching takes place may be optional. In case of layer dependency, the highest layer available for a picture if the layer selected for that picture, and any layer which this predetermined layer depends on, are also gathered at the decoder. In any case, the layers enable a derivation of a reference picture of one layer based on a temporally aligned, i.e. co-temporal, picture of another layer. That is, the layers relate to the same picture content at, for example, different quality. Thus, the layer coded into the data stream, or up to which the layers are coded into the data stream, can be switched and at such a switching point a picture associated with a first layer can be referenced by a picture associated with a second layer, for sake of a motion-compensated prediction of the picture associated with the second layer. The reference picture using which the picture associated with the second layer has actually been encoded at the encoder side, may then be estimated or approximated by use of the picture associated with the first layer which is co-temporal with respect to the missing the reference picture.

In accordance with an embodiment of the second aspect, the group of layers is independently coded, ought to select out of this layer group exactly one layer per picture. In that case, the resulting data stream, though relating to several layers, namely the group of layers, is similar to a single-layer data stream in the sense of having only one layer per picture. Drift may occur at the decoder in case of having encoded the layers independently among the layers, with encoding each layer's pictures solely based on reference pictures of that layer, i.e. if performing the encoding agnostic with respect to the switching among the layers by selecting for each picture only one layer of the layers subsequently such as by adaptive streaming or the like. In case of changing the layer on the fly during encoding, and performing the reference picture derivation of missing layer's reference pictures also at the side of the encoder, this drift does not even occur. Even if drift occurs, measures may be taken to keep these coding drifts reasonable low. Thus, a high efficiency in an adaptation of a layer variation over time can be achieved by gathering portions of a multi-layered video data stream associated with, for each picture, one layer of a group of layers. The group of layers may, for example, comprise layers leading to no drift or only a small drift at a switching among these layers. Layers of the group of layers, for example, differ with respect to quality of encoded pictures of a video. Each layer, for example, may be encoded independently. In case of independently coded layers, for each picture of the video, only portions belonging to one layer, i.e. a predetermined layer, out of the

group of layers might be gathered and decoded. The predetermined layer can change over time, e.g. at some switching points. Again, compared thereto, if the layer were coded in a scalable manner a number of layers may be encoded into a multi-layered video data stream per picture, wherein the decoder can be configured to choose as the predetermined layer
5 always the highest layer present for a picture. In case of mutually independent layers, the selection of the one predetermined layer per picture, might be made by some entity other than encoder and other than decoder such as by an adaptive streaming client, with the decoder "realizing" the predetermined layer simply by collecting for each picture all portions such as NAL units, which belong to any of the group of layers, thereby realizing that for each picture,
10 the portions having this picture encoded thereinto, only relate to one layer, i.e. the predetermined layer. Thus, the decoder realizes for each picture as to which layer is the predetermined layer for the respective picture and may, thus, determine whenever a reference picture of a picture has a different predetermined layer than the picture referencing that reference picture. When this occurs, the actually missing reference picture, namely the version
15 thereof which is of the same layer as the predetermined layer of the referencing picture, i.e. the picture indicating the reference picture to be used for a prediction of the picture, is derived based on the actually coded version of that reference picture, namely the picture of the coded predetermined layer, which is co-temporal to the missing reference picture. If the predetermined coded layer of the referencing picture corresponds to a finer resolution than the
20 predetermined layer which has been selected for the reference picture, and at which the reference picture is coded into the data stream, up-sampling is used to form a substitute for the actual reference picture version which would be of the same layer than the predetermined layer of the referencing picture, and if the predetermined coded layer of the referencing picture corresponds to a lower resolution than the predetermined layer which has been selected for
25 the reference picture, and at which the reference picture is coded into the data stream, down-sampling may be used to form a substitute for the actual reference picture version which would be of the same layer than the predetermined layer of the referencing picture. The substitute is then used for the motion compensated prediction of the referencing picture.

30 Accordingly, in accordance with a second aspect of the present application, a decoder for decoding a multi-layered video data stream which is partitioned into portions each of which has an associated picture of a video, or a part thereof, encoded thereinto and comprises a layer indication indicating a layer the respective portion belongs to, is configured to read, from the multi-layered video data stream, a layer grouping information which groups the layers the
35 portions belong to, into which the video is coded, into a group of layers and gather portions of the multi-layered video data stream, which belong to the group of layers, so that each picture of the video is associated with one layer of the group of layers, and for each picture of the

video, the one or more portions belonging to the layer associated with the respective picture are gathered. The gathering may, as outlined above, merely include the decoder's collection of all stream portions belonging to any of the layer group. In case of the independently coded layers, the decoder may then realize, however, that the data stream merely comprises portions of one predetermined layer out of the layer group for each picture, with this one predetermined picture varying from one picture to another. The selection has been done before by another entity as described above. The decoder realizes as to which predetermined layer has been chosen for each picture based on the gathering result. The decoder might deduce from the layer grouping information that the data stream merely comprises portions of one predetermined layer for each picture and might even regard the data stream to be erroneous if there were portions of two of layers of the layer group. Alternatively, the decoder always appoints the highest layer among the layers any of the portions of a certain picture in the data stream belongs to, as the predetermined layer and gathers only those portions. In case of dependent layers, the decoder may accompany these portions with the portions of any reference layer. Additionally, the decoder is configured to decode the video from the gathered portions by use of motion-compensated prediction with supporting that pictures decoded from portions associated with a first layer of the group of layers being referenced, for the motion-compensated prediction, by portions associated with a second layer of the group of layers, different from the first layer. In other words, the portions associated with the second layer refer to the one or more pictures decoded from the portions associated with the first layer, such that the first layer pictures are used for motion-compensated prediction of second layer pictures. The supporting may involve the decoder estimating/reconstructing such portions based on inter-layer prediction.

According to another interpretation, the decoder for decoding a multi-layered video data stream which is partitioned into portions each of which has an associated picture of a video, or a part thereof, encoded therein and comprises a layer indication indicating a layer the respective portion belongs to, wherein each picture of the video is associated with one layer of the group of layers, and for each picture of the video, the one or more portions belonging to the layer associated with the respective picture are gathered, is configured to read, from the multi-layered video data stream, a layer grouping information which groups the layers the portions belong to, into which the video is coded, into a group of layers; and decode the video from the gathered portions by use of motion-compensated prediction with supporting that pictures decoded from portions associated with a first layer of the group of layers are referenced, for the motion-compensated prediction, by portions associated with a second layer of the group of layers, different from the first layer.

Accordingly, in accordance with a second aspect of the present application, an apparatus for forming a multi-layered video data stream having a video encoded thereinto, is configured to encode, using motion-compensated prediction, pictures of the video into portions of the multi-layered video data stream so that each portion has an associated picture of the video, or a portion thereof, encoded thereinto and comprises a layer indication indicating a layer the respective portion belongs to, and so that the portions into which the video is encoded belong to different layers, wherein the apparatus is configured so that the multi-layered video data stream comprises portions associated with a predetermined layer of the group of layers, which reference, for the motion-compensated prediction, pictures for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream. Additionally, the apparatus is configured to insert, into the multi-layered video data stream, a layer grouping information which groups the layers the portions belong to, into which the video is encoded, into a group of layers.

Accordingly, in accordance with a second aspect of the present application, a Multi-layered video data stream having a video encoded thereinto is provided. Portions of the multi-layered video data stream have, using motion-compensated prediction, pictures of the video encoded thereinto so that each portion has an associated picture of the video, or a portion thereof, encoded thereinto and comprises a layer indication indicating a layer the respective portion belongs to, and so that the portions into which the video is encoded belong to different layers, wherein the multi-layered video data stream comprises portions associated with a predetermined layer of the group of layers, which reference, for the motion-compensated prediction, pictures for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream. Additionally, the multi-layered video data stream comprises a layer grouping information which groups the layers the portions belong to, into which the video is encoded, into a group of layers.

Each picture of a video can be encoded into one or more portions of a multi-layered video data stream in different qualities, i.e. versions, wherein each quality is assigned a different layer out of a group of layers. A picture of a certain quality can be reconstructed based on a reference picture. A third aspect of the present invention is based on the idea, that the reference picture does not necessarily have to belong to the same layer as a predetermined picture which one or more portions, i.e. one or more predetermined portions, are predicted from the reference picture. An advantageous indication for the actually used layer associated with the reference picture can be indicated in the data stream or read from the data stream. It is possible, that the layer associated with the actually used reference picture is encoded for each picture of the video but that portions of the reference picture belonging to this layer are not already decoded

by a decoder. For this case, a subset of layers can be signaled for the predetermined picture in the multi-layered video data stream, wherein the subset indicates layers out of the group of layers representing an allowed or preferred substitute of the actually used version of the reference picture. If the subset is signaled and the version of the reference picture associated with the predetermined layer is not present, a decoder can be configured to decode the predetermined picture from a version of the reference picture associated with one of the layers out of the subset. Alternatively, a layer ranking can be signaled for the predetermined picture in the multi-layered video data stream, wherein the layer ranking, for example, indicates preferences for two or more layer out of the group of layers for the reference picture. Different preferences are assigned to different layers. The decoder tries to use a version of the reference picture associated with a most preferred layer (based on the layer ranking) first, after realizing, that the version of the reference picture associated with the predetermined layer is not possible or present. If the usage of the version of the reference picture associated with the most preferred layer also not possible, the next layer according to the ranking is checked by the decoder and so forth. Alternatively, an allowance or disallowance can be signaled for each layer of the group of layers as a possible layer associated with the reference picture for the motion-compensated prediction of a picture of the video.

Accordingly, in accordance with a third aspect of the present application, a decoder for decoding a multi-layered video data stream having a video encoded thereinto, is configured to decode, using motion-compensated prediction, pictures of the video from portions of the multi-layered video data stream. Each portion has an associated picture of the video, or a part thereof, encoded thereinto and comprises a layer indication indicating a layer the respective portion belongs to, and each picture is, for each of at least one out of a group of layers, encoded into one or more portions, which belong to the respective layer, so that a version of the respective picture, which is associated with the respective layer, is reconstructible from the one or more portions, which belong to the respective layer. Additionally, the decoder is configured to read, from the multi-layered video data stream, a reference picture information which indicates for one or more predetermined portions which have, using motion-compensated prediction from a reference picture, a predetermined picture encoded thereinto, and which belong to a predetermined layer, one or more of

the layer, e.g., a reference layer, with which an actually used version of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions using motion-compensated prediction; and

a subset of layers out of the group of layers which includes, or excludes, the layer, e.g., the reference layer, with which the actually used version of the reference picture is

associated, wherein a version of any layer of the subset, other than the layer with which the actually used version of the reference picture is associated, represents an allowed or preferred substitute of the actually used version of the reference picture in decoding the one or more predetermined portions; and

5 a layer ranking indicating a preference ranking among the layers of the group of layers for using the versions of the reference picture associated with the layers for decoding the one or more predetermined portions using motion-compensated prediction, whether any other layer's version of the reference picture other than the predetermined layer is allowed to be used in decoding the one or more predetermined portions.

10 Additionally, the decoder is configured to decode the predetermined picture at the predetermined layer from the multi-layered video data stream using motion compensated prediction from a version of the reference picture selected depending on the reference picture information.

15 Accordingly, in accordance with a third aspect of the present application, an apparatus for forming a multi-layered video data stream having a video encoded therein, is configured to encode, using motion-compensated prediction, pictures of a video into portions of the multi-layered video data stream so that each portion has an associated picture of the video, or a part thereof, encoded therein and comprises a layer indication indicating a layer the
20 respective portion belongs to, and so that each picture is, for each of at least one out of a group of layers, encoded into one or more portions, which belong to the respective layer, so that a version of the respective picture, which is associated with the respective layer, is reconstructible from the one or more portions, which belong to the respective layer. Additionally, the apparatus is configured to insert, into the multi-layered video data stream, a
25 reference picture information which indicates for one or more predetermined portions which have, using motion-compensated prediction from a reference picture, a predetermined picture encoded therein, and which belong to a predetermined layer, one or more of

the layer, e.g., a reference layer, with which an actually used version of the reference picture is associated, from a reconstruction of which the predetermined picture is
30 encoded into the one or more predetermined portions using motion-compensated prediction; and

a subset of layers out of the group of layers which includes, or excludes, the layer, e.g., the reference layer, with which the actually used version of the reference picture is associated, wherein a version of any layer of the subset, other than the layer with which
35 the actually used version of the reference picture is associated, represents an allowed or preferred substitute of the actually used version of the reference picture in decoding the one or more predetermined portions; and

a layer ranking indicating a preference ranking among the layers of the group of layers for using the versions of the reference picture associated with the layers for decoding the one or more predetermined portions using motion-compensated prediction, whether any other layer than the predetermined layer is allowed to be used in decoding the one or more predetermined portions.

Accordingly, in accordance with a third aspect of the present application, a multi-layered video data stream having a video encoded therein is provided, wherein, using motion-compensated prediction, pictures of a video are encoded into portions of the multi-layered video data stream so that each portion has an associated picture of the video, or a part thereof, encoded therein and comprises a layer indication indicating a layer the respective portion belongs to, and so that each picture is, for each of at least one out of a group of layers, encoded into one or more portions, which belong to the respective layer, so that a version of the respective picture, which is associated with the respective layer, is reconstructible from the one or more portions, which belong to the respective layer. Additionally, the multi-layered video data stream comprises a reference picture information which indicates for one or more predetermined portions which have, using motion-compensated prediction from a reference picture, a predetermined picture encoded therein, and which belong to a predetermined layer, one or more of

the layer, e.g., a reference layer, with which an actually used version of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions using motion-compensated prediction; and

a subset of layers out of the group of layers which includes, or excludes, the layer, e.g., the reference layer, with which the actually used version of the reference picture is associated, wherein a version of any layer of the subset, other than the layer with which the actually used version of the reference picture is associated, represents an allowed or preferred substitute of the actually used version of the reference picture in decoding the one or more predetermined portions; and

a layer ranking indicating a preference ranking among the layers of the group of layers for using the versions of the reference picture associated with the layers for decoding the one or more predetermined portions using motion-compensated prediction, whether any other layer than the predetermined layer is allowed to be used in decoding the one or more predetermined portions.

Brief Description of the Drawings

The drawings are not necessarily to scale, emphasis instead generally being placed upon illustrating the principles of the invention. In the following description, various embodiments of the invention are described with reference to the following drawings, in which:

Fig. 1 shows a schematic view of an encoder and a decoder and a multi-layer video data stream comprising portions comprising a layer indication and associated with picture regions, according to an embodiment of the present invention;

Fig. 2 shows a table illustrating a layer indication information, according to an embodiment of the present invention;

Fig. 3 shows a schematic view of an encoder and a decoder and a multi-layer video data stream comprising one or more portions for each quality, for each picture of a video, according to an embodiment of the present invention;

Fig. 4 shows a schematic view of an initial data stream and examples of a "stripped" data stream associated with intra-layer prediction, according to an embodiment of the present invention;

Fig. 5 shows a schematic view of an initial data stream and examples of a "stripped" data stream associated with inter-layer prediction, according to an embodiment of the present invention;

Fig. 6a shows a table illustrating a picture region arrangement information, according to an embodiment of the present invention;

Fig. 6b shows a schematic view of a formed video data stream out of a multi-layered video data stream and illustrates a relative position information of picture parts and a picture region arrangement information, according to an embodiment of the present invention; and

Fig. 7 shows a schematic view of a video with different temporal layer encoded at different bitrates.

Detailed Description of the Embodiments

Equal or equivalent elements or elements with equal or equivalent functionality are denoted in the following description by equal or equivalent reference numerals even if occurring in different figures.

In the following description, a plurality of details is set forth to provide a more throughout explanation of embodiments of the present invention. However, it will be apparent to those skilled in the art that embodiments of the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form rather than in detail in order to avoid obscuring embodiments of the present invention. In addition, features of the different embodiments described herein after may be combined with each other, unless specifically noted otherwise.

1. Multi-layer for 360 video regions

For devices that are able to synchronize all output regions (i.e. they have an API or hardware support), the multi-layer approach with parallel decoders is used. As aforementioned, separate encoding and decoding from different layers related to different regions allows for a more "flexible" encoding so that no constraints or specific configuration equal for all layers is required. For instance, GOP sizes might be different for different layers, no motion vector constraints are required as for each region a picture boundary extension is performed, CTU sizes might be different for each layer, etc. This requires that devices implement a output pictures synchronization module that puts all output pictures together properly.

However, many devices out there miss such an API or hardware support and are not capable of synchronizing several videos together when the number of videos becomes higher, e.g. > 20 regions. For such cases, typically, the several different bitstreams (or in the example described layers) are merged together before decoding, i.e. a rewriting operation is needed, so that a single-decoder can be used and a single output picture is output with all regions inside. As discussed, for such a case, a layered merging process is required following, for instance, the indication of how to merge the bitstreams in [EP18194348.1, HHI - 2018P61504EP, FH180906PEP]. The problem is that, besides rewriting the parameter sets, each single NAL unit header would be needed to be rewritten changing the values of layer_id to a single common value of layer_id, which is a burden.

- The first invention consists of defining a VirtualLayerId to which a given set of layer_ids are mapped. All layer_ids that belong to the VirtualLayerId are treated in conjunction as a single layer. The invention consists of signaling the layers that can be interpreted as a single virtual layer, which means that the content has been prepared in such a way that the several layers interpreted as single layer is a single layer conforming bitstream. E.g., the CU addresses of the several NAL units of different layer_ids lead to a conforming bitstream when treated as a single VirtualLayerId, CTU sizes are the same, POC across different layers are aligned, the joint decoding leads to a rectangular picture, the chroma format is the same, etc.
- 10 The bitstream generation at the encoder side and bitstream signaling takes care of making sure that all constraints are in place that allow to interpret several layer ids as a single layer id. At the decoder side, the decoder choses a target VirtualLayerId, and decodes all layer_ids that belong to that VirtualLayerId and ignores any other layer_ids.
- 15 An example embodiment for the VirtualLayerId definition is shown in the following in the SPS. However, this syntax could also exist in a VPS or any other parameter set with a scope of multiple layers, i.e. different values of layer_id).

	seq_parameter_set_rbsp() {	Descriptor
	sps_decoding_parameter_set_id	u(4)
	sps_max_sub_layers_minus1	u(3)
	sps_reserved_zero_5bits	u(5)
	...	
42	sps_virtual_layer_enabled_flag	u(1)
	if (sps_virtual_layer_enabled_flag) {	
	sps_num_virtual_layers_minus1	ue(v)
	for(i = 0; i < sps_num_virtual_layers_minus1 + 1; i++) {	
	sps_num_layers_in_virtual_layer_minus1[i]	ue(v)
	for(j = 0; j < sps_num_layers_in_virtual_layer_minus1[i] + 1; j++)	72;
	sps_layer_id_in_virtual_layer[i][j]	u(7)
	}	
	}	
	...	
	}	

- 20 The VirtualLayerId could be derived or explicitly signaled in the parameter set.

It is problematic, when layers grouped together through a virtual layer use different chroma formats (e.g. 4:2:0 to 4:4:4). As its not trivial to arrange pictures with different chroma formats in a common buffer, measures have to be taken to solve this problem. In another embodiment of the invention, it is a requirement of bitstream conformance that all layers within a virtual
5 layer have the same chroma format, i.e. chroma subsampling and chroma position. Thereby, arrangement of layer pictures in a common picture can happen equally in each component buffer as the layer picture dimensions are equal for each component.

2. Adaptive Streaming with independent layers

10 Contrary to the use-case described above, a bitstream may consist of multiple representations of the same content (i.e. for multiple resolutions), each representation indicated by a different layer_id, multiplexed into the same bitstream and bundled through a virtual layer id as described in aspect 1.

15 In a system, e.g. adaptive streaming or RTP point-to-point communication, only one layer would be transmitted at a time / per segment / per picture depending on receiver decoder capabilities, display device, network characteristics and so on.

20 Hence, the layer to be decoded, may change over time, e.g. for ARC/RPR, which could be used for switching target resolutions, e.g. when Open GOP structures are used, or at some pre-defined switching points, where the drift for switching is under control. That is, as further outlined below, it might be that layer switching happens at some points only, called switching points, for instance. The encoder may signal some points where this referencing a different
25 layer can be done.

While virtual layers can be used to indicate a set of target layers to be expected in the bitstream, it is required to indicate that the layers belonging to the same virtual layer are not to be decoded simultaneously, but only one at a time. The invention therefore is to provide
30 measures to indicate that only a single layer is to be decoded and further to ensure correct selection/detection of the corresponding layer in the decoding process.

In a first embodiment of the invention can be found in the following, with changes to the aspect 1 syntax highlighted in yellow.

		Descriptor
	seq_parameter_set_rbsp() {	
	sps_decoding_parameter_set_id	u(4)
	sps_max_sub_layers_minus1	u(3)
	sps_reserved_zero_5bits	u(5)
	...	
	sps_virtual_layer_enabled_flag	u(1)
42	if (sps_virtual_layer_enabled_flag) {	
120	sps_num_virtual_layers_minus1	ue(v)
	for(i = 0; i < sps_num_virtual_layers_minus1 + 1; i++) {	
	sps_num_layers_in_virtual_layer_minus1[i]	ue(v)
	sps_exclusive_layer_in_virtual_layer[i]	u(1)
	for(j = 0; j < sps_num_layers_in_virtual_layer_minus1[i] + 1; j++)	
	sps_layer_id_in_virtual_layer[i][j]	u(7)
	}	
	}	
	...	
	}	

sps_exclusive_layer_in_virtual_layer[i] Indicates that only a single value for layer_id in the i-th virtual layer shall be decoded per AU.

- 5 In other words, a decoder encountering a bitstream that according to the invention either:
- changes resolution through concatenating two sub-streams with different layer_id, both belonging to the targeted virtual layer id, or
 - containing the data of layers with multiple layer_id values, all belonging to a targeted virtual layer, even for the same AU

10 shall be aware that the two values of layer_id belong to the same virtual layer it is currently decoding, potentially spatially resample reference picture, and decode only pictures of one layer_id value per AU to be found in the bitstream of the target virtual layer id given as decoder control input.

15 In this embodiment, it is a requirement of bitstream conformance or there is a flag in the bitstream indicating that no access unit in a bitstream (as received by the decoder) shall contain slice NAL units with more than a single layer_id value in their header per access unit.

20 In another embodiment, the decoder shall decode the highest value of layer_id in the virtual layer id to be found in the bitstream at a certain access unit.

The decoder in this case, taking into account the signaling in the bitstream, is aware that only one layer_id per AU pertaining to the virtual layer with VirtualLayerId is to be decoded, either because there is only one layer_id per AU belonging to the virtual layer or selecting the highest available one that matched the decoder capabilities.

As discussed above, the switch could happen at any picture. In such a case, the encoder would take care of carrying out a more constraint encoding, with any version that can be used for reference if a change should happen at any AU would lead to same values of e.g., MV when TMVP is used. Alternatively, in another embodiment, the encoder would decide to encode some sub-set of pictures in time, e.g every 8th picture as a switching point and make sure that TMVP is not used or that when TMVP is used the reference MVs used for prediction of TMVP lead to the same result, or the DMVR (Decoder-side Motion Vector Derivation), i.e. the encoder exploits that the decoder performs motion estimation based on available picture data in the decoded picture buffer, just as the encoder does, and the encoder reduces, accordingly, the motion vector side information, is not used. I.e., there would be some signaling indicating pictures that are valid for switching layer, while still being able to decode a version that is drift-controlled and leads to a good quality.

3. Error resiliency and enhanced referencing

Using references from higher layers could improve the efficiency of lower layer when decoding all the layers. Therefore, under some circumstances where the efficiency penalty of layered coding is too high, it could be beneficial to predict lower layer from higher layer pictures (pictures with higher fidelity) instead of using the corresponding picture of lower layers. However, when on-the-fly adaptation is used and the reference used at the encoder is not available at the decoder, the decoding process could use the corresponding picture available at a lower layer, albeit the incurring drift issues. As known from previous standards SVC (see MGS) some drift might be acceptable as long as there are measures to keep it under control. This requires some picture to break the drift, by making sure that for some picture the referencing is carried out from a picture that cannot be dropped, e.g. explicitly indicating that the reference picture belongs to the same layer or to the lowest layer.

The invention is to carry related indication in the reference picture list signaling to enable a decoder to make the right (same as encoder) or under sub-optimal circumstances (given not all data is available) the best decision. In embodiments of the invention, the reference picture list in addition to indicating the POC of a reference picture:

- Indicates the layer of a reference picture, i.e. the actual used picture on encoder-side
- Indicates the virtual layer of a reference picture, i.e. a group of visually similar pictures
- Indicates the "preferred" layer in the virtual layer of a reference picture, i.e. the actual used layer in a group of virtual pictures
- 5 - Indicates a ranking for all the layers in the virtual layer of a reference picture, i.e. for a decoder to select among alternatives in absence of the actual used layer.

An embodiment of the invention can be found in the following with a syntax example in the SPS (although it could be VPS which is more adequate for layer information).

10

	seq_parameter_set_rbsp() {	Descriptor
	sps_decoding_parameter_set_id	u(4)
	sps_max_sub_layers_minus1	u(3)
	sps_reserved_zero_5bits	u(5)
	...	
	sps_cross_layer_reference_enabled_flag	u(1)
	if (sps_cross_layer_reference_enabled_flag) {	
132	sps_cross_layer_reference_sets_minus1	ue(v)
	for(i = 0; i < sps_cross_layer_reference_sets_minus1 + 1; i++) {	
	sps_target_layer_for_cross_layer_set[i]	ue(v)
	sps_num_layers_in_cross_layer_set_minus1[i]	u(1)
	for(j = 0; j < sps_num_layers_in_cross_layer_set_minus1[i] + 1; j++)	
	sps_layer_id_in_cross_layer_set[i][j]	u(7)
	}	
	}	
	...	
	}	

130

With the syntax as an example it could be signaled that a layer uses any layer with a layer_id value equal to one of sps_layer_id_in_cross_layer_set[i][j] as reference, or only the same layer_id value. The order in the for loop for sps_layer_id_in_cross_layer_set[i][j] could be interpreted as a ranking and priority (or preference) of which version should be used.

15 Alternatively, the raking could be explicitly signaled.

The syntax in the parameter sets would apply to the whole bitstream but there could be updates specific to some given AUs. Such a signaling should be done for each NAL unit, e.g. at the slice header:

20

134

slice_header() {	Descriptor

if(sps_cross_layer_reference_enabled_flag) {	
cross_layer_reference_update_flag	u(1)
if(cross_layer_reference_update_flag)	
current_layer_reference_flag	u(1)
}	
if((NalUnitType != IDR_W_RADL && NalUnitType != IDR_N_LP)	
sps_idr_rpl_present_flag) {	
for(i = 0; i < 2; i++) {	
if(num_ref_pic_lists_in_sps[i] > 0 &&	
(i == 0 (i == 1 && rpl1_idx_present_flag)))	
ref_pic_list_sps_flag[i]	u(1)
if(ref_pic_list_sps_flag[i]) {	
if(num_ref_pic_lists_in_sps[i] > 1 &&	
(i == 0 (i == 1 && rpl1_idx_present_flag)))	
ref_pic_list_idx[i]	u(v)
} else	
ref_pic_list_struct(i, num_ref_pic_lists_in_sps[i])	
for(j = 0; j < NumLtrpEntries[i][RplIdx[i]]; j++) {	
if(ltrp_in_slice_header_flag[i][RplIdx[i]])	
slice_poc_lsb_ltrp[i][j]	u(v)
delta_poc_msb_present_flag[i][j]	u(1)
if(delta_poc_msb_present_flag[i][j])	
delta_poc_msb_cycle_ltrp[i][j]	ue(v)
}	
}	

}	

}	

In the example above, for a given AU a layer would indicate at 134 that irrespective of the presence of cross layer reference sets, the current AU uses for a given layer only the pictures from a specific layer, e.g. the same layer as reference or some other specific layer.

5 Alternatively, the syntax could update the set of layers and rankings that can be used for

reference and reduce the set, e.g., remove one layer from the set of allowed layer_ids for the current AU.

5 Similar to what has been discussed for aspect 2, whenever some pictures are allowed to use a different reference picture (with a different value of layer_id), the invention here might require making sure that TMVP or some other kind of syntax prediction (non-sample prediction) is the same when using any of the references with different layer_id values. Alternatively, TMVP or some other kind of syntax prediction (non-sample prediction) would not be performed for such pictures or those would be done only from a specific picture, potentially indicated or being the
10 lowest layer_id of the group or the layer_id with the same value as the picture being decoded.

The later of the aspects listed above could be used for instance for an adaptive streaming following scenario.

15 According to an embodiment shown in Fig. 7 only pictures belonging to the lowest temporal layer representation are encoded at several bitrates 12_1 , 12_2 and 12_3 and the higher temporal layers are encoded once, i.e. one layer contains pictures of the higher temporal layers whereas all layers contain pictures of the lowest temporal layer.

20 If all layers are included, only the layer_id of the layer containing version 12_1 , i.e. the corresponding dark blue picture (or a default value), is decoded when decoding the (sub)bitstream containing the pictures as in the example. If used in adaptive streaming, typically only one of the versions of the three in the example would be present. In such a case the parameter sets could indicate which is the layer present or the virtualLayer concept
25 discussed above could be used, where layers are exclusively present.

If such a concept is to be applied in a system, it is clear that additional coding constraints are beneficial to limit the amount of drift when switching between exchangeable pictures on client side.
30

In one embodiment, exchangeable pictures are not used as source for TMVP or ATMVP. Thereby, if a client performs a switch between these variants of the exchangeable pictures, no faulty TMVP or ATMVP based inter-prediction occurs.

35 In another embodiment, BDOF and DMVR are prohibited to be applied based on exchangeable pictures to limit possible drift likewise.

In another embodiment, the parameters used for adaptive loop filter (ALF), i.e. the set of filter kernel, optimized for the picture content, is supposed to be the same for all variants of such an external picture. In other words, all exchangeable picture of the same time instance only apply filters of a common set of ALF filters, e.g. carried in one or more APS associated with the exchangeable picture.

In addition, with the following applying to both, the present aspect as well as the previous one, there may be a special treatment or usage of ALF (applying to aspects regarding layers corresponding to different qualities of pictures). This relates to ALF. The ALF filters can be signalled for more than one picture. In the examples where a different layer is used for reference (either for case 2 or 3 (applying to aspects regarding layers corresponding to different qualities of pictures)), it could be mandated that the ALF filters used for any picture are signalled with the pictures for which more than one reference layer might be used alternatively and the ALF parameters of the used reference layer are used for the referencing pictures.

In another embodiment, in addition to the above, all pictures between two exchangeable pictures only apply filters of a common set of ALF filters, e.g. carried in one or more APS associated with the first exchangeable picture.

The above aspect **"multi-layer for 360 video regions"** is briefly summarized in the following.

Fig. 1 shows an example for encoder and decoder and a data stream sent from encoder to decoder. The encoder encodes a video 6 which is composed of pictures 12 into the multi-layer video data stream 32. The video 6 is partitioned into picture regions 54 and the encoder 8 independently encodes each picture region 54. That is, all the pictures 12 of video 6 are subdivided into regions 54 in the same manner, i.e., so that the region 54 into which pictures 12 are subdivided coincide in terms of region boundaries. The term "picture region" is used both, to denote the region 54 of a picture as well as the collections of mutually co-located regions 54 of all pictures with the meaning getting clear from the corresponding context where the respective term is used.

Each picture region 54 is encoded by encoder 8 into corresponding video data stream portions 16. That is, each data stream portion 16 is associated with, and has encoded therein, a corresponding picture region 54 of one picture. More precisely, each portion 16 has only the associated region 54 of a certain picture 12 encoded therein, or a part thereof. In Fig. 1, for instance, it is shown that each picture 12 of video 6 is encoded into a sequence of portions 16

which, together, form an access unit 30, which access unit corresponds to the respective picture. In Fig. 1, this is illustrated by using a corresponding number of apostrophes. The encoder embeds the video data stream portions 16 into the data stream 32 which is a multi-layered video data stream. Each portion 16 is provided with a layer indication 18 which indicates the layer the respective portion 16 belongs to such as by indicating the layer ID of that layer. The encoder associates portion 16 with the layers of data stream 32 in a manner so that portion 16 having one picture region 54 encoded thereinto belong to one layer which is different from a layer to which portions 16 belong which have a different (off-set) picture region 54 encoded thereinto.

The encoder 8 provides the multi-layer video data stream 32 with a layer grouping information 40 which indicates a virtual layer with which the layers which portions 16 belong to, are associated. According to an embodiment, the layer grouping information 40 indicates a grouping of the layers which the video data stream portions 16 belong to, into one or more groups 72 of layers, with a virtual layer being associated with each group of layers. In other words, the encoder 8 encodes each picture region 54 of each picture 12 into one or more corresponding portions 16 of data stream 32. Coding dependencies merely exist between portions 16 corresponding to the same region 54, or collocated regions 54 in case of the regions belonging to different pictures 12. The encoder 8 associates a different layer with each video region 54 video 6 is subdivided into. The number of different layers which the portion 16 are associated with and which have video 6 encoded thereinto, thus corresponds to the numbers of video regions 54. An example for the layer indication information 40 is shown in Fig. 2. It groups the layers 70, i.e., the layers which any of portion 16 having video encoded thereto belongs to, into one or more groups 72 and indicates or associates, a virtual layer for/with each layer group 72.

The effect of the above-outlined procedure becomes apparent when looking at the decoder behavior. The decoder is also shown in Fig. 1. The decoder is for decoding the multi-layer video data stream 32 which is partitioned into the partitions 16 each of which comprises the layer indication 18 indicating the layer the respective portion belongs to. The decoder 28 reads, by way of a layer grouping indication reader 44 from data stream 32 the layer grouping information 40 which indicates a grouping of the layers, which the portions 16 of data stream 32 belong to, into one or more groups 72 of layers and associates a virtual layer 42 with each layer group, e.g. associate a virtual layer 1 with layer group 72' and a virtual layer 2 with layer group 72" as shown in Fig. 2. The decoder then forms, by way of a formatter 46, a video data stream 48 out of those portions 16 of data stream 32, which belong to the group of layers associated with a predetermined virtual layer 50 which is set, for instance, by another entity

such as a user, an application or the like. The formation is done by taking over these portions 16 from data stream 32 into data stream 48 with leaving the layer indication 18 of the portions 16 thus taken over unchanged. Data stream 48 is then subject to a single layer decoding by forwarding same to a single virtual layer decoder 52 which decodes data stream 48 as a single layer data stream although the portion 16 contained therein are still associated with different layers, namely the layers belonging to the selected/predetermined group of layers correspond to the virtual layer 50, but decoder 52 disregards the different layer association as decoder 52 expects the portions taken over into data stream 48 to belong to one virtual layer and to, thus, form a single-layer-like data stream.

Fig. 2 shows two examples for groups of layers. Group 72', for instance, comprises all layers which any of portions 16 having video 6 encoded thereinto, belongs to. All portions 16 would, thus, be taken over into data stream 48 in case of the virtual layer 50, i.e. the predetermined virtual layer, indicating that group 72'. The formation of data stream 48 would, thus, be very easy, at least as far as the formation of data stream 48 is concerned as the portions 16 may simply be taken over from data stream 32 to data stream 48.

Further measures may have been taken by the encoder to guarantee that the formation is kept simple. For instance, the encoder may set position information in the portion 16 so that they are valid when the portions 16 put together to form the data stream. The position information may relate to CU (coding unit) addresses. the CU addresses, i.e., the addresses contained in each portion 16 and explaining as to which part of the corresponding picture 12 is encoded into the respective portion 16 would have been coded by encoder 8 in such a manner that, when the portion 16 is taken over from data stream 32 to data stream 48, and when the data stream 48 is deemed as a single layer data stream by decoder 52, which has the video 6 with the corresponding regions 54 encoded thereinto, i.e., the regions belonging to the layer group of the virtual layer 50, then no conflict results, i.e., the location of these parts are correct. The position information in portions 16 may, for instance, relate to a certain default/reference position of pictures 12 such as the upper left corner. Alternatively, the locations of the parts coded into each portion 16 may be indicated by the location information in each portion 16 in a manner relative to a default/reference position of the region 54 containing the encoded part of the respective portion 16 and the formation 56 would merely involve accompanying the portions 16 taken over from data stream 32 to data stream 48 with a parameter set explaining/describing the spatial distribution of the default positions of those regions 54 coded by the portions 16 which belong to any of the group of layers associated with the virtual layer 50. The just-mentioned parameter set may comprise a picture region arrangement information indicating, for each virtual layer 42 indicated in information 40, an output picture position of the

regions within an output video composed of one or more regions coded into the video data stream portions 16 belonging to the group of layers associated with the respective virtual layer. For instance, the output pictures of the output picture in case of the layer group corresponding to the virtual layer 50 merely comprising a proper subset of layers in set 20, would be smaller than the pictures 12 and the positions of regions 54 relative to the reference position of the output pictures may change relative to their positions within pictures 12, while the position of the coded parts, coded in each portion 16, relative to the reference position of the region 54 containing that part, remains the same.

The example of Fig. 2, for instance, shows that such a group of layers associated with a certain virtual layer may likewise comprise merely a proper subset of all layers, such as layer group 72". In that case, in forming data stream 48, merely a proper subset of all portions 16 would be taken over from data stream 32 to data stream 48, namely only those belonging to the group of layers associated with the virtual layer 50.

This is shown in Fig. 6b, for the predetermined virtual layer 50 with virtual layer ID 2 associated with the layer group 72". In this case, only the portions 16 of the multi-layered video data stream 32 relating to picture regions 54 indicated by the triangle and by the upside-down triangle are taken over to the video data stream 48. This is due to the fact, that the layer group 72", as shown in Fig. 2, comprises only picture regions 54 indicated by the triangle and by the upside-down triangle associated with the layer ID's 1 and 2.

Each video data stream portion 16 is, for example, provided with a relative position information 17 indicative of a picture position of a picture part, e.g. of the picture position 117₁ of the picture part 116₁ or of the picture position 117₂ of the picture part 116₂, which is coded into the respective video data stream portion 16 in a manner so that the relative position information 17 is indicative of the picture position relative to the picture region 54 the picture part 116₁/116₂ is located in. The picture position 117₁, i.e. location, of the picture part 116₁ coded into a portion 16 may be indicated in a manner relative to a reference position of the region 54, indicated by the triangle, containing the encoded part of the respective portion 16. The reference position is, for example, the upper left corner of the picture region 54.

The relative position information 17 of a portion 16 may indicate that the top left corner of the picture part 116₁ encoded into the respective portion 16 does not deviate from the position of the top left corner of the picture region 54 the picture part 116₁ is located, according to the embodiment shown in Fig. 6b. The relative position information 17 of another portion 16 may indicate that the top left corner of the picture part 116₂ encoded into the respective portion 16

is offset to the position of the top left corner of the picture region 54 the picture part 116₂ is located, according to the embodiment shown in Fig. 6b.

The relative position information 17 is, for example, provided by an encoder 8 for each portion 16 and a decoder 28 is, for example, configured to read this relative position information 17 from each portion 16.

The multi-layered video data stream 32 may further comprise a picture region arrangement information 110 indicating, for each virtual layer 42, an output picture position of the regions 54 within an output video composed of one or more regions 54 coded into the video data stream portions 16 belonging to the group 72 of layers associated with the respective virtual layer 42. Alternatively, the multi-layered video data stream 32 may further comprise a picture region arrangement information 110 indicating positions of the picture regions 54 within the pictures of the video.

The picture region arrangement information 110 is, for example, provided by an encoder 8 in the multi-layered video data stream 32 and a decoder 28 is, for example, configured to read this picture region arrangement information 110 from the multi-layered video data stream 32.

An example for the picture region arrangement information is illustrated in Fig. 6a for the example of Fig. 2 and indicated using reference sign 110. The output pictures' 112 size depends on the virtual layer, just as the positions 114 of regions 54 corresponding to the virtual layer's layer set 72 within the output picture 112 do. Alternatively, the picture region arrangement information 110 merely comprises the positions of regions 54 within the original pictures 12 and a correct parameter set with the positions 114 of regions within the output picture 112 are determined by formatter 46 based on an analysis of subset of regions 54 corresponding to the subset of layers associated with the virtual layer 50.

Beyond the positional constraints relating to the position information, e.g. slice addresses or CU addresses, in the portions 16, further constraints might be obeyed by the encoder. For instance, for each virtual layer, the corresponding layer set 72 might be restricted to ones yielding a rectangular output picture shape of the output pictures 112. In other words, for each virtual layer, the group of layers, i.e. the layer set 72, associated with the respective virtual layer yields a video data stream 48 out of portions 16 belonging to the group of layers associated with the respective virtual layer which has an output video of output pictures of rectangular output picture shape encoded therein. Alternatively or additionally, it might be that all layers within a virtual layer's layer set 72 have to have the same chroma format, i.e.

coincide in chroma subsampling and chroma position, and that the encoder keeps care that this requirement is met. In other words, for each virtual layer, the portions 16 of all layers within the group of layers, i.e. the layer set 72, associated with the respective virtual layer have the same chroma format. Alternatively or additionally, it might be that the CTU (coding tree units) sizes, i.e. the sizes of the blocks into which each picture is sub-divided regularly in rows and columns before each CTU is further subdivided by recursive multi-tree subdivision into CUs, may have to have the same size in all layers within a virtual layer's layer set 72 and that the encoder keeps care that this requirement is met. In other words, for each virtual layer, the portions 16 of all layers within the group of layers associated with the respective virtual layer are encoded using a coding tree units subdivision in CTUs of sizes which are equal among the layers within the group of layers associated with the respective virtual layer. Alternatively or additionally, it might be that the POC assignment to the pictures 12 may have to be the same in all layers within a virtual layer's layer set 72 and that the encoder keeps care that this requirement is met. In other words, for each virtual layer, the portions 16 of all layers within the group of layers associated with the respective virtual layer are encoded a POC to picture assignment to the pictures 12 which is the same among the layers within the group of layers associated with the respective virtual layer. Each picture 12 comprises different picture regions 54 which are associated with individual layers 70. Portions 16 associated with the same picture 12 but belonging to different layers 70 out of the group 72 of layers have, for example, the same POC assignment. However, these examples are merely optional as, for example, the video codec may allow for output pictures with other shapes, the CU subdivision may be fixed anyway, the video may be monochrome, and/or the POC assignment may be inevitably adapted to each other.

We now briefly explain the aspect **"adaptive streaming with independent layers"**. Let us inspect Fig. 3. In accordance with this aspect, an encoder 80 encodes a video 82 of pictures 12 in different qualities such as different spatial resolution. Encoder 80 may have encoded video 82 in a manner with encoding each quality thereof independently. That is, encoder 80 would encode, using motion-compensated prediction, the pictures 12 of video 82 into portions 16 of a multi-layer data stream 84 so that each portion 16 has an associated picture 12 of the video 82, or a part thereof, encoded therein. Each portion 16 comprises a layer indication indicating the layer the respective portion belongs to. That is, the multi-layer data stream 84 output by encoder 80 would comprise one or more portions 16 for each quality, for each picture. Each portion belonging to a certain quality would belong to a corresponding layer. All the layers which any of these portions belongs to, would, thus, correspond to the number of qualities at which encoder 80 encodes video 82. In Fig. 3, this collection of layers is indicated as layer set 20. The encoder 80, additionally, inserts into multi-layered video data stream 84, a layer

grouping information 50 which groups the layers of the portions in data stream 84 into layer group 20, i.e., indicates them as forming a group. For instance, further layers may be conveyed by the multi-layer video data stream. Further, more than one group may be indicated by layer grouping information 50 as outlined in the syntax examples presented above.

5

Let us refer to Fig. 4 which shows an example for the initial data stream 84 again. As shown, for each of a set of layers 20, the encoder 80 encodes the video into corresponding portions 16 which belong to the respective layer, each portion coding a corresponding picture 12, or a part thereof. The coding is done using temporal or motion-compensative prediction as illustrated in Fig. 4 using horizontal arrows 24 which point from picture to another picture within the same layer or, to be more precise, from one version of a picture to a version of another picture, both versions corresponding to the corresponding layer. That is, when encoding a version 12_i' of a picture 12' into corresponding portion(s), the encoder 80 performs motion-compensative prediction from a version 12_j' of another picture 12'', the so-called reference picture, that is reconstructable from corresponding portion 16 which belong to the same layer i as those currently coded. Thus, no inter-layer prediction is used by encoder 80 and, thus, in Fig. 4, no vertical arrows point from lower layer picture versions 12_j to higher layer picture versions 12_{i>j}. Other than the example of Figs. 1 and 2, each layer has the complete video encoded therein, i.e., any one of these layers of set 20 is sufficient in order to represent the video. Insofar it would, accordingly, be superfluous to convey data stream 84 completely to a decoder as merely one layer or, alternatively speaking, the portion 16 belonging to one layer would be of interest. As already denoted above, the representation of the video by any of the layers may differ among the layers with respect to quality such as in terms of spatial resolution, bit depth but any other differences are also within the scope of the present embodiment.

25

As shown in Fig. 3, the data stream 14 finally forwarded to scalable decoder 90, a decoder 90 capable of handling the multi-layer data stream 14, may have been modified relative to the initial data stream 84 by way of intermediate entities of apparatus 10 such as a pair of server 92 and client 94 which may co-operate according to an adaptive streaming protocol such as DASH, so as to forward from the server 92 to the client 94 for each of temporal fragments into which the video 82 is temporally subdivided such as consecutive sequences of pictures, merely those portions 16 of data stream 84 which belong to one of the layers out of set 20, namely one chosen or selected adaptively for the corresponding fragment. Thus, insofar, the server 92 acts as an issuer of the data stream 84 and the client 94 is a retriever. For each temporal fragment, the server 92 issues merely portions 16 having the corresponding temporal fragment encoded therein, which belong to one of the layer set 20, namely the one selected for the corresponding temporal fragment, and the retriever or client 94 receives the corresponding

35

portions 16. The layer selected for the corresponding temporal fragments varies over time. Information 50 remains in the data stream 14.

The decoder 90 receives the "stripped" data stream 14, i.e., the incomplete data stream compared to data stream 84, namely a data stream 14 where at some point in time, the subset of layers changes the portions 16 belonging thereto have been stripped off from the original data stream 84. The variation in layer and, thus, quality at which the video is represented over time may, for instance, be adapted to changes in the transmission capacity between server 92 and client 94. In performing the adaptation most efficiently, the association of a layer to each picture 12 would be done in a manner so that exactly one layer out of set 20 is chosen for each picture, but as illustrated in Fig. 3 for picture 12' and the corresponding access unit 30' and in Fig. 4 in the example illustrated with the dashed dotted lines, it might be that for selected ones of the pictures or, alternatively speaking, how selected ones of temporal fragments of the video, more than one layer is present in data stream 14. It could be, however, a requirement for data stream conformance of data stream 14, that one layer is present in data stream 14 for each picture 12 exclusively. Irrespective of this requirement being posed onto data stream 14 or not, decoder 90 treats layers within set 20 as mutual substitutes from one another and as being coded independently from each other, and accordingly, whenever a certain portion 16 is to be decoded by motion-compensative prediction on the basis of a reference picture for which, however, there is no portion 16 present in the data stream, i.e., no portion 16 has encoded therein this reference picture 16, which would additionally belong to the same layer as the portion currently to be decoded, then the decoder 90 simply derives the motion-compensative prediction signal based on a substitute version of this reference picture, namely on the basis of the version of the reference picture derivable from any portion 16 in the data stream present for that reference picture and belonging to a different layer. Up-sampling or down-sampling may be used to this end in case of, for instance, the two layers being of different spatial resolution. In other words, the decoder 90 is configured to up-sample and/or down-sample the pictures decoded from the portions 16 associated with a first layer of the group 20 of layers to decode, by the motion-compensated prediction, the portions 16 associated with a second layer of the group 20 of layers. In any case, the decoder 90 does not detect any fault. Rather, the decoder 90 is "prepared" of such situations. In even other words, decoder 90 reads, from the multi-layer video data stream 14, the layer grouping information 50 and derives therefrom the knowledge about layer group 20. In other words, the layer grouping information 50 groups the layers the portions 16 belong to, into which the video is coded, into a group of layers, i.e. the layer group 20. The scalable decoder 90 gathers portions of the multi-layer video data stream 14, which belongs to layer group 20. As already denoted above, if portions 16 of other layers were present in data stream 14, they would be stripped off now internally within decoder 90.

In gathering portion 16, decoder 90 prepares the gathering such that each picture of the video is associated with one layer of layer group 20 and for each picture of the video, the one or more portions 16 belonging to this one selected layer are gathered, e.g., gather only those in within the dashed lines 62, as shown in Fig. 4, although all in the dash-dotted lines are arriving.

For instance, decoder 90 could be configured to, by default, always select the highest available layer within layer set 20. For instance, if portions of layers according to the dashed dotted lines of Fig. 4 were arriving as data stream 14 at decoder 90, same could, for instance, be configured to select for each picture one layer, thereby arriving at gathered portions 16 belonging to layers as illustrated by dashed lines 62 in Fig. 4.

Decoder 90 then decodes the video from the gathered portions by use of motion-compensated prediction with supporting that pictures decoded from portion associated with a first layer of the group of layers being referenced, for the motion-compensated prediction by portions associated with a second layer of a group of layers, different from the first layer.

As explained above, both of the above briefly summarized aspects may be combined: A flag 120 may signal whether a certain virtual layer pertains to a region layer set 72 or a set 20 in terms of Fig. 4 forming a set of mutually exchangeable layers between which the layer selection may vary even between referencing and referenced pictures. As shown in the above syntax example, even more than one such virtual layer's mutually exchangeable layer set 20, may be indicated in information 50, one per virtual layer 42.

In other words, the decoder 90 is, for example, configured to read a predetermined layer group handling indication, i.e. the flag 120, from the multi-layered video data stream 84. If the predetermined layer group handling indication 120 has a first state, the decoder 90 is configured to perform the gathering of the portions 16 of the multi-layered video data stream 84, which belong to the group 20 of layers, so that each picture 12 of the video is associated with one layer of the group 20 of layers, and for each picture 12 of the video, the one or more portions 16 belonging to the layer associated with the respective picture 12 are gathered. Furthermore, if the predetermined layer group handling indication 120 has a first state, the decoder 90 is configured to decode the video from the gathered portions by use of motion-compensated prediction with supporting that pictures decoded from portions associated with a first layer of the group 20 of layers being referenced, for the motion-compensated prediction, by portions associated with a second layer of the group 20 of layers, different from the first layer. If the predetermined layer group handling indication 120 has a second state, the decoder 90 is configured to form a video data stream 48 out of portions 16 of the multi-layered video data stream 32, which belong to the group 72 of layers, by taking over the portions 16 into the

video data stream 48 with leaving the layer indication 18 of the portions 16 unchanged so that each portion 16 of the multi-layered video data stream 32, which belongs to the group 72 of layers, is present in the video data stream 48. Furthermore, if the predetermined layer group handling indication 120 has a second state, the decoder 90 is configured to decode the video data stream 48 as a single layer data stream, the group of layers being coded independent from each other and relating to mutually different picture regions 54 of a video represented by the single layer data.

It should be noted that the above embodiment may be varied in a manner where the encoder 80 adapts cross-layer motion compensated prediction on the fly so that the encoder does the same as described so far with respect to the decoder when the reference picture's version used for motion compensated prediction deviates from the currently coded/decoded portion's layer, namely performing, for instance, up/downsampling in case of spatial resolution differences between the layers or the like. Here, no drift would occur as decoder and encoder would act the same.

It should be noted that it may be a constraint that layer switching is only performed/allowed at some switch points, for which, e.g. TMVP is not used or the predictors used for deriving motion-vectors are the same in any version. A corresponding signaling may be used.

The multi-layered video data stream, for example, comprises a layer switching information which informs a decoder 90 at which pictures 12 of the video the portions 16 associated with the predetermined layer of the group 20 of layers, which reference, for the motion-compensated prediction, pictures for which portions, which belong to the predetermined layer of the group 20 of layers, are absent in the multi-layered video data stream, are allowed to occur. This layer switching information, for example, is signaled in the multi-layered video data stream by the encoder 80. The decoder 90 or any other data stream handling apparatus, may read from the multi-layered video data stream the layer switching information and derive therefrom at which pictures of the video a layer switching is allowed. Thus, from the layer switching information the pictures may be derived for which stream portions of a reference picture, referenced in terms of motion-compensated prediction, can also belong to another layer than the layer of portions into which the pictures are coded into the data stream. Thus, at the switch points, no reference associated with the same layer as the portions of the pictures at the switch points are needed for the motion-compensated prediction. In other words, at the switch points, the decoder 90 can be configured to decode that pictures from portions associated with a first layer of the group of layers being referenced, for the motion-compensated prediction, by portions associated with a second layer of the group of layers,

different from the first layer. The decoder uses to this end, the reconstructed version of the reference pictures of the different layer than the referencing picture's stream portions such as by additionally using up- or down-sampling or the like. A stream handling apparatus may take advantage of such layer switching information in case of this information already persisting in a data stream which comprises all layers for all pictures: The stream handling apparatus may then use this information to fragmentize the complete data stream into fragments temporally starting or abutting at the indicated pictures, and fragmented into fragments of different layers so as to be subject to a selection by an adaptive streaming client with respect to fragments of different layers, but belonging to the same time interval.

According to an embodiment, the layer switching information can indicate the pictures at which a layer switching is allowed for a sequence of pictures as occurring at a regular pattern such as every n^{th} picture with n being an integer, and/or picture individually.

According to an embodiment, the decoder 90 is configured to decode the pictures 12 of the video from the portions 16 of the multi-layered video data stream using an adaptive loop filter (ALF). Furthermore, the decoder 90, for example, is configured to derive from the portions 16 associated with the predetermined layer of the group 20 of layers, which reference, for the motion-compensated prediction, pictures, e.g. 12'', for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream, ALF parameters for use with respect to the predetermined layer and for use with at least one other layer and filter the pictures decoded from the portions associated with the first layer of the group of layers to decode, by the motion-compensated prediction, the portions associated with the second layer of the group of layers, using ALF parameters derived from the portions associated with the second layer of the group of layers for the first layer. With regard to Fig. 4 this can be interpreted the following way, wherein picture 12₃'' is to be decoded. Picture 12₃'' is associated with the layer 3 as the predetermined layer representing the aforementioned second layer of the group 20 of layers. For the motion-compensated prediction, picture 12₃' would be the preferred reference picture, or, to be more precise, might have been the actual motion compensated reference basis at the encoder side. However, for picture 12' portions 16, which belong to the predetermined layer 3 of the group 20 of layers, are absent in the multi-layered video data stream as indicated by the dashed-dotted lines. The decoder 20 uses adaptive loop filtering to filter the basis for motion compensated prediction. ALF parameters are coded in the data stream for each motion predicted picture. Picture 12₃'' is assumed to be such a picture and has first ALF parameters coded into data stream for the ALF filtering. They had been defined by the encoder for improving the motion compensated prediction of picture 12₃'' by filtering picture 12₃'. However, the data stream also has second ALF parameters coded

thereinto for picture $12_3''$, which are defined by the encoder for improving the motion compensated prediction of picture $12_3''$ in case of using picture $12_1'$ to form the motion compensated prediction source, namely in order to improve that source by filtering picture $12_1'$ or the up-sampled version thereof. Thus, the decoder 90 is, for example, configured to derive
5 different ALF parameters from portions 16 into which picture $12_3''$ is encoded, namely ones intended for intra-layer use and at least ones for inter-layer motion compensated prediction use, and uses the latter ALF parameters for filtering picture $12_1'$ which corresponds to portions 16 belong to layer 1 representing the aforementioned first layer of the group 20 of layers, or filtering an up-sampled version of that picture $12_1'$. For the motion-compensated prediction of
10 the picture $12_3''$ associated with the second layer of the group 20 of layers the adaptive loop filtered picture $12_1'$ associated with the first layer of the group 20 of layers can be used as reference.

It should be noted that the feature described above may be varied in a manner where the
15 encoder 80 encodes the pictures of the video into the portions 16 of the multi-layered video data stream 14 using an adaptive loop filter (ALF) and signal for the portions associated with the predetermined layer of the group of layers, which reference, for the motion-compensated prediction, pictures, e.g. $12''$, for which portions, which belong to the predetermined layer of the group 20 of layers, are absent in the multi-layered video data stream 14, the ALF
20 parameters for use with respect to the predetermined layer and for use with at least one other layer, e.g. for which portions having the referenced pictures encoded thereinto and belonging to the at least one other layer are present in the multi-layered data stream.

The following description briefly summarizes the above-aspect **“error resiliency and enhanced referencing”**. This aspect is explained with respect to Figs. 3 and 5. The base
25 situation is pretty similar to the one outlined above with respect to Figs. 3 and 4. That is, an apparatus forms a multi-layer video data stream 14 having a video 82 encoded thereinto and in doing so, an encoder 80 of apparatus 10 encodes, using motion-compensated prediction 24, pictures 12 of the video into portion 16 of a multi-layered video data stream 84. Each
30 portion 16 has an associated picture 12 of the video, or a part thereof, encoded thereinto and comprises a layer indication 18 indicating a layer the respective portion 16 belongs to. Each picture 12 is, for each of at least one layer out of a layer group 20, encoded into one or more portions 16 which belong to the respective layer, so that a corresponding version 12_i of the respective picture, which is associated with a respective layer i , is reconstructable from the
35 one or more portions 16, which belong to the respective layer i . However, other than the description brought forward above with respect to Figs. 3 and 4, inter-layer prediction is allowed. This is indicated by dashed lines 100 in Fig. 5. That is, encoder 80 uses inter-layer

prediction in order to encode each picture 12 of the video 82 into data stream. That is, in order to encode a picture 12 into one or more portions 16, which belong to a layer i , from which, accordingly, a version 12_i of picture 12 is reconstructable, the encoder 80 uses inter-layer prediction from one or more portions 16 of the data stream which belong to a lower layer $j < i$ which have the same picture 12 encoded therein, but from which a lower quality version 12_j is reconstructable, or inter-layer prediction from the corresponding lower quality version 12_j . As far as the motion-compensated prediction is concerned, encoder 80 may, for instance, act as depicted in Fig. 5 and as it was the case in Fig. 4: in encoding a certain picture $12'$ into data stream 84 with respect to a certain layer i , i.e., into corresponding one or more portions 16 of layer i from which a corresponding version $12'_i$ is reconstructable, the encoder 80 may use for the motion-compensated prediction a version of a reference picture $12''$ which of the same layer i , i.e., version $12''_i$. This is merely an example, however, but for ease of understanding, Fig. 5 illustrates this example. In forwarding data stream 84 to decoder 90, similar to the description of Fig. 3, brought forward above with respect to the example outlined with respect to the combination of Figs. 3 and 4, some portion 16 of the data stream might have been stripped off or left off. In case of the usage of inter-layer prediction, though, server 92 and client 94 would, for instance, obey the inter-layer dependencies and for a picture, for which one or more portions 16 of a layer i are present in data stream 14, all one or more portions 16 of the one or more layers, which have also this picture 12 encoded therein, are also present in data stream 14 in which the one or more portions having this picture encoded therein and belonging to layer i depend by way of the inter-layer prediction 100. Two examples for a stripping-off or leaving-off of some portion 16 from the initial data stream 84 to data stream 14 received by decoder 90 are presented in Fig. 5 by way of dashed lines and dashed-dotted lines.

The encoder 80 accompanies data stream 84 with a reference picture information 22 which information 22 is left inside in data stream 14. This picture information 22 indicates to the decoder 90, optionally, set 20. Most importantly, however, information 22 indicates to decoder 90 certain hints with respect to versions to be used for reference pictures. Again, when the encoder 80 encoded the video into portions 16, certain versions of reference pictures were used for the encoding, but these versions might be not available any longer in data stream 14 due to the described omission when transitioning from data stream 84 to data stream 14. Information 22 provides the decoder with hints or instructions as to which version of such reference picture should preferably be used instead, or may be used as a substitute. For instance, the information 22 indicates for one or more predetermined portions 16 which have, using motion-compensated prediction from a reference picture $12''$, a predetermined picture $12'$ encoded therein, and which belong to a predetermined layer, let's say layer 3, i.e., for

the one or more predetermined portions 16 from which version 12₃' is reconstructable, one or more of the following:

1) For instance, information 22 could indicate the actual version or the reference picture 12'' having been used for encoding, i.e., 12₃'' in the present example of Fig. 5. In other words, information 22 could indicate the layer, e.g. layer 3, with which the actually used version 12₃'' of the reference picture 12'' is associated, from a reconstruction of which the predetermined picture 12' has been encoded into the one or more predetermined portions 16 using motion-compensated prediction, which belong to the predetermined layer 3. Above, for example, syntax element 134 indicated that the reference picture level is the same level as that of the currently decoded portion 16 and that merely that layer is allowed to be used.

2) Additionally or alternatively, information 22 could indicate a subset of layers, e.g. layer 2 and layer 3, out of layer group 20, a reconstructable version of which layer for reference picture 12'' may be used for the motion-compensated prediction. This subset may or may not comprise the actually used version of the reference picture. In other words, the subset includes, or excludes, the layer with which the actually used version of the reference picture is associated. A version of any layer of the subset, other than the layer with which the actually used version of the reference picture is associated, represents an allowed or preferred substitute of the actually used version 12₃'' of the reference picture in decoding the one or more portions 16 for coding version 12₃'. For instance, this subset may indicate that version 12₂'' may be used as a substitute for the actually used version 12₃'' and accordingly, even in the example #1 of data stream 14, where the one or more portions 12 having version 12₃'' encoded therein are missing, the decoder 90 would be able to decode the data stream 14, namely by using version 12₂'' for the motion-compensated prediction for decoding the one or more portions 16 relating to version 12₃', such as by upsampling in case of layers 2 and 3 differing in sample resolution. Such a subset was indicated above at 130. There, this indication has been provided in a manner for each layer 132, and for a sequence of pictures. That is, for any portion 16 relating to a picture pertaining to that sequence of pictures, and belong to that layer, the subset 130 was used.

3) Even additionally or alternatively, information 22 may indicate a layer ranking, e.g. 3→2→1, indicating a preference ranking among the layers of layer group 20 for using the corresponding layer versions 12_i'' of the reference picture in decoding the one or more portions relating to version 12₃'. For instance, the preference could indicate that, preferably, the same layer version is to be used and, if this one is not available, the next lower version $i - 1$ and so forth. Such a ranking was indicated above at 130. There, this indication has been provided in a

manner for each layer 132, and for a sequence of pictures. That is, for any portion 16 relating to a picture pertaining to that sequence of pictures, and belong to that layer, the ranking 130 is used. In other words, the information 22 may indicate a layer ranking indicating a preference ranking among the layers of the group 20 of layers for using the versions of the reference picture associated with the layers for decoding the one or more predetermined portions 16 using motion-compensated prediction.

4) Additionally or alternatively, information 22 could indicate the pure allowance or disallowance whether the usage of any other layer's version than the version of the same layer as the one of the currently decoded portion 16 is allowed to be used as reference picture version. Flag 134 was an example for this. In other words, the information 22 could indicate whether any other layer than the predetermined layer, e.g. layer 3, is allowed to be used in decoding the one or more predetermined portions 16.

The indication 22 may, as outlined above, be signaled in the data stream 14 in a staggered manner: large scope signaling such as in a SPS or VPS may indicate subsets of substitute layers or a layer ranking for a sequence of pictures, and this signaling may be modified for individual pictures or portion 16 which may be slices in any of the embodiments presented herein, such as by signaling that, for that picture or slice, merely the currently decoded portion's 16 layer is allowed to be used.

The reference picture information 22, for example, is inserted into the data stream 14 or read from the data stream 14 individually for the predetermined picture and/or for a picture sequence including the predetermined picture.

The reference picture information 22, for example, is inserted into the data stream 14 in form of a layer specific indication describing for the predetermined layer as well as one or more further layers, the layer (e.g. 3), with which an actually used version (12₃) of the reference picture is associated, the subset of layers (e.g. 2+3) out of the group (20) of layers, the layer ranking and/or whether any other layer is allowed to be used.

As noted above, it may be that for pictures allowed to use alternative references TMVP is not used or the predictors used for deriving motion-vectors are the same in any version. Or in this case, alternative references are only used for sample prediction but all the rest (i.e. syntax or MV prediction) is done from a unique version, e.g. lowest or same layer_id. The encoder 80, for example, is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream signal using temporal motion vector prediction TMVP, and for

portions 16 for which any other layer's version, other than the layer, e.g. layer 3, the portions 16 belong to, is allowed to be used in decoding same, not use TMVP, or restrict TMVP to a derivation of motion vector predictors from one or more portions belonging to a default layer such as a lowest layer or base layer. This usage of TMVP can be signaled in the data stream

5 14.

According to an embodiment, the encoder 80 is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream 14 using an adaptive loop filter (ALF), and signal for the portions 16 associated with a predetermined layer of the group

10 20 of layers, which reference, for the motion-compensated prediction, pictures, e.g. 12'', for which portions 16, which belong to the predetermined layer of the group 20 of layers, are absent in the multi-layered video data stream 14, ALF parameters for use with respect to the predetermined layer and for use with at least one other layer.

15 Similarly to the encoder 80, the decoder 90 for decoding a multi-layered video data stream 14 having a video 82 encoded therein, for example, is configured to decode, using motion-compensated prediction 24, pictures 12 of the video 82 from portions 16 of the multi-layered video data stream 14. Each portion 16 has an associated picture 12 of the video 82, or a part thereof, encoded therein and comprises a layer indication 18 indicating a layer the respective

20 portion 16 belongs to. Furthermore, each picture is, for each of at least one out of the group 20 of layers, encoded into one or more portions 16, which belong to the respective layer i , so that a version 12_i of the respective picture, which is associated with the respective layer i , is reconstructible from the one or more portions 16, which belong to the respective layer i . Additionally, the decoder 90 is, for example, configured to read, from the multi-layered video

25 data stream 14, the reference picture information 22 which indicates for one or more predetermined portions 16 which have, using motion-compensated prediction from a reference picture 12'', a predetermined picture 12' encoded therein, and which belong to a predetermined layer, e.g. layer 3, for the one or more predetermined portions 16 from which version $12_3'$ is reconstructable one or more of the above described options 1 to 4, like the

30 actual layer, e.g. layer 3, associated with the used reference picture 12'', e.g. $12_3''$, the subset of layers, e.g. 2+3, out of the group 20 of layers, the layer ranking, e.g. $3 \rightarrow 2 \rightarrow 1$ and/or the allowance or disallowance of the usage of any other layer's version than the version of the predetermined layer. Additionally, the decoder 90 is configured to decode the predetermined picture, e.g. $12_3'$, at the predetermined layer, e.g. layer 3, from the multi-layered video data

35 stream 14 using motion compensated prediction from a version $12_i''$ of the reference picture selected depending on the reference picture information 22.

The decoder 90, for example, is configured to perform inter-layer quality adaptation in order to perform the motion compensated prediction from the version of the reference picture selected depending on the reference picture information 22 in case of said layer deviating from the predetermined layer.

5

The decoder 90 may be configured to decode the pictures 12 of the video 82 from the portions 16 of the multi-layered video data stream 14 using an adaptive loop filter (ALF), and derive from portions 16 associated with a predetermined layer of the group 20 of layers, which reference, for the motion-compensated prediction 24, pictures, e.g. 12", for which portions 16, which belong to the predetermined layer of the group 20 of layers, are absent in the multi-layered video data stream 14, ALF parameters for use with respect to the predetermined layer and for use with at least one other layer and filter the pictures decoded from the portions 16 associated with the predetermined layer of the group 20 of layers to decode, by the motion-compensated prediction, the portions 16 associated with the predetermined layer of the group of layers, using ALF parameters derived from the portions associated with the predetermined layer of the group of layers for the layer of the version selected depending on the reference picture information layer.

According to an embodiment, the encoder 80 is configured to adaptively encode the video 82 into the multi-layered video data stream 14. That is, the encoder directly forms the data stream with the selected/predetermined layer out of the layer group varying from one picture to another such as owing to transmission bandwidth situations. The encoder is aware of the change in selected layer and only codes the layer selected for each picture. If layers between reference picture and referencing picture deviate, the encoder does the same as the decoder will do, i.e. it uses a substitute reference picture for motion compensated prediction. The encoder 80, for example, is configured to encode, using motion-compensated prediction, pictures 12 of the video 82 into portions 16 of the multi-layered video data stream 14 so that each portion 16 has an associated picture 12 of the video, or a portion thereof, encoded therein and comprises a layer indication 18 indicating a layer the respective portion belongs to, and so that the portions into which the video is encoded belong to different layers, i.e. the multi-layered video data stream 14 comprises portions belonging to different layers, wherein portions associated with pictures encoded in the same quality belong to the same layer and portions associated with pictures encoded in a another quality belong to another layer. Furthermore, the encoder 80 is configured so that the multi-layered video data stream 14 comprises portions 16 associated with a predetermined layer of the group 20 of layers, which reference, for the motion-compensated prediction, pictures, e.g. 12", for which portions, which belong to the

predetermined layer of the group 20 of layers, are absent in the multi-layered video data stream 14.

According to the example of the multi-layered video data stream 14 shown in the dashed lines 62 in Fig. 4, the encoder 80 can use the picture $12_1'$ as a reference for the motion-compensated prediction of the picture $12_3''$, even though the picture $12_1'$ is associated with the layer 1 out of the group 20 of layers and the picture $12''$ is encoded into portions belonging to the predetermined layer being layer 3 of group 20 of layers. For picture $12'$ portions, which belong to the predetermined layer being layer 3 of the group 20 of layers, are absent in the multi-layered video data stream 14. Additionally, the encoder 80 is configured to insert, into the multi-layered video data stream 14, a layer grouping information 50 which groups the layers the portions 16 belong to, into which the video 82 is encoded, into the group 20 of layers.

The encoder 80, for example, is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream 14 so that each picture 12 has one layer associated therewith and is exclusively encoded into one or more portions 16 which belong to the one layer which is associated with the respective picture. Thus each picture is encoded only in one quality, as shown in Fig. 4 for the multi-layered video data stream 14 in the dashed lines 62. More than one picture can be associated with the same layer. As shown in Fig. 4, one or more portions 16 into which the picture $12_3''$ is encoded belong to the same layer as one or more portions 16 into which the picture $12_3'''$ is encoded.

According to an embodiment, the encoder 80 is configured to encode at least one or more of the portions 16 associated with the predetermined layer of the group 20 of layers, which reference, for the motion-compensated prediction, pictures for which portions 16, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream 14, by obtaining motion-compensated predictions for the one or more portions 16 using up-sampling and/or down-sampling from version of the pictures referenced by the one or more pictures which are present in the data stream. As shown in Fig. 5, for example, the picture $12_1'$ is up-sampled, as indicated by the arrow 100, to obtain a substitute of the version $12_3'$ of the picture $12'$ which is referenced, for the motion-compensated prediction, by picture $12_3''$ to be encoded into portions 16 belonging to layer 3 of the group 20 of layers.

According to an embodiment, the encoder 80 is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream 14 using temporal motion vector prediction (TMVP), and not use TMVP for encoding the pictures where the portions 16 associated with the predetermined layer of the group 20 of layers, which reference, for the

motion-compensated prediction, pictures for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream, are allowed to occur. Thus TMVP is, for example, not used at layer switching points, where a version of a picture associated with a first layer of the group 20 of layers references a version of another picture associated with a second layer of the group 20 of layers. TMVP should not be used, when a switching between exchangeable pictures is possible. TMVP is, for example, only used when pictures where the portions associated with the predetermined layer of the group of layers, for the motion-compensated prediction, reference pictures where the portions are also associated with the predetermined layer of the group of layers.

According to an embodiment, the encoder 80 is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream 14 using temporal motion vector prediction (TMVP), and not use the pictures which are referenced by pictures of the video associated with the predetermined layer of the group of layers and for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream, as a source for TMVP. Thus exchangeable pictures are not used as source for TMVP. A picture associated with a second layer of the group 20 of layers, for the motion-compensated prediction, referenced by a picture associated with a first layer of the group 20 of layers, for example, is not used as a source for TMVP.

According to an embodiment, the encoder 80 is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream 14 using decoder-side motion vector derivation (DMVR), and not use DMVR for encoding the pictures where the portions associated with the predetermined layer of the group of layers, which reference, for the motion-compensated prediction, pictures for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream, are allowed to occur. Thus DMVR is, for example, not used at layer switching points, where a version of a picture associated with a first layer of the group 20 of layers references a version of another picture associated with a second layer of the group 20 of layers. DMVR should not be used, when a switching between exchangeable pictures is possible. DMVR is, for example, only used when pictures where the portions associated with the predetermined layer of the group of layers, for the motion-compensated prediction, reference pictures where the portions are also associated with the predetermined layer of the group of layers.

According to an embodiment, the encoder 80 is configured to encode the pictures 12 of the video 82 into the portions 16 of the multi-layered video data stream 14 using decoder-side motion vector derivation (DMVR), and not use the pictures which are referenced by pictures of

the video associated with the predetermined layer of the group of layers and for which portions, which belong to the predetermined layer of the group of layers, are absent in the multi-layered video data stream, as a source for DMVR. Thus exchangeable pictures are not used as source for DMVR. A picture associated with a second layer of the group 20 of layers, for the motion-compensated prediction, referenced by a picture associated with a first layer of the group 20 of layers, for example, is not used as a source for DMVR.

Although some aspects have been described in the context of an apparatus, it is clear that these aspects also represent a description of the corresponding method, where a block or device corresponds to a method step or a feature of a method step. Analogously, aspects described in the context of a method step also represent a description of a corresponding block or item or feature of a corresponding apparatus. Some or all of the method steps may be executed by (or using) a hardware apparatus, like for example, a microprocessor, a programmable computer or an electronic circuit. In some embodiments, one or more of the most important method steps may be executed by such an apparatus.

Depending on certain implementation requirements, embodiments of the invention can be implemented in hardware or in software. The implementation can be performed using a digital storage medium, for example a floppy disk, a DVD, a Blu-Ray, a CD, a ROM, a PROM, an EPROM, an EEPROM or a FLASH memory, having electronically readable control signals stored thereon, which cooperate (or are capable of cooperating) with a programmable computer system such that the respective method is performed. Therefore, the digital storage medium may be computer readable.

Some embodiments according to the invention comprise a data carrier having electronically readable control signals, which are capable of cooperating with a programmable computer system, such that one of the methods described herein is performed.

Generally, embodiments of the present invention can be implemented as a computer program product with a program code, the program code being operative for performing one of the methods when the computer program product runs on a computer. The program code may for example be stored on a machine-readable carrier.

Other embodiments comprise the computer program for performing one of the methods described herein, stored on a machine-readable carrier.

In other words, an embodiment of the inventive method is, therefore, a computer program having a program code for performing one of the methods described herein, when the computer program runs on a computer.

A further embodiment of the inventive methods is, therefore, a data carrier (or a digital storage medium, or a computer-readable medium) comprising, recorded thereon, the computer program for performing one of the methods described herein. The data carrier, the digital storage medium or the recorded medium are typically tangible and/or non-transitionary.

A further embodiment of the inventive method is, therefore, a data stream or a sequence of signals representing the computer program for performing one of the methods described herein. The data stream or the sequence of signals may for example be configured to be transferred via a data communication connection, for example via the Internet.

A further embodiment comprises a processing means, for example a computer, or a programmable logic device, configured to or adapted to perform one of the methods described herein.

A further embodiment comprises a computer having installed thereon the computer program for performing one of the methods described herein.

A further embodiment according to the invention comprises an apparatus or a system configured to transfer (for example, electronically or optically) a computer program for performing one of the methods described herein to a receiver. The receiver may, for example, be a computer, a mobile device, a memory device or the like. The apparatus or system may, for example, comprise a file server for transferring the computer program to the receiver.

In some embodiments, a programmable logic device (for example a field programmable gate array) may be used to perform some or all of the functionalities of the methods described herein. In some embodiments, a field programmable gate array may cooperate with a microprocessor in order to perform one of the methods described herein. Generally, the methods are preferably performed by any hardware apparatus.

The apparatus described herein may be implemented using a hardware apparatus, or using a computer, or using a combination of a hardware apparatus and a computer.

The apparatus described herein, or any components of the apparatus described herein, may be implemented at least partially in hardware and/or in software.

The methods described herein may be performed using a hardware apparatus, or using a computer, or using a combination of a hardware apparatus and a computer.

The methods described herein, or any components of the apparatus described herein, may be performed at least partially by hardware and/or by software.

5 The above described embodiments are merely illustrative for the principles of the present invention. It is understood that modifications and variations of the arrangements and the details described herein will be apparent to others skilled in the art. It is the intent, therefore, to be limited only by the scope of the impending patent claims and not by the specific details presented by way of description and explanation of the embodiments herein.

Claims

1. Decoder (28) for decoding a multi-layered video data stream (32) which is partitioned into portions (16) each of which comprises a layer indication (18) indicating a layer (70) the respective portion (16) belongs to, configured to

read (44), from the multi-layered video data stream (32), a layer grouping information (40) which indicates a grouping of layers (70) which the portions (16) of the multi-layered video data stream (32) belong to, into one or more groups (72) of layers, a virtual layer (42) being associated with each group (72) of layers,

form (46) a video data stream (48) out of portions (16) of the multi-layered video data stream (32), which belong to the group (72) of layers associated with a predetermined virtual layer (50), by taking over the portions (16) from the multi-layered video data stream (32) into the video data stream (48) with leaving the layer indication (18) of the portions (16) unchanged,

decode (52) the video data stream (48) as a single layer data stream, the group of layers being coded independent from each other and relating to mutually different picture regions (54) of a video (6) represented by the single layer data stream.

2. Decoder of claim 1, configured to read from each portion (16), a relative position information (17) indicative of a picture position (117₁₋₃) of a picture part (116₁, 116₂) which is coded into the respective portion (16) in a manner so that the relative position information (17) is indicative of the picture position (117₁₋₃) relative to the picture region (54) the picture part (116₁, 116₂) is located in, and

from the multi-layered video data stream (48), a picture region arrangement information (110) indicating,

for each virtual layer (42), an output picture position of the region (54) within an output video composed of one or more regions (54) coded into the video data stream (48) portions (16) belonging to the group (72) of layers associated with the respective virtual layer (42), or

positions (114) of the picture regions (54) within the pictures (112) of the video (6).

3. Encoder for encoding a video (6), comprising

independently encoding mutually different picture regions (54) of the video (6) into video data stream portions (16) and embedding the video data stream portions (16) into a multi-layered video data stream (32) by providing each video data stream portion (16) with a layer indication (18) indicating a layer (70) the respective video data stream portion (16) belongs to, so that video data stream portions (16) having one picture region (54) of the different picture regions (54) encoded therein belong to a different layer (70) than video data stream portions (16) having another picture region (54) of the different picture regions (54) encoded therein,

providing the multi-layered video data stream (32) with a layer grouping information (40) which indicates a grouping of the layers (70) which the video data stream portions (16) belong to, into one or more groups (72) of layers, a virtual layer (42) being associated with each group (72) of layers.

4. Encoder of claim 3, configured to provide

each video data stream portion (16) with a relative position information (17) indicative of a picture position (117₁₋₃) of a picture part (116₁, 116₂) which is coded into the respective video data stream portion (16) in a manner so that the relative position information (17) is indicative of the picture position (117₁₋₃) relative to the picture region (54) the picture part (116₁, 116₂) is located in, and

the multi-layered video data stream (48) with a picture region arrangement information (110) indicating,

for each virtual layer (42), an output picture position of the region (54) within an output video composed of one or more regions (54) coded into the video data stream portions (16) belonging to the group (72) of layers associated with the respective virtual layer (42), or

positions (114) of the picture regions (54) within the pictures (112) of the video (6).

5. Encoder of claim 3, configured to obey a constraint in encoding the video (6) which includes one or more of the following

for each virtual layer (42), the group (72) of layers associated with the respective virtual layer (42) yields a video data stream (48) out of portions (16) belonging to the group (72) of layers associated with the respective virtual layer (42) which has an output video of output pictures (112) of rectangular output picture shape encoded thereinto,

for each virtual layer (42), the portions (16) of all layers (70) within the group (72) of layers associated with the respective virtual layer (42) have the same chroma format,

for each virtual layer (42), the portions (16) of all layers (70) within the group (72) of layers associated with the respective virtual layer (42) are encoded using a coding tree units (CTU) subdivision in CTUs of sizes which are equal among the layers (70) within the group (72) of layers associated with the respective virtual layer (42),

for each virtual layer (42), the portions (16) of all layers (70) within the group (72) of layers associated with the respective virtual layer (42) are encoded a POC to picture assignment to the pictures (12) which is the same among the layers (70) within the group (72) of layers associated with the respective virtual layer (42).

6. Multi-layered video data stream (32) comprising

video data stream portions (16) having mutually different picture regions (54) of the video (6) independently encoded thereinto, wherein each video data stream portion (16) comprises a layer indication (18) indicating a layer (70) the respective video data stream portion (16) belongs to, wherein video data stream portions (16) having one picture region (54) of the different picture regions (54) encoded thereinto belong to a different layer (70) than video data stream portions (16) having another picture region (54) of the different picture regions (54) encoded thereinto,

a layer grouping information (40) which indicates a grouping of the layers (70) which the video data stream portions (16) belong to, into one or more groups (72) of layers, a virtual layer (42) being associated with each group (72) of layers.

7. Multi-layered video data stream (32) of claim 6, wherein

each video data stream portion (16) is provided with a relative position information (17) indicative of a picture position (117₁₋₃) of a picture part (116₁, 116₂) which is coded into the respective video data stream portion (16) in a manner so that the relative position information (17) is indicative of the picture position (117₁₋₃) relative to the picture region (54) the picture part (116₁, 116₂) is located in, and

the multi-layered video data stream (48) further comprises a picture region arrangement information (110) indicating,

for each virtual layer (42), an output picture position of the regions (54) within an output video composed of one or more regions (54) coded into the video data stream portions (16) belonging to the group (72) of layers associated with the respective virtual layer (42), or

positions (114) of the picture regions (54) within the pictures (112) of the video (6).

8. Multi-layered video data stream (32) of claim 6 or 7, wherein

for each virtual layer (42), the group (72) of layers associated with the respective virtual layer (42) yields a video data stream (48) out of portions (16) belonging to the group (72) of layers associated with the respective virtual layer (42) which has an output video of output pictures (112) of rectangular output picture shape encoded thereinto, and/or

for each virtual layer (42), the portions (16) of all layers (70) within the group (72) of layers associated with the respective virtual layer (42) have the same chroma format, and/or

for each virtual layer (42), the portions (16) of all layers (70) within the group (72) of layers associated with the respective virtual layer (42) are encoded using a coding tree units (CTU) subdivision in CTUs of sizes which are equal among the layers (70) within the group (72) of layers associated with the respective virtual layer (42), and/or

for each virtual layer (42), the portions (16) of all layers (70) within the group (72) of layers associated with the respective virtual layer (42) are encoded a POC to picture assignment to the pictures (12) which is the same among the layers (70) within the group (72) of layers associated with the respective virtual layer (42).

9. Decoder (60) for decoding a multi-layered video data stream (14) which is partitioned into portions (16) each of which has an associated picture (12) of a video (82), or a part thereof, encoded thereinto and comprises a layer indication (18) indicating a layer the
5 respective portion (16) belongs to, configured to

read, from the multi-layered video data stream (14), a layer grouping information (50) which groups the layers the portions (16) belong to, into which the video (82) is coded, into a group (20) of layers,

10 gather portions (16) of the multi-layered video data stream (14), which belong to the group (20) of layers, so that each picture (12) of the video (82) is associated with one layer of the group (20) of layers, and for each picture (12) of the video (82), the one or more portions (16) belonging to the layer associated with the respective picture (12) are
15 gathered,

decode the video (82) from the gathered portions (16) by use of motion-compensated prediction (24) with supporting that pictures (12) decoded from portions (16) associated with a first layer of the group (20) of layers are referenced, for the motion-compensated
20 prediction (24), by portions (16) associated with a second layer of the group (20) of layers, different from the first layer.

10. Decoder of claim 9, configured to up-sample and/or down-sample the pictures (12) decoded from the portions (16) associated with the first layer of the group (20) of layers
25 to decode, by the motion-compensated prediction (24), the portions (16) associated with the second layer of the group (20) of layers.

11. Decoder of claim 9 or 10, configured to

30 read a predetermined layer group handling indication from the multi-layered video data stream (14),

if the predetermined layer group handling indication has a first state,

35 perform the gathering portions (16) of the multi-layered video data stream (14), which belong to the group (20) of layers, so that each picture (12) of the video (82) is associated with one layer of the group (20) of layers, and for each picture (12) of the

video (82), the one or more portions (16) belonging to the layer associated with the respective picture (12) are gathered and the decoding the video (82) from the gathered portions (16) by use of motion-compensated prediction (24) with supporting that pictures (12) decoded from portions (16) associated with a first layer of the group (20) of layers being referenced, for the motion-compensated prediction (24), by portions (16) associated with a second layer of the group (20) of layers, different from the first layer, and

if the predetermined layer group handling indication has a second state,

form a video data stream (48) out of portions (16) of the multi-layered video data stream (14), which belong to the group (72) of layers, by taking over the portions (16) into the video data stream (48) with leaving the layer indication (18) of the portions (16) unchanged so that each portion (16) of the multi-layered video data stream (14), which belongs to the group (72) of layers, is present in the video data stream (48) and

decode the video data stream (48) as a single layer data stream, the group (72) of layers being coded independent from each other and relating to mutually different picture regions (54) of a video (6) represented by the single layer data.

12. Decoder of any of claims 9 to 11, configured to

read from the multi-layered video data stream (14) a layer switching information and derive therefrom at which pictures (12) of the video (82) the portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur.

13. Decoder of claim 12, wherein

read from the multi-layered video data stream (14) the layer switching information and derive therefrom the pictures (12) of the video (82) where the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur,

for a sequence of pictures (12) as occurring at a regular pattern such as every n^{th} picture (12) with n being an integer, and/or

picture (12) individually.

5

14. Decoder of any of claims 9 to 13, configured to

decode the pictures (12) of the video (82) from the portions (16) of the multi-layered video data stream (14) using an adaptive loop filter (ALF), and

10

derive from the portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), ALF parameters for use with respect to the predetermined layer and for use with at least one other layer and filter the pictures (12) decoded from the portions (16) associated with the first layer of the group (20) of layers to decode, by the motion-compensated prediction (24), the portions (16) associated with the second layer of the group (20) of layers, using ALF parameters derived from the portions (16) associated with the second layer of the group (20) of layers for the first layer.

15

20

15. Apparatus for forming a multi-layered video data stream (14) having a video (82) encoded therein, configured to

25

encode, using motion-compensated prediction (24), pictures (12) of the video (82) into portions (16) of the multi-layered video data stream (14) so that each portion (16) has an associated picture (12) of the video (82), or a portion (16) thereof, encoded therein and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and so that the portions (16) into which the video (82) is encoded belong to different layers, wherein the apparatus is configured so that the multi-layered video data stream (14) comprises portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14); and

30

35

insert, into the multi-layered video data stream (14), a layer grouping information (50) which groups the layers the portions (16) belong to, into which the video (82) is encoded, into a group (20) of layers.

- 5 16. Apparatus of claim 15, configured to encode at least one or more of the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), by obtaining motion-compensated predictions (24) for
10 the one or more portions (16) using up-sampling and/or down-sampling from version (12₁₋₃) of the pictures (12) referenced by the one or more pictures (12) which are present in the data stream.
- 15 17. Apparatus of claim 15 or 16, configured to
encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream (14) so that each picture (12) has one layer associated therewith and is exclusively encoded into one or more portions (16) which belong to the one layer which is associated with the respective picture (12).
20
18. Apparatus of any of claims 15 to 17, configured to
signal in the multi-layered video data stream (14) a layer switching information which informs a decoder at which pictures (12) of the video (82) the portions (16) associated
25 with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur.
- 30 19. Apparatus of claim 18, configured to
signal in the multi-layered video data stream (14) the layer switching information which indicates the pictures (12) of the video (82) where the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the
35 predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur,

for a sequence of pictures (12) as occurring at a regular pattern such as every n^{th} picture (12) with n being an integer, and/or

picture (12) individually.

5

20. Apparatus of any of claims 18 to 19, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream (14) signal using temporal motion vector prediction (TMVP), and

10

not use TMVP for encoding the pictures (12) where the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur.

15

21. Apparatus of any of claims 18 to 20, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream (14) signal using decoder-side motion vector derivation (DMVR), and

20

not use DMVR for encoding the pictures (12) where the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur.

25

22. Apparatus of any of claims 15 to 21, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream signal using temporal motion vector prediction (TMVP), and

30

not use the pictures (12) which are referenced by pictures (12) of the video (82) associated with the predetermined layer of the group (20) of layers and for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), as a source for TMVP.

35

23. Apparatus of any of claims 15 to 22, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream signal using decoder-side motion vector derivation (DMVR), and

5

not use the pictures (12) which are referenced by pictures (12) of the video (82) associated with the predetermined layer of the group (20) of layers and for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), as a source for DMVR.

10

24. Apparatus of any of claims 15 to 23, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream signal using an adaptive loop filter (ALF), and

15

signal for the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), ALF parameters for use with respect to the predetermined layer and for use with at least one other layer.

20

25. Multi-layered video data stream (14) having a video (82) encoded thereinto,

wherein portions (16) of the multi-layered video data stream (14) have, using motion-compensated prediction (24), pictures (12) of the video (82) encoded thereinto so that each portion (16) has an associated picture (12) of the video (82), or a portion (16) thereof, encoded thereinto and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and so that the portions (16) into which the video (82) is encoded belong to different layers, wherein the multi-layered video data stream (14) comprises portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14); and

25

30

the multi-layered video data stream (14) comprising a layer grouping information (50) which groups the layers the portions (16) belong to, into which the video (82) is encoded, into a group (20) of layers.

35

26. Multi-layered video data stream (14) of claim 25, having at least one or more of the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), encoded by obtaining motion-compensated predictions (24) for the one or more portions (16) using up-sampling and/or down-sampling from version (12₁₋₃) of the pictures (12) referenced by the one or more pictures (12) which are present in the data stream.

27. Multi-layered video data stream (14) of claim 25 or 26, wherein

each picture (12) has one layer associated therewith and is exclusively encoded into one or more portions (16) which belong to the one layer which is associated with the respective picture (12).

28. Multi-layered video data stream (14) of any of claims 25 to 27, comprising

a layer switching information which informs a decoder at which pictures (12) of the video (82) the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur.

29. Multi-layered video data stream (14) of claim 28, wherein

the layer switching information indicates the pictures (12) of the video (82) where the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are allowed to occur,

for a sequence of pictures (12) as occurring at a regular pattern such as every n^{th} picture (12) with n being an integer, and/or

picture (12) individually.

30. Multi-layered video data stream (14) of any of claims 28 to 29,

wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream (14) signal using temporal motion vector prediction (TMVP), and

wherein the pictures (12) which are referenced by pictures (12) of the video (82) associated with the predetermined layer of the group (20) of layers and for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are encoded without TMVP.

31. Multi-layered video data stream (14) of any of claims 28 to 30,

wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream (14) signal using decoder-side motion vector derivation (DMVR), and

wherein the pictures (12) which are referenced by pictures (12) of the video (82) associated with the predetermined layer of the group (20) of layers and for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are encoded without DMVR.

32. Multi-layered video data stream (14) of any of claims 25 to 31,

wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream (14) signal using temporal motion vector prediction (TMVP), and

wherein the pictures (12) which are referenced by pictures (12) of the video (82) associated with the predetermined layer of the group (20) of layers and for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are not used as a source for TMVP.

33. Multi-layered video data stream (14) of any of claims 25 to 32,

wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream (14) signal using decoder-side motion vector derivation (DMVR), and

wherein the pictures (12) which are referenced by pictures (12) of the video (82) associated with the predetermined layer of the group (20) of layers and for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are not used as a source for DMVR.

34. Multi-layered video data stream (14) of any of claims 25 to 33,

wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream (14) signal using an adaptive loop filter (ALF), and

comprising, for the portions (16) associated with the predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), ALF parameters for use with respect to the predetermined layer and for use with at least one other layer.

35. Decoder for decoding a multi-layered video data stream (14) having a video (82) encoded therein, configured to

decode, using motion-compensated prediction (24), pictures (12) of the video (82) from portions (16) of the multi-layered video data stream (14)

wherein each portion (16) has an associated picture (12) of the video (82), or a part thereof, encoded therein and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and

wherein each picture (12) is, for each of at least one out of a group (20) of layers, encoded into one or more portions (16), which belong to the respective layer, so that a version (12₁₋₃) of the respective picture (12), which is associated with the respective layer, is reconstructible from the one or more portions (16), which belong to the respective layer, and

read, from the multi-layered video data stream (14), a reference picture information (22) which indicates

for one or more predetermined portions (16) which have, using motion-compensated prediction (24) from a reference picture, a predetermined picture encoded therein, and which belong to a predetermined layer, one or more of

the layer, with which an actually used version (12₁₋₃) of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions (16) using motion-compensated prediction (24), and

a subset of layers out of the group (20) of layers which includes, or excludes, the layer with which the actually used version (12₁₋₃) of the reference picture is associated, wherein a version (12₁₋₃) of any layer of the subset, other than the layer with which the actually used version (12₁₋₃) of the reference picture is associated, represents an allowed or preferred substitute of the actually used version (12₁₋₃) of the reference picture in decoding the one or more predetermined portions (16); and

a layer ranking indicating a preference ranking among the layers of the group (20) of layers for using the versions (12₁₋₃) of the reference picture associated with the layers for decoding the one or more predetermined portions (16) using motion-compensated prediction (24), and

whether any other layer's version (12₁₋₃) of the reference picture other than the predetermined layer is allowed to be used in decoding the one or more predetermined portions (16),

decode the predetermined picture at the predetermined layer from the multi-layered video data stream (14) using motion compensated prediction from a version (12₁₋₃) of the reference picture selected depending on the reference picture information.

36. Decoder of claim 35, configured to perform inter-layer quality adaptation in order to perform the motion compensated prediction from the version (12₁₋₃) of the reference picture selected depending on the reference picture information in case of said layer deviating from the predetermined layer.

37. Decoder of claim 35 or 36, configured to read the reference picture information (22) individually for the predetermined picture and/or for a picture sequence including the predetermined picture.

38. Decoder of claim any of claims 35 to 37, configured to read the reference picture information (22) in form of a layer specific indication describing for the predetermined layer as well as one or more further layers, the layer, with which an actually used version (12₁₋₃) of the

reference picture is associated, the subset of layers out of the group (20) of layers, the layer ranking and/or whether any other layer is allowed to be used.

39. Decoder of any of claims 35 to 38, configured to

decode the pictures (12) of the video (82) from the portions (16) of the multi-layered video data stream (14) signal using an adaptive loop filter (ALF), and

derive from portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), ALF parameters for use with respect to the predetermined layer and for use with at least one other layer and filter the pictures (12) decoded from the portions (16) associated with the predetermined layer of the group (20) of layers to decode, by the motion-compensated prediction (24), the portions (16) associated with the predetermined layer of the group (20) of layers, using ALF parameters derived from the portions (16) associated with the predetermined layer of the group (20) of layers for the layer of the version (12₁₋₃) selected depending on the reference picture information layer.

40. Apparatus for forming a multi-layered video data stream (14) having a video (82) encoded therein, configured to

encode, using motion-compensated prediction (24), pictures (12) of a video (82) into portions (16) of the multi-layered video data stream (14)

so that each portion (16) has an associated picture (12) of the video (82), or a part thereof, encoded therein and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and

so that each picture (12) is, for each of at least one out of a group (20) of layers, encoded into one or more portions (16), which belong to the respective layer, so that a version (12₁₋₃) of the respective picture (12), which is associated with the respective layer, is reconstructible from the one or more portions (16), which belong to the respective layer, and

insert, into the multi-layered video data stream (14), a reference picture information (22) which indicates

for one or more predetermined portions (16) which have, using motion-compensated prediction (24) from a reference picture, a predetermined picture encoded thereinto, and which belong to a predetermined layer, one or more of

the layer, with which an actually used version (12₁₋₃) of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions (16) using motion-compensated prediction (24), and

a subset of layers out of the group (20) of layers which includes, or excludes, the layer with which the actually used version (12₁₋₃) of the reference picture is associated, wherein a version (12₁₋₃) of any layer of the subset, other than the layer with which the actually used version (12₁₋₃) of the reference picture is associated, represents an allowed or preferred substitute of the actually used version (12₁₋₃) of the reference picture in decoding the one or more predetermined portions (16); and

a layer ranking indicating a preference ranking among the layers of the group (20) of layers for using the versions (12₁₋₃) of the reference picture associated with the layers for decoding the one or more predetermined portions (16) using motion-compensated prediction (24), and

whether any other layer than the predetermined layer (e.g. 3) is allowed to be used in decoding the one or more predetermined portions (16).

41. Encoder of claim 40, configured to insert the reference picture information (22) individually for the predetermined picture and/or for a picture sequence including the predetermined picture.

42. Encoder of claim 40 or 41, configured to insert the reference picture information (22) in form of a layer specific indication describing for the predetermined layer as well as one or more further layers, the layer, with which an actually used version (12₁₋₃) of the reference picture is associated, the subset of layers out of the group (20) of layers, the layer ranking and/or whether any other layer is allowed to be used.

43. Encoder of any of claims 40 to 42, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream (14) signal using temporal motion vector prediction (TMVP), and

for portions (16) for which any other layer's version (12₁₋₃), other than the layer the portions (16) belong to, is allowed to be used in decoding same,

5 not use TMVP, or

restrict TMVP to a derivation of motion vector predictors from one or more portions (16) belonging to a default layer such as a lowest layer.

10 44. Encoder of any of claims 40 to 43, configured to

encode the pictures (12) of the video (82) into the portions (16) of the multi-layered video data stream signal using an adaptive loop filter (ALF), and

15 provide portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), ALF parameters for use with respect to the predetermined layer and for use with at least one other layer.

20

45. A multi-layered video data stream (14) having a video (82) encoded therein,

wherein, using motion-compensated prediction (24), pictures (12) of a video (82) are encoded into portions (16) of the multi-layered video data stream (14)

25 so that each portion (16) has an associated picture (12) of the video (82), or a part thereof, encoded therein and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and

so that each picture (12) is, for each of at least one out of a group (20) of layers, encoded into one or more portions (16), which belong to the respective layer,

30 so that a version (12₁₋₃) of the respective picture (12), which is associated with the respective layer, is reconstructible from the one or more portions (16), which belong to the respective layer, and

the multi-layered video data stream (14) comprises a reference picture information (22)
35 which indicates

for one or more predetermined portions (16) which have, using motion-compensated prediction (24) from a reference picture, a predetermined picture encoded thereinto, and which belong to a predetermined layer, one or more of

the layer, with which an actually used version (12₁₋₃) of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions (16) using motion-compensated prediction (24), and

a subset of layers out of the group (20) of layers which includes, or excludes, the layer with which the actually used version (12₁₋₃) of the reference picture is associated, wherein a version (12₁₋₃) of any layer of the subset, other than the layer with which the actually used version (12₁₋₃) of the reference picture is associated, represents an allowed or preferred substitute of the actually used version (12₁₋₃) of the reference picture in decoding the one or more predetermined portions (16); and

a layer ranking indicating a preference ranking among the layers of the group (20) of layers for using the versions (12₁₋₃) of the reference picture associated with the layers for decoding the one or more predetermined portions (16) using motion-compensated prediction (24),

whether any other layer than the predetermined layer is allowed to be used in decoding the one or more predetermined portions (16).

46. Multi-layered video data stream (14) of claim 45, having the reference picture information (22) inserted individually for the predetermined picture and/or for a picture sequence including the predetermined picture.

47. Multi-layered video data stream (14) of claim 45 or 46, having the reference picture information (22) inserted in form of a layer specific indication describing for the predetermined layer as well as one or more further layers, the layer, with which an actually used version (12₁₋₃) of the reference picture is associated, the subset of layers out of the group (20) of layers, the layer ranking and/or whether any other layer is allowed to be used.

48. Multi-layered video data stream (14) of any of claims 45 to 47,

wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream (14) signal using temporal motion vector prediction (TMVP), and

- 5 for portions (16) for which any other layer's version (12₁₋₃), other than the layer the portions (16) belong to, is allowed to be used in decoding same,

TMVP is not used, or

- 10 TMVP is limited to a derivation of motion vector predictors from one or more portions (16) belonging to a default layer such as a lowest layer.

49. Multi-layered video data stream (14) of any of claims 45 to 48,

- 15 wherein the pictures (12) of the video (82) are encoded into the portions (16) of the multi-layered video data stream signal using an adaptive loop filter (ALF), and

- portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions
20 (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14), are provided with ALF parameters for use with respect to the predetermined layer and for use with at least one other layer.

50. Method for decoding a multi-layered video data stream (32) which is partitioned into
25 portions (16) each of which comprises a layer indication (18) indicating a layer (70) the respective portion (16) belongs to, comprising

- reading (44), from the multi-layered video data stream (32), a layer grouping information (40) which indicates a grouping of layers (70) which the portions (16) of the multi-layered
30 video data stream (32) belong to, into one or more groups (72) of layers, a virtual layer (42) being associated with each group (72) of layers,

- forming (46) a video data stream (48) out of portions (16) of the multi-layered video data stream (32), which belong to the group (72) of layers associated with a predetermined
35 virtual layer (50), by taking over the portions (16) from the multi-layered video data stream (32) into the video data stream (48) with leaving the layer indication (18) of the portions (16) unchanged, and

decoding (52) the video data stream (48) as a single layer data stream, the group of layers being coded independent from each other and relating to mutually different picture regions (54) of a video (6) represented by the single layer data stream.

5

51. Method for encoding a video (6), comprising

independently encoding mutually different picture regions (54) of the video (6) into video data stream portions (16) and embedding the video data stream portions (16) into a multi-layered video data stream (32) by providing each video data stream portion (16) with a layer indication (18) indicating a layer (70) the respective video data stream portion (16) belongs to, so that video data stream portions (16) having one picture region (54) of the different picture regions (54) encoded thereinto belong to a different layer (70) than video data stream portions (16) having another picture region (54) of the different picture regions (54) encoded thereinto; and

15

providing the multi-layered video data stream (32) with a layer grouping information (40) which indicates a grouping of the layers (70) which the video data stream portions (16) belong to, into one or more groups (72) of layers, a virtual layer (42) being associated with each group (72) of layers.

20

52. Method for decoding a multi-layered video data stream (14) which is partitioned into portions (16) each of which has an associated picture (12) of a video (82), or a part thereof, encoded thereinto and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, comprising

25

reading, from the multi-layered video data stream (14), a layer grouping information (50) which groups the layers the portions (16) belong to, into which the video (82) is coded, into a group (20) of layers,

30

gathering portions (16) of the multi-layered video data stream (14), which belong to the group (20) of layers, so that each picture (12) of the video (82) is associated with one layer of the group (20) of layers, and for each picture (12) of the video (82), the one or more portions (16) belonging to the layer associated with the respective picture (12) are gathered, and

35

decoding the video (82) from the gathered portions (16) by use of motion-compensated prediction (24) with supporting that pictures (12) decoded from portions (16) associated with a first layer of the group (20) of layers being referenced, for the motion-compensated prediction (24), by portions (16) associated with a second layer of the group (20) of layers, different from the first layer.

53. Method for forming a multi-layered video data stream (14) having a video (82) encoded therein, comprising

encoding, using motion-compensated prediction (24), pictures (12) of the video (82) into portions (16) of the multi-layered video data stream (14) so that each portion (16) has an associated picture (12) of the video (82), or a portion (16) thereof, encoded therein and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and so that the portions (16) into which the video (82) is encoded belong to different layers, and so that the multi-layered video data stream (14) comprises portions (16) associated with a predetermined layer of the group (20) of layers, which reference, for the motion-compensated prediction (24), pictures (12) for which portions (16), which belong to the predetermined layer of the group (20) of layers, are absent in the multi-layered video data stream (14); and

insert, into the multi-layered video data stream (14), a layer grouping information (50) which groups the layers the portions (16) belong to, into which the video (82) is encoded, into a group (20) of layers.

54. Method for decoding a multi-layered video data stream (14) having a video (82) encoded therein, comprising

decoding, using motion-compensated prediction (24), pictures (12) of the video (82) from portions (16) of the multi-layered video data stream (14)

wherein each portion (16) has an associated picture (12) of the video (82), or a part thereof, encoded therein and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and

wherein each picture (12) is, for each of at least one out of a group (20) of layers, encoded into one or more portions (16), which belong to the respective layer, so that a version (12₁₋₃) of the respective picture (12), which is associated with the respective layer, is reconstructible from the one or more portions (16), which belong to the respective layer, and

reading, from the multi-layered video data stream (14), a reference picture information (22) which indicates

for one or more predetermined portions (16) which have, using motion-compensated prediction (24) from a reference picture, a predetermined picture encoded thereinto, and which belong to a predetermined layer, one or more of

the layer, with which an actually used version (12_{1-3}) of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions (16) using motion-compensated prediction (24), and

a subset of layers out of the group (20) of layers which includes, or excludes, the layer with which the actually used version (12_{1-3}) of the reference picture is associated, wherein a version (12_{1-3}) of any layer of the subset, other than the layer with which the actually used version (12_{1-3}) of the reference picture is associated, represents an allowed or preferred substitute of the actually used version (12_{1-3}) of the reference picture in decoding the one or more predetermined portions (16); and

a layer ranking indicating a preference ranking among the layers of the group (20) of layers for using the versions (12_{1-3}) of the reference picture associated with the layers for decoding the one or more predetermined portions (16) using motion-compensated prediction (24), and

whether any other layer's version (12_{1-3}) of the reference picture other than the predetermined layer is allowed to be used in decoding the one or more predetermined portions (16),

decoding the predetermined picture at the predetermined layer from the multi-layered video data stream (14) using motion compensated prediction from a version (12_{1-3}) of the reference picture selected depending on the reference picture information.

55. Method for forming a multi-layered video data stream (14) having a video (82) encoded thereinto, comprising

encoding, using motion-compensated prediction (24), pictures (12) of a video (82) into portions (16) of the multi-layered video data stream (14)

so that each portion (16) has an associated picture (12) of the video (82), or a part thereof, encoded thereinto and comprises a layer indication (18) indicating a layer the respective portion (16) belongs to, and

so that each picture (12) is, for each of at least one out of a group (20) of layers, encoded into one or more portions (16), which belong to the respective layer, so that a version (12₁₋₃) of the respective picture (12), which is associated with the respective layer, is reconstructible from the one or more portions (16), which belong to the respective layer, and

inserting, into the multi-layered video data stream (14), a reference picture information (22) which indicates

for one or more predetermined portions (16) which have, using motion-compensated prediction (24) from a reference picture, a predetermined picture encoded thereinto, and which belong to a predetermined layer, one or more of

the layer, with which an actually used version (12₁₋₃) of the reference picture is associated, from a reconstruction of which the predetermined picture is encoded into the one or more predetermined portions (16) using motion-compensated prediction (24), and

a subset of layers out of the group (20) of layers which includes, or excludes, the layer with which the actually used version (12₁₋₃) of the reference picture is associated, wherein a version (12₁₋₃) of any layer of the subset, other than the layer with which the actually used version (12₁₋₃) of the reference picture is associated, represents an allowed or preferred substitute of the actually used version (12₁₋₃) of the reference picture in decoding the one or more predetermined portions (16); and

a layer ranking indicating a preference ranking among the layers of the group (20) of layers for using the versions (12₁₋₃) of the reference picture associated with the layers for decoding the one or more predetermined portions (16) using motion-compensated prediction (24), and

whether any other layer than the predetermined layer (e.g. 3) is allowed to be used in decoding the one or more predetermined portions (16).

56. Computer program having a program code for performing, when running on a computer, a method of any of claims 50 to 55.

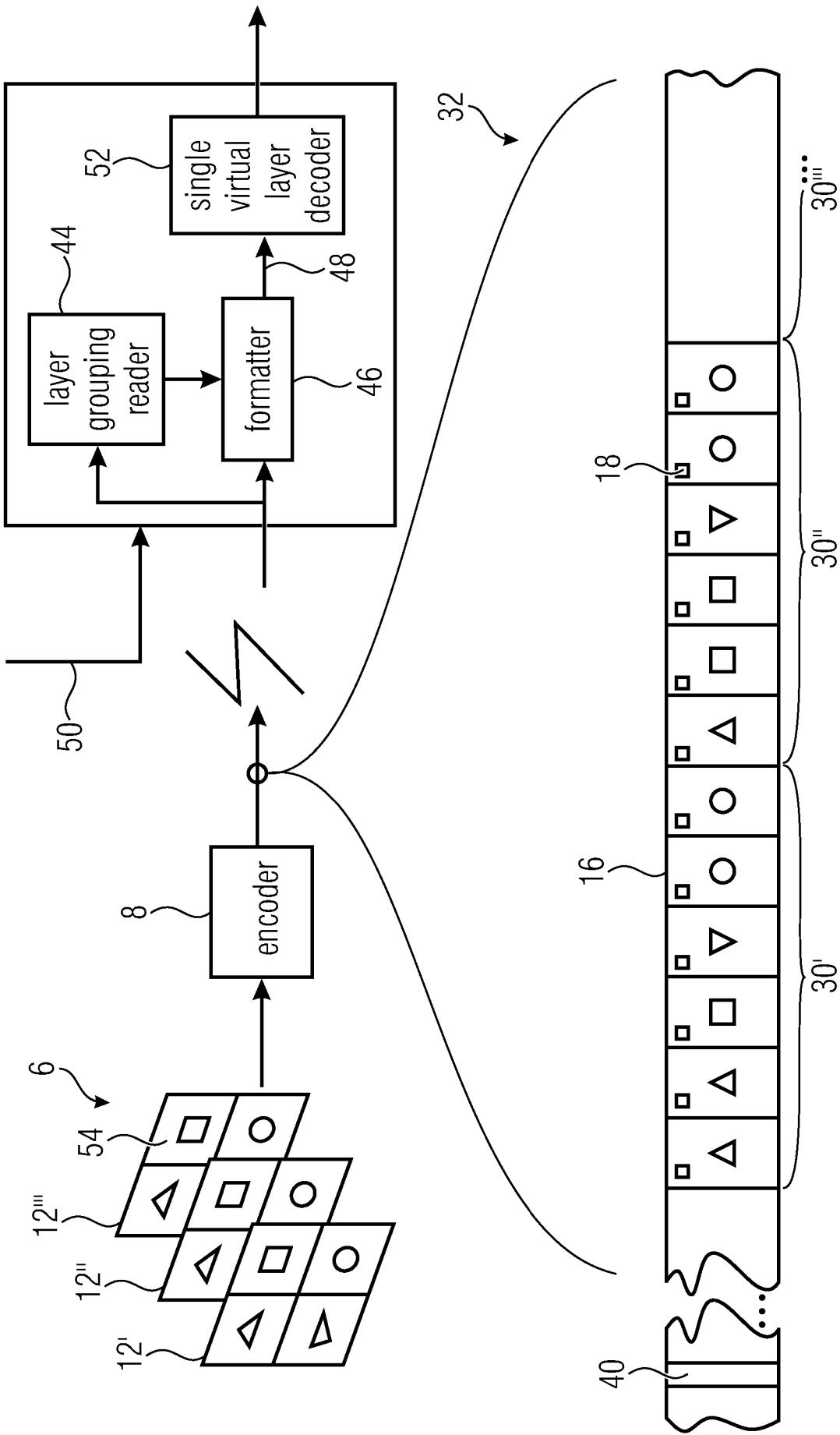


Fig. 1

	layer 10	virtual layer 10(s)	
Δ	1	1	2
▽	2	1	2
□	3	1	
○	4	1	

42

40

70

72'

72''

Fig. 2

3/8

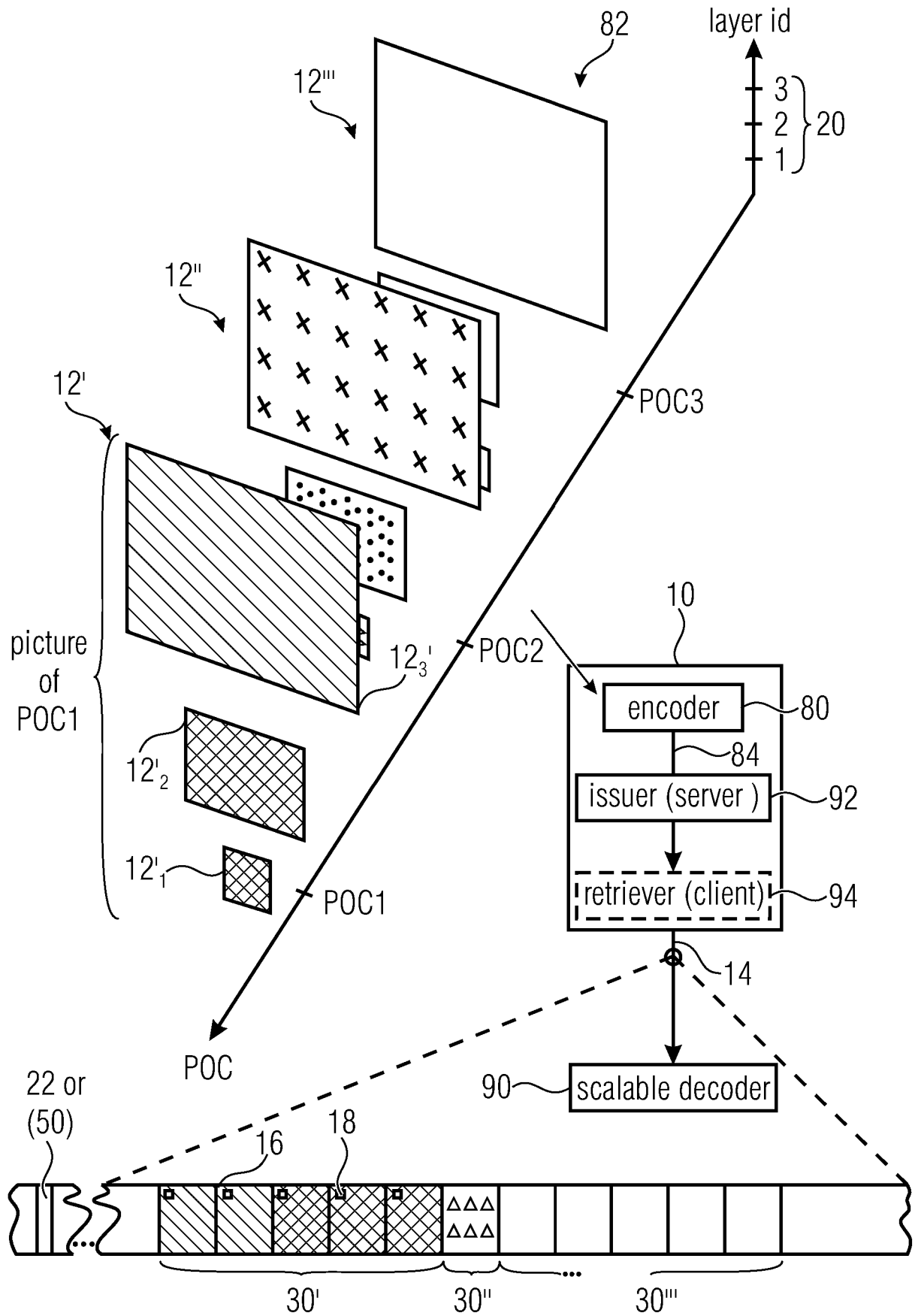


Fig. 3

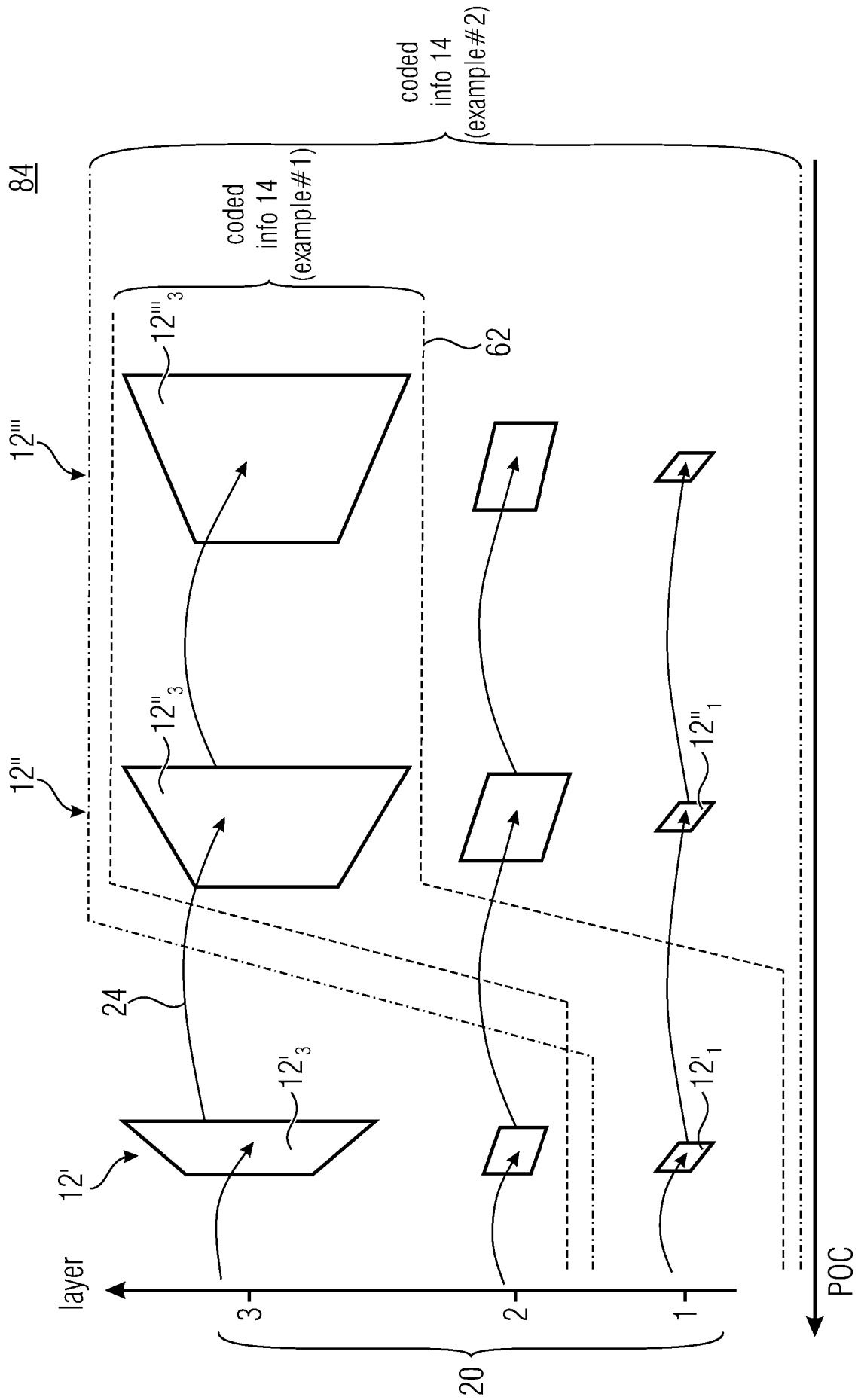


Fig. 4

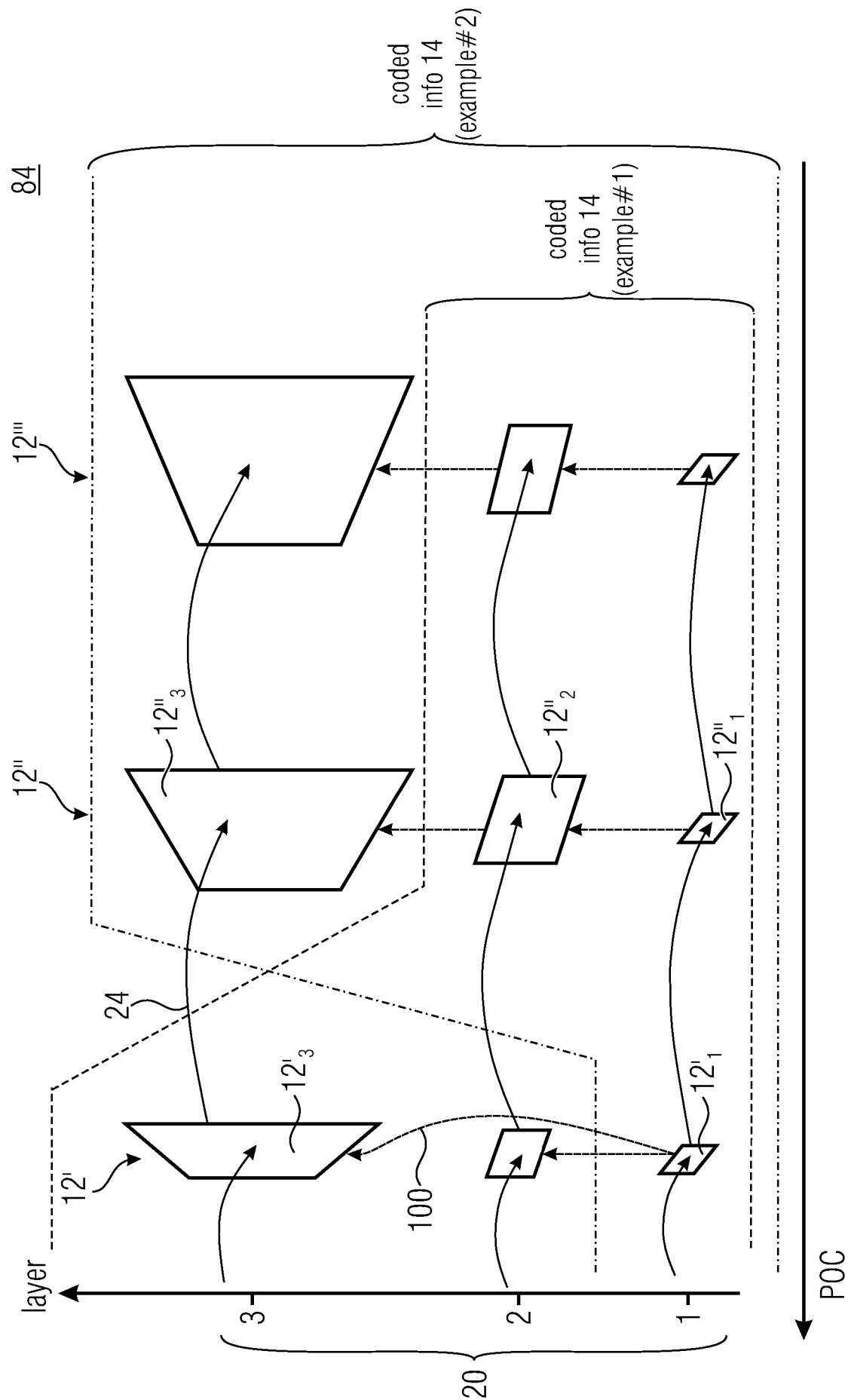


Fig. 5

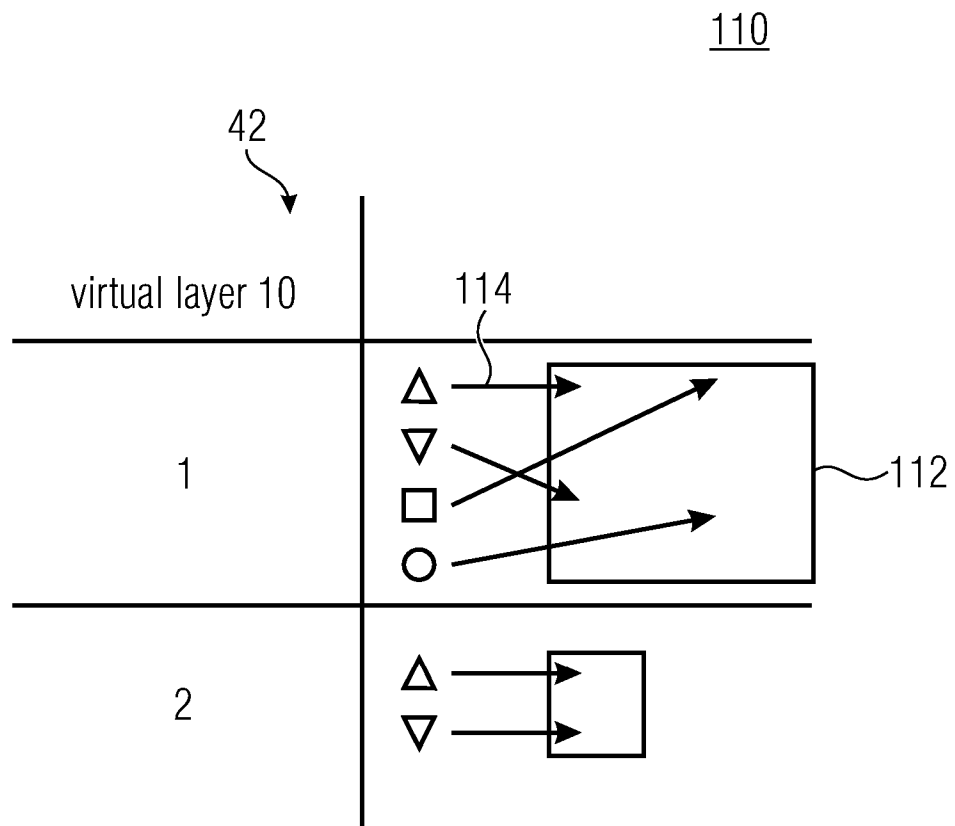


Fig. 6a

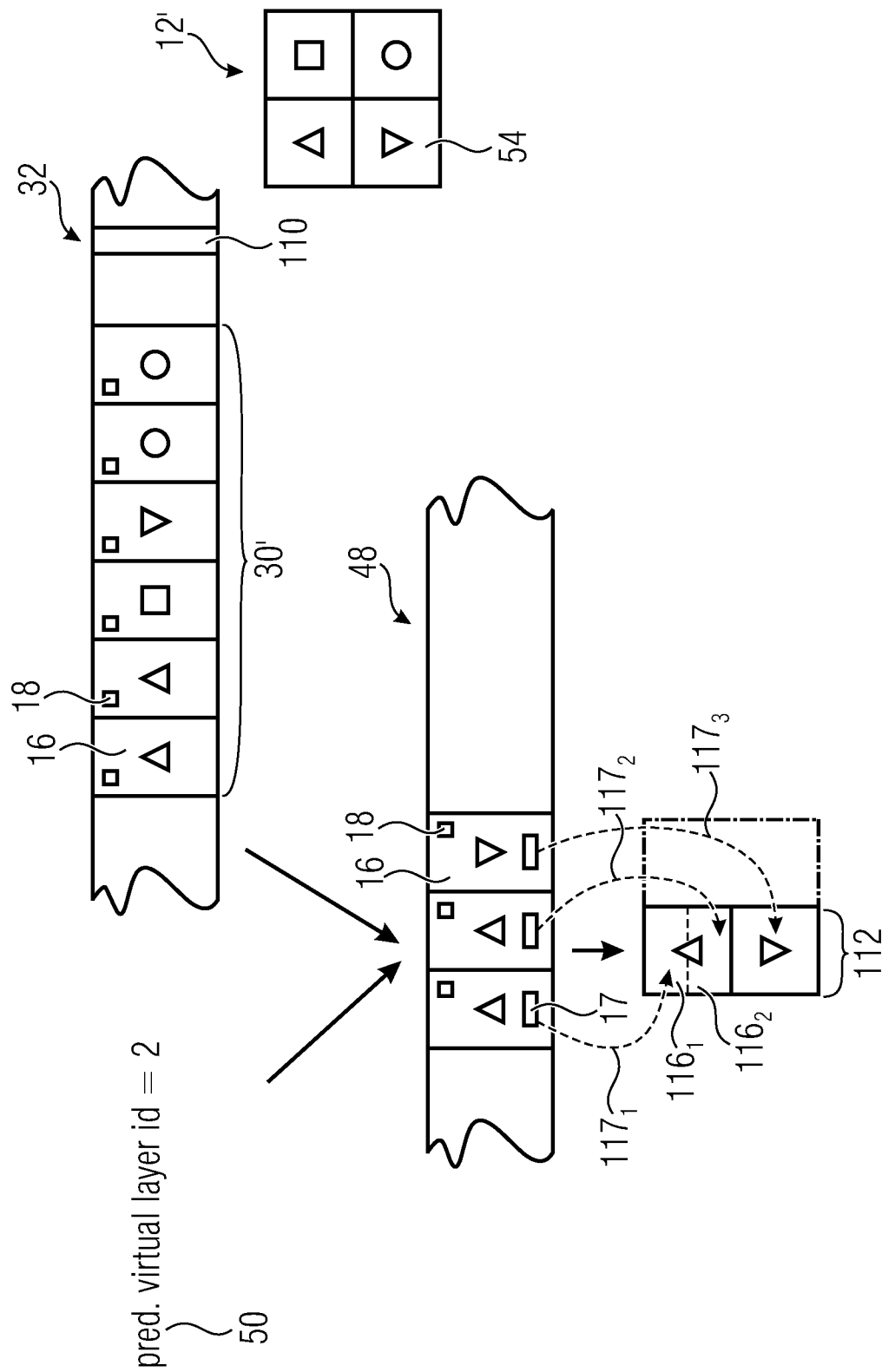


Fig. 6b

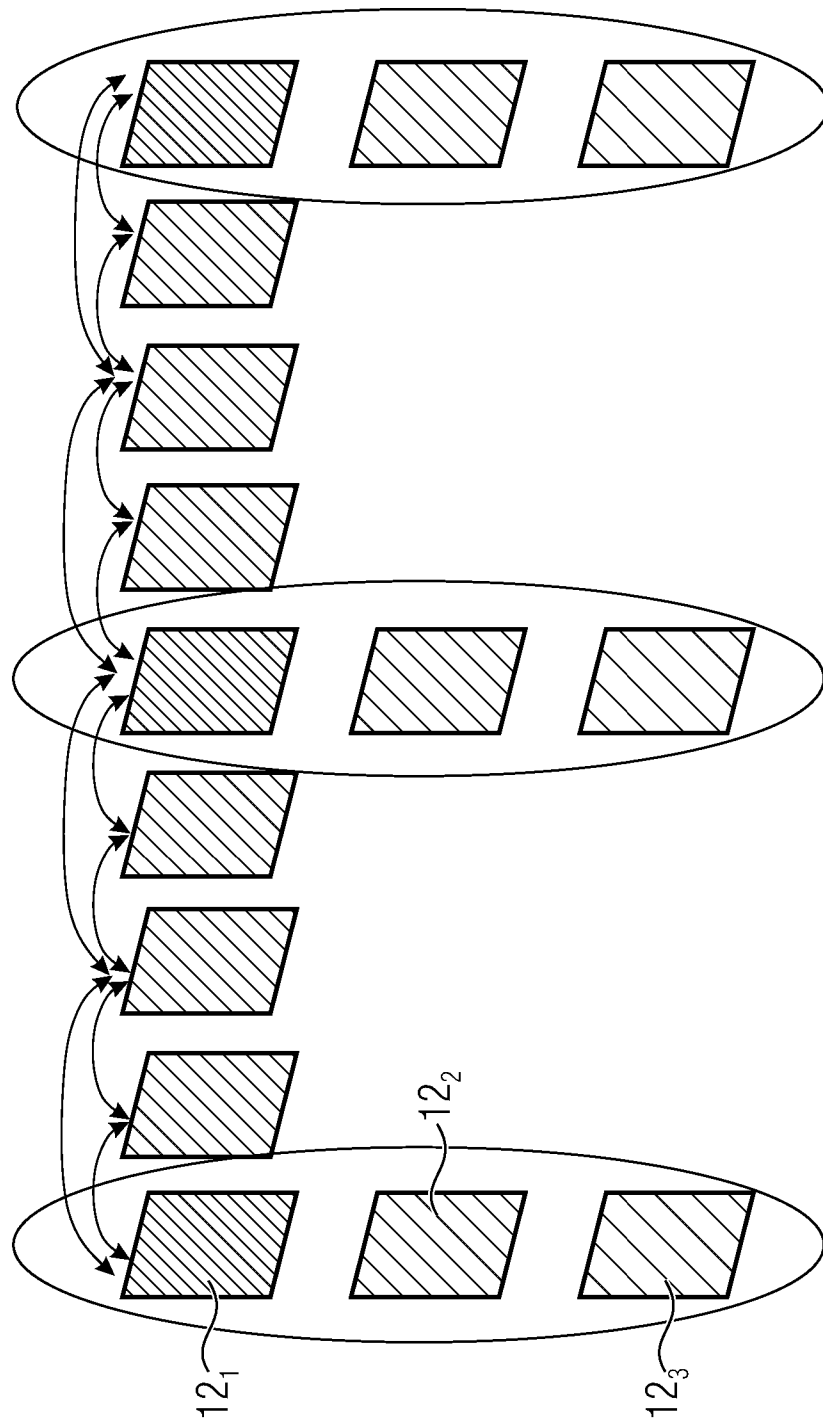


Fig. 7

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/065681

A. CLASSIFICATION OF SUBJECT MATTER INV. H04N19/30 H04N19/70 H04N19/46 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) H04N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	H. SCHWARZ ET AL: "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, vol. 17, no. 9, 1 September 2007 (2007-09-01), pages 1103-1120, XP055378169, US ISSN: 1051-8215, DOI: 10.1109/TCSVT.2007.905532 sections I-II, V figures 1, 4 ----- -/--	1-8,50, 51,56
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 7 September 2020		Date of mailing of the international search report 11/11/2020
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Mayer, Claudia

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2020/065681

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MASAYUKI INOUE ET AL: "Interactive panoramic video streaming system over restricted bandwidth network", PROCEEDINGS OF THE ACM MULTIMEDIA 2010 INTERNATIONAL CONFERENCE : ACM MM'10 & CO-LOCATED WORKSHOPS ; OCTOBER 25 - 29, FIRENZE, ITALY, ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, NY, USA, 25 October 2010 (2010-10-25), pages 1191-1194, XP058390276, DOI: 10.1145/1873951.1874184 ISBN: 978-1-60558-933-6 sections 2 and 3.1 abstract figures 3-6 -----	1-8,50, 51,56
X	SÁNCHEZ DE LA FUENTE Y ET AL: "Video processing for panoramic streaming using HEVC and its scalable extensions", MULTIMEDIA TOOLS AND APPLICATIONS, KLUWER ACADEMIC PUBLISHERS, BOSTON, US, vol. 76, no. 4, 1 December 2016 (2016-12-01), pages 5631-5659, XP036179119, ISSN: 1380-7501, DOI: 10.1007/S11042-016-4097-4 [retrieved on 2016-12-01] section 2-4 figure 3 abstract -----	1,3,6, 50,51,56

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2020/065681

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-8, 50, 51(completely); 56(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-8, 50, 51(completely); 56(partially)

Enabling decoding of a group of mutually different video picture regions as a single layer data stream.

2. claims: 9-34, 52, 53(completely); 56(partially)

Motion-compensated prediction using a different spatial/quality layer of the reference picture at the decoder than the spatial/quality layer referenced by the encoder.

3. claims: 35-49, 54, 55(completely); 56(partially)

Signaling reference picture information indicating which spatial/quality layer or layers of a reference picture are allowed to be used for motion-compensated prediction at a decoder.
