(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 March 2014 (06.03.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/033255 A1

(51) International Patent Classification:
H04N 7/26 (2006.01) *H04N 7/36* (2006.01)

(21) International Application Number:
PCT/EP2013/067983

(22) International Filing Date:
30 August 2013 (30.08.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1215430.8 30 August 2012 (30.08.2012) GB

(71) Applicant: **CANON KABUSHIKI KAISHA** [JP/JP]; 3-30-2, Shimomaruko, 3-chome, Ohta-ku, Tokyo (JP).

(72) Inventors; and
(71) Applicants *(for US only)*: **LE LEANNEC, Fabrice** [FR/FR]; 4, rue des Magnolias, Mouaze, 35250 (FR). **LASSERRE, Sébastien** [FR/FR]; 11 bd Saint Conwoïon, Rennes, 35000 (FR).

(74) Agent: **SANTARELLI**; P.O.Box 237, 14 avenue de la Grande-Armée, F-75822 Paris Cedex 17 (FR).

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

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

Published:
— *with international search report (Art. 21(3))*

(54) **Title**: METHOD AND DEVICE FOR DETETERMINING PREDICTION INFORMATION FOR ENCODING OR DECODING AT LEAST PART OF AN IMAGE

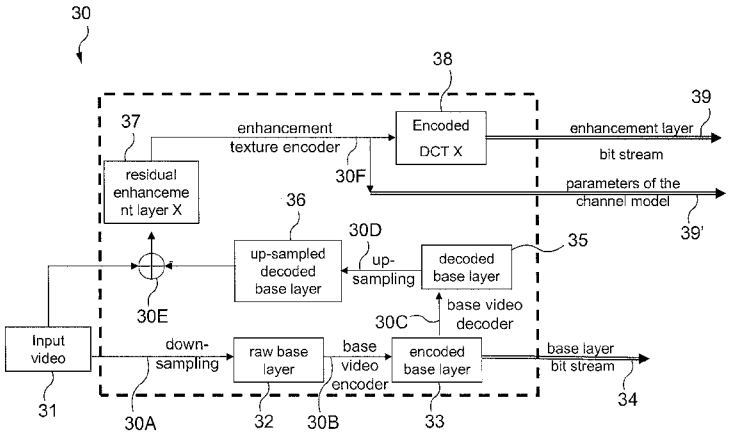


Fig. 3

(57) **Abstract**: A method of determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the step of deriving enhancement layer prediction information comprising for a processing block: in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, deriving the enhancement prediction information from the base layer prediction information of said one elementary prediction unit; otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, deriving the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected according to a predetermined criterion; wherein the predetermined criterion is based on at least one of the relative location of said one of said plurality of elementary prediction units with respect to the other elementary prediction units of said plurality of elementary prediction units and the base layer prediction information of the said elementary prediction units.

WO 2014/033255 A1

METHOD AND DEVICE FOR DETETERMINING PREDICTION INFORMATION FOR ENCODING OR DECODING AT LEAST PART OF AN IMAGE

The present invention concerns a method and device for determining prediction information for encoding or decoding at least part of an image. The present invention further concerns a method and a device for encoding at least part of an image and a method and device for decoding at least part of an image.

Embodiments of the invention relate to the field of scalable video coding, in particular to scalable video coding in which the High Efficiency Video Coding (HEVC) standard may be applied.

BACKGROUND OF THE INVENTION

Video data is typically composed of a series of still images which are shown rapidly in succession as a video sequence to give the idea of a moving image. Video applications are continuously moving towards higher and higher resolution. A large quantity of video material is distributed in digital form over broadcast channels, digital networks and packaged media, with a continuous evolution towards higher quality and resolution (e.g. higher number of pixels per frame, higher frame rate, higher bit-depth or extended color gamut). This technological evolution puts higher pressure on the distribution networks that are already facing difficulties in bringing HDTV resolution and high data rates economically to the end user.

Video coding techniques typically use spatial and temporal redundancies of images in order to generate data bit streams of reduced size compared with the video sequences. Spatial prediction techniques (also referred to as Intra coding) exploit the mutual correlation between neighbouring image pixels, while temporal prediction techniques (also referred to as INTER coding) exploit the correlation between images of sequential images. Such compression techniques render the transmission and/or storage of the video sequences more effective since they reduce the capacity required of a transfer network, or storage device, to transmit or store the bit-stream code.

An original video sequence to be encoded or decoded generally comprises a succession of digital images which may be represented by one or more matrices

the coefficients of which represent pixels. An encoding device is used to code the video images, with an associated decoding device being available to reconstruct the bit stream for display and viewing.

Common standardized approaches have been adopted for the format and method of the coding process. One of the more recent standards is Scalable Video Coding (SVC) wherein a video image is split into smaller sections (called macroblocks or blocks) and treated as being comprised of hierarchical layers. The hierarchical layers include a base layer, corresponding to low quality images (or frames) of the original video sequence, and one or more enhancement layers (also known as refinement layers) providing quality, spatial and/or temporal enhancement images to base layer images.. SVC is a scalable extension of the H.264/AVC video compression standard. In SVC, compression efficiency can be obtained by exploiting the redundancy between the base layer and the enhancement layers.

A further video standard being standardized is HEVC, in which the macroblocks are replaced by so-called Coding Units and are partitioned and adjusted according to the characteristics of the original image segment under consideration. This allows more detailed coding of areas of the video image which contain relatively more information and less coding effort for those areas with fewer features.

In general, the more information that can be compressed at a given visual quality, the better the performance in terms of compression efficiency.

The present invention has been devised to address one or more of the foregoing concerns.

According to a first aspect of the invention there is provided a method of determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the step of deriving enhancement layer prediction information comprising for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, deriving the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, deriving the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected according to a predetermined criterion;

wherein the predetermined criterion is based on at least one of the relative location of said one of said plurality of elementary prediction units with respect to the other elementary prediction units of said plurality of elementary prediction units and the base layer prediction information of the said elementary prediction units.

An elementary prediction unit may be an elementary coding element such as a coding unit or a coding block or an element making up part of a coding unit or coding block

In an embodiment, the predetermined criterion is based on the raster scan ordering of the plurality of the elementary prediction units.

In an embodiment, the predetermined criterion determines that the prediction information of the last elementary prediction unit in raster scan order is selected.

In an embodiment, if the plurality of elementary prediction units are distributed in a horizontal direction the predetermined criterion determines that the prediction information of the elementary prediction unit located most right with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

In an embodiment, if the plurality of elementary prediction units are distributed in a vertical direction, the predetermined criterion determines that the prediction information of the bottom elementary prediction unit with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

In an embodiment, if the plurality of elementary prediction units are distributed in both vertical and horizontal directions the predetermined criterion determines that the prediction information of the elementary prediction unit located

at the right bottom part with respect to the other elementary prediction units of said plurality of elementary prediction units.

In an embodiment, if the base layer prediction information of the elementary prediction units corresponding to said region is the same for said elementary prediction units, the enhancement prediction information is derived from said common base layer prediction information.

In an embodiment, the predetermined criterion determines that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected

In an embodiment, the spatial scaling between an image of the enhancement layer and a corresponding image of the base layer is a non-integer ratio. For example, the non-integer ratio is 1.5.

In an embodiment, the corresponding base layer image is a base layer image temporally coincident with the enhancement layer image.

In an embodiment, the processing block has a 2NX2N pixel size, N being an integer. In an embodiment, the processing block has a 4X4 pixel size.

In an embodiment, the method further comprises grouping together, into a common processing unit for encoding, a plurality of neighbouring processing units having the same enhancement layer prediction information.

In an embodiment, the common processing unit is considered as a processing unit of 2Nx2N size

In an embodiment the method further includes grouping together, into a larger common processing unit for encoding, a plurality of neighbouring common processing units having the same enhancement layer prediction information.

In an embodiment, the larger common processing unit is considered as a processing unit of 2Nx2N size.

In an embodiment, the depth value of each common processing unit of the larger common processing unit is decremented.

In an embodiment, a prior step of partitioning the enhancement layer into coding units is provided such that each coding unit has the highest possible depth value; and partitioning each coding unit into said processing blocks.

According to a second aspect of the invention there is provided a device for determining prediction information for encoding at least part of an image of an

enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information,

the device comprising prediction information derivation means for deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the prediction information derivation means being operable to derive for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected according to a predetermined criterion;

wherein the predetermined criterion is based on at least one of the relative location of said one of said plurality of elementary prediction units with respect to the other elementary prediction units of said plurality of elementary prediction units and the base layer prediction information of the said elementary prediction units.

In an embodiment, the predetermined criterion is based on the raster scan ordering of the plurality of the elementary prediction units.

In an embodiment, the predetermined criterion determines that the prediction information of the last elementary prediction unit in raster scan order is selected.

In an embodiment, if the plurality of elementary prediction units are distributed in a horizontal direction the predetermined criterion is such that the prediction information of the elementary prediction unit located most right with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

In an embodiment, if the plurality of elementary prediction units are distributed in a vertical direction, the predetermined criterion is such that the prediction information of the bottom elementary prediction unit with respect to the

other elementary prediction units of said plurality of elementary prediction units is selected.

In an embodiment, if the plurality of elementary prediction units are distributed in both vertical and horizontal directions the predetermined criterion is such that the prediction information of the elementary prediction unit located at the right bottom with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

In an embodiment, if the base layer prediction information of the elementary prediction units corresponding to said region is the same for said elementary prediction units, the prediction information derivation means is operable to derive the enhancement prediction information from said common base layer prediction information.

In an embodiment, the predetermined criterion is such that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected

In an embodiment, the spatial scaling between an image of the enhancement layer and a corresponding image of the base layer is a non-integer ratio; for example the non-integer ratio is 1.5.

In an embodiment, the corresponding base layer image is a base layer image temporally coincident with the enhancement layer image.

In an embodiment, the processing block has a 2NX2N pixel size, N being an integer.

In an embodiment, the processing block has a 4X4 pixel size.

In an embodiment, grouping means are provided for grouping together, into a common processing unit for encoding, a plurality of neighbouring processing units having the same enhancement layer prediction information.

In an embodiment, the common processing unit is considered as a processing unit of 2Nx2N size.

In an embodiment, the grouping means is further configured to group together, into a larger common processing unit for encoding, a plurality of neighbouring common processing units having the same enhancement layer prediction information.

In an embodiment, the larger common processing unit is considered as a processing unit of 2Nx2N size.

In an embodiment, the depth value of each common processing unit of the larger common processing unit is decremented.

In an embodiment, means are provided for partitioning the enhancement layer into elementary prediction units such that each elementary prediction unit has the highest possible depth value; and partitioning each elementary prediction unit into said processing blocks.

According to third aspect of the invention there is provided a method of determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the step of deriving enhancement layer prediction information comprising for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, deriving the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, deriving the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected so that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected.

In an embodiment, the spatial scaling between an image of the enhancement layer and a corresponding image of the base layer is a non-integer ratio. For example, the non-integer ratio is 1.5.

In an embodiment, the corresponding base layer image is a base layer image temporally coincident with the enhancement layer image.

In an embodiment, the processing block has a 2NX2N pixel size, N being an integer.

In an embodiment, the processing block has a 4X4 pixel size.

In an embodiment the method comprises grouping together, into a common processing unit for encoding, a plurality of neighbouring processing units having the same enhancement layer prediction information.

In an embodiment, the common processing unit is considered as a processing unit of 2Nx2N size

In an embodiment, the method comprises grouping together, into a larger common processing unit for encoding, a plurality of neighbouring common processing units having the same enhancement layer prediction information.

In an embodiment, the larger common processing unit is considered as a processing unit of 2Nx2N size.

In an embodiment, the depth value of each common processing unit of the larger common processing unit is decremented.

In an embodiment, a prior step of partitioning the enhancement layer into elementary prediction units is provided such that each elementary prediction unit has the highest possible depth value; and partitioning each elementary prediction unit into said processing blocks.

According to a fourth aspect of the invention there is provided a device for determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information,

the device comprising prediction information derivation means for deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the prediction information derivation means being operable to derive for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected so that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected.

According to a further aspect of the invention there is provided method of encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising

determining enhancement layer prediction information for a processing block of the enhancement layer according to the method of embodiment of the first aspect of the invention or any embodiment of the third aspect of the invention; and

encoding the processing unit into an encoded video bitstream using said enhancement layer prediction information.

According to another aspect of the invention there is provided a device for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the device comprising

a device for determining enhancement layer prediction information for a processing block of the enhancement layer according to the method of any embodiment of the second aspect of the invention or the fourth aspect of the invention and

an encoder for encoding the processing unit into an encoded video bitstream using said enhancement layer prediction information.

At least parts of the methods according to the invention may be computer implemented. Accordingly, the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and

hardware aspects that may all generally be referred to herein as a "circuit", "module" or "system". Furthermore, the present invention may take the form of a computer program product embodied in any tangible medium of expression having computer usable program code embodied in the medium.

Since the present invention can be implemented in software, the present invention can be embodied as computer readable code for provision to a programmable apparatus on any suitable carrier medium. A tangible carrier medium may comprise a storage medium such as a floppy disk, a CD-ROM, a hard disk drive, a magnetic tape device or a solid state memory device and the like. A transient carrier medium may include a signal such as an electrical signal, an electronic signal, an optical signal, an acoustic signal, a magnetic signal or an electromagnetic signal, e.g. a microwave or RF signal.

Embodiments of the invention will now be described, by way of example only, and with reference to the following drawings in which:

Fig. 1A schematically illustrates a data communication system in which one or more embodiments of the invention may be implemented;

Fig. 1B is a schematic block diagram illustrating a processing device configured to implement at least one embodiment of the present invention;

Fig. 2 illustrates an example of an all-INTRA configuration for scalable video coding (SVC).

Fig. 3 illustrates anexemplary scalable video encoder architecture in all-INTRA mode according to at least one embodiment.

Fig. 4 illustrates anexemplary scalable video decoder architecture, associated with the scalable video encoder architecture for all-INTRA mode (as shown in Fig. 3) according to at least one embodiment.

Fig. 5 illustrates an encoding process associated with the residuals of an enhancement layer according to at least one embodiment of the invention

Fig. 6 illustrates the decoding process consistent with the encoding process of Fig. 5 according to at least one embodiment of the invention.

Fig. 7 illustrates an exemplary low- delay temporal coding structure according to the HEVC standard.

Fig. 8 illustrates an exemplary random access temporal coding structure according to the HEVC standard.

Fig. 9 illustrates an exemplary standard video encoder, compliant with the HEVC standard for video compression.

Fig. 10 is a block diagram of an exemplary scalable video encoder, compliant with the HEVC standard in the compression of the base layer.

Fig. 11 is a block diagram of an exemplary decoder, compliant with standard HEVC or H.264/AVC and reciprocal to the encoder of Fig.9.

Fig. 12 is a block diagram of an exemplary scalable decoder, compliant with standard HEVC or H.264/AVC in the decoding of the base layer, and reciprocal to the encoder of Fig.10.

Fig. 13 schematically illustrates an encoder architecture according to an embodiment of the invention.

Fig. 14 schematically illustrates elementary prediction units and prediction unit concepts specified in the HEVC standard.

Fig. 15 schematically illustrates prediction modes suitable for the scalable codec architecture, according to an embodiment of the invention.

Fig. 16 schematically illustrates an architecture of a scalable video decoder according to an embodiment of the invention.

Fig. 17 schematically illustrates an example of the prediction information up-sampling process according to an embodiment of the invention.

Fig. 18 illustrates the construction of a Base Mode prediction picture according to an embodiment of the invention .

Fig. 19 is a flow chart of an algorithm according to an embodiment of the invention used to encode an INTER picture.

Fig. 20 is a flow chart of an algorithm according to the invention used to decode an INTER picture, complementary to the encoding algorithm of Fig. 19.

Fig. 21 schematically illustrates an example of a coding derivation process applicable to one or more embodiments of the invention;

Fig. 22 schematically illustrates a method of up-sampling prediction unit partitions applicable to one or more embodiments of the invention;

Fig. 23 schematically illustrates a method of up-sampling prediction unit partitions applicable to one or more embodiments of the invention;

Fig. 24 is a flow chart illustrating steps of a method of up-sampling prediction information applicable to one or more embodiments of the invention;

Fig. 25 is a flow chart illustrating steps of a method of deriving i prediction information between layers applicable to one or more embodiments of the invention;

Fig. 26 schematically illustrates prediction information up-sampling according to an embodiment of the invention in the case of a non-integer scaling ratio;

Fig. 27 schematically illustrates inter-layer derivation of prediction information for 4x4 enhancement layer blocks in accordance with an embodiment of the invention;

Fig. 28 schematically illustrates an example of the correspondence between spatial scalability layers in the case of a 1.5 spatial ratio between layers,

Fig. 29 schematically illustrates a method of grouping enhancement layer blocks according to their derived prediction information in accordance with an embodiment of the invention;

Fig. 30 is a flow chart of steps of a method of deriving prediction information according to an embodiment of the invention,

Fig. 31 is a flow chart of steps of a method of deriving prediction information for 4x4 enhancement layer blocks according to an embodiment of the invention,

Fig. 32 is a flow chart of a method of merging prediction units according to an embodiment of the invention, and

Fig. 33 is flow chart of a method of merging elementary prediction units according to an embodiment of the invention.

Detailed Description

Figure 1A illustrates a data communication system in which one or more embodiments of the invention may be implemented. The data communication system comprises a sending device, in this case a server 11, which is operable to transmit data packets of a data stream to a receiving device, in this case a client terminal 12, via a data communication network 10. The data communication network 10 may be a Wide Area Network (WAN) or a Local Area Network (LAN). Such a network may be for example a wireless network (Wifi / 802.11a or b or g or n), an Ethernet network, an Internet network or a mixed network composed of several different networks. In a particular embodiment of the invention the data

communication system may be, for example, a digital television broadcast system in which the server 11 sends the same data content to multiple clients.

The data stream 14 provided by the server 11 may be composed of multimedia data representing video and audio data. Audio and video data streams may, in some embodiments, be captured by the server 11 using a microphone and a camera respectively. In some embodiments data streams may be stored on the server 11 or received by the server 11 from another data provider. The video and audio streams are coded by an encoder of the server 11 in particular for them to be compressed for transmission.

In order to obtain a better ratio of the quality of transmitted data to quantity of transmitted data, the compression of the video data may be of motion compensation type, for example in accordance with the HEVC type format or H.264/AVC type format.

A decoder of the client 12 decodes the reconstructed data stream received by the network 10. The reconstructed images may be displayed by a display device and received audio data may be reproduced by a loud speaker.

Figure 1B schematically illustrates a device 100, in which one or more embodiments of the invention may be implemented. The exemplary device as illustrated is arranged in cooperation with a digital camera 101, a microphone 124 connected to a card input/output 122, a telecommunications network 340 and a disk 116. The device 100 includes a communication bus 102 to which are connected:

- a central processing CPU 103 provided, for example in the form of a microprocessor
- a read only memory (ROM) 104 comprising a computer program 104A whose execution enables methods according to one or more embodiments of the invention to be performed. This memory 104 may be a flash memory or EEPROM, for example;
- a random access memory (RAM) 106 which, after powering up of the device 100, contains the executable code of the program 104A necessary for the implementation of one or more embodiments of the invention. The memory 106, being of a random access type, provides more rapid access compared to ROM 104. In addition the RAM 106 may be operable to store images and blocks of

pixels as processing of images of the video sequences is carried out on the video sequences (transform, quantization, storage of reference images etc.);

- a screen 108 for displaying data, in particular video and/or serving as a graphical interface with the user, who may thus interact with the programs according to embodiments of the invention, using a keyboard 110 or any other means e.g. a mouse (not shown) or pointing device (not shown);
- a hard disk 112 or a storage memory, such as a memory of compact flash type, able to contain the programs of embodiments of the invention as well as data used or produced on implementation of the invention;
- an optional disc drive 114, or another reader for a removable data carrier, adapted to receive a disc 116 and to read/write thereon data processed, or to be processed, in accordance with embodiments of the invention and;
- a communication interface 118 connected to a telecommunications network 34
- connection to a digital camera 101; It will be appreciated that in some embodiments of the invention the digital camera and the microphone may be integrated into the device 100 itself.

The communication bus 102 permits communication and interoperability between the different elements included in the device 100 or connected to it. The representation of the communication bus 102 given here is not limiting. In particular, the CPU 103 may communicate instructions to any element of the device 100 directly or by means of another element of the device 100.

The disc 116 can be replaced by any information carrier such as a compact disc (CD-ROM), either writable or rewritable, a ZIP disc, a memory card or a USB key. Generally, an information storage means, which can be read by a micro-computer or microprocessor, which may optionally be integrated in the device 100 for processing a video sequence, is adapted to store one or more programs whose execution permits the implementation of the method according to the invention.

The executable code enabling a coding device to implement one or more embodiments of the invention may be stored in ROM 104, on the hard disc 112 or on a removable digital medium such as a disc 116.

The CPU 103 controls and directs the execution of the instructions or portions of software code of the program or programs of embodiments of the

invention, the instructions or portions of software code being stored in one of the aforementioned storage means. On powering up of the device 100, the program or programs stored in non-volatile memory, e.g. hard disc 112 or ROM 104, are transferred into the RAM 106, which then contains the executable code of the program or programs of embodiments of the invention, as well as registers for storing the variables and parameters necessary for implementation of embodiments of the invention.

It may be noted that the device implementing one or more embodiments of the invention, or incorporating it, may be implemented in the form of a programmed apparatus. For example, such a device may then contain the code of the computer program or programs in a fixed form in an application specific integrated circuit (ASIC).

The exemplary device 100 described here and, particularly, the CPU 103, may implement all or part of the processing operations as described in what follows.

Figure 2 schematically illustrates an example of the structure of a scalable video stream 20 in the context of one or more embodiments of the invention, when all images are encoded in INTRA mode. As shown, an all-INTRA coding structure includes a series of images which are encoded independently from each other. The base layer 21 of the scalable video stream 20 is illustrated at the bottom of the figure. In this base layer, each image is INTRA coded and is usually referred to as an "I" image. INTRA coding involves predicting a macroblock or block from its directly neighbouring macroblocks or blocks within a single image or frame.

A spatial enhancement layer 22 is encoded on top of the base layer 21 as illustrated at the top of Fig. 2. This spatial enhancement layer 22 introduces some spatial refinement information over the base layer. In other words, the decoding of this spatial layer leads to a decoded video sequence that has a higher spatial resolution than the base layer. The higher spatial resolution adds to the quality of the reproduced images.

As illustrated in the figure, each enhancement image, denoted an 'EI' image, is intra coded. An enhancement INTRA image is encoded independently from any other enhancement image. It is coded in a predictive way, by predicting it only from the temporally coincident image in the base layer.

Figure 3 schematically illustrates an example of a particular type of scalable video encoder architecture 30 for all-INTRA mode, referred to herein as an INTRA LCC encoder. This coder is dedicated to the encoding of a spatial or SNR (signal to noise) enhancement layer on top of a standard coded base layer. The base layer is compliant with the HEVC or H.264/AVC video compression standard.

The overall architecture of the INTRA encoder 30 is described as follows. The input full resolution original image 31 is down-sampled 30A to the base layer resolution level 32 and is encoded 30B with HEVC 33. This produces a base layer bit-stream 34. The image 31 is now represented by a base layer which is essentially at a lower resolution than the original. Then the base layer image 33 is reconstructed 30C to produce a decoded base layer image 35 and up-sampled 30D to the top layer resolution in case of spatial scalability to produce an image 36. Thus information from only one (base) layer of the original image 31 is now available. The difference 30E with the original image constitutes a spatial residual image 37. The residual image 37 is now subjected to the normal encoding process 30F which comprises transformation, quantisation and entropy operations. The processing is performed sequentially on macroblocks using a DCT (Discret Cosine Transform) function, to produce a DCT profile over the global image area. Quantisation is performed by fitting with GGD (Generalised Gaussian Distribution) functions the values taken by DCT coefficients, per DCT channel. Use of such functions allows flexibility in the quantisation step, with a smaller step being available for more central regions of the curve. An optimal centroid position per quantisation step may also be applied to optimise the quantisation process. Entropy coding is then applied (e.g. using arithmetic coding) for the quantised data. The result is the enhancement layer 38 associated to the coding of the original image 31. The enhancement layer is also converted into a bit-stream 39 with its associated parameters 39' (39 prime) used to model DCT parameters of the residual image.

For down sampling, H.264/SVC down-sampling filters are used and for up sampling, the DCTIF interpolation filters used for the quarter-pel motion compensation in HEVC are used.

The resulting residual image is encoded using DCT and quantization, which will be further described with reference to Fig. 5. The resulting coded

enhancement layer 38 consists of coded residual data as well as parameters used to model DCT channels of the residual image.

As can be seen, this global architecture corresponds to classical scalable INTRA coding, where the spatial intra prediction and coding mode decision steps have been removed. The only prediction mode used in this INTRA scalable coder is the known inter-layer intra prediction mode.

Fig. 4 illustrates an exemplary scalable video decoder 40 associated with the type of scalable video encoder architecture 30 for all-INTRA mode (as shown in Fig. 3). The inputs to the decoder 40 are equivalent to the base layer bit-stream 34 and the enhancement layer bit-stream 39, with its associated parameters 39'. The input bit-stream to that decoder comprises the HEVC-coded base layer 33, enhancement residual coded data 38, and parameters 39' of the DCT channels in the enhancement residual image. First, the base layer is decoded 40A, which provides a reconstructed base image 41. The reconstructed base image 41 is up-sampled 40B to the enhancement layer resolution to produce an image 42. Then, the enhancement layer 38 is decoded 40C as follows. The residual data decoding process 40C is further described in association with Fig. 6. This process is invoked, which provides successive de-quantized DCT blocks 43. These DCT blocks are then inverse transformed and added to their co-located up-sampled block 40D. The so-reconstructed enhancement image 44 finally undergoes HEVC post-filtering processes 40E, i.e. de-blocking filter, sample adaptive offset (SAO), and Adaptive Loop Filter (ALF). A filtered reconstructed image 45 is produced.

Fig. 5 schematically illustrates an exemplary coding process 50 to which one or more embodiments of the invention may be applied, associated with the residuals of an enhancement layer, an example of which is image 37 shown in Fig. 3. The coding process comprises transformation by DCT function, quantisation and entropy coding. This process, embodying the invention, is also referred to as texture encoding. It may be noted that this process applies on a complete residual image, and does not proceed block by block, like in classical H.264/AVC or HEVC intra coding.

The input to the encoder 37 consists of a set of DCT blocks. Several DCT transform sizes are supported in the transform process: 16, 8 and 4. The transform size is flexible and is decided 50A according to the characteristics of the input data. The input residual image 37 is first divided into 16x16 macroblocks. The

transform size is decided for each macroblock as a function of its activity level in the pixel domain. Then the transform is applied 50B, which provides a frame of DCT block 51. The transforms used are the 4x4, 8x8 and 16x16 DCT, as defined in the HEVC standard.

The next coding step comprises computing, by channel modelling 50C, a statistical model of each DCT channel 52. A DCT channel consists of the set of values taken by samples from all image blocks at the same DCT coefficient position, for a given transform size. DCT coefficients are modelled by a Generalized Gaussian Distribution (GGD). For such a distribution, each DCT channel is assigned a quantifier. This non-uniform scalar quantifier 53 is defined by a set of quantization intervals and associated de-quantized sample values. A pool of such quantifiers 54 is available on both the encoder and on the decoder side. Various quantifiers are pre-computed off-line, through the Chou-Lookabaugh-Gray rate distortion optimization process.

The selection of the rate distortion optimal quantifier for a given DCT channel proceeds as follows. Given input coding parameters, a distortion target 55 is determined for the DCT channel under consideration. To do so, a distortion target allocation among various DCT channels, and among various block sizes, is performed. The distortion allocation ensures that each DCT channel of each block size is encoded at a level that corresponds to an identical rate distortion slope among all coded DCT channels. This rate distortion slope depends on an input quality parameter, given by the user.

Once the distortion target 55 is obtained for each DCT channel, the right quantifier 53 to use is chosen 50D. As the rate distortion curve associated with each pre-computed quantifier is known (tabulated), this merely consists in choosing the quantifier that provides minimal bitrate for given distortion target.

Then DCT coefficients are quantized 50E to produce quantised DCT X_Q values 56, and entropy coded 50F to produce a set of values $H(X_Q)$ 57. The entropy coder used consists of a simple, non-contextual, non-adaptive arithmetic coder. The arithmetic coding employs, for each DCT channel, a set of fixed probabilities, respectively associated with each pre-computed quantization interval. Therefore, these probabilities are entirely calculated off-line, together with the rate distortion optimal quantifiers. Probability values are not updated during the encoding or decoding processes, and are fixed for the whole image being processed. In

particular, this ensures the spatial random access feature, and also makes the decoding process highly parallelizable.

As a result of the proposed INTRA enhancement coding scheme, the enhancement layer bit-stream is made of the following syntax elements.

- Parameters of each coded DCT channel model 39'. Two parameters are needed to fully specify a generalized Gaussian distribution. Therefore, two parameters are sent for each encoded DCT channel. These are sent only once for each image.
- Chosen block sizes 58 are arithmetic encoded 50F. The probabilities used for their arithmetic coding are computed during the selection of transform sizes, are quantized and fixed-length coded into the output bit-stream. These probabilities are fixed for the whole image.
- Coded residual data 39 results from the entropy coding of quantized DCT coefficients.

It may be noted that the above syntax elements represent the content of coded slice data in the scalable extension of HEVC. The NAL unit container of HEVC can be used to encapsulate a slice that is coded according to the coding scheme of Figure 5.

Fig. 6 depicts an exemplary INTRA decoding process 60 to which embodiments of the invention may be applied and which corresponds to the encoding process illustrated in Fig. 5. The input to the decoder includes the coded residual data 39 and the parametric model of DCT channels 39' (39 prime), for the input image 37.

First, following a process similar to that effected in the encoder, the decoder determines the distortion target 55 of each DCT channel, given the parametric model of each coded DCT channel 39'. Then, the choice of optimal quantizers (or quantifiers) 50D for each DCT channel is performed in exactly the same way as on the encoder side. Given the chosen quantifiers 53, and thus probabilities of all quantized DCT symbols, the arithmetic decoder is able to decode the input coded residual data 39. This provides successive quantized DCT blocks, which are then inverse quantized 60A and inverse transformed 60B. The transform size of each DCT block 58 is obtained from the entropy decoding step 60C.

Fig. 7 and Fig. 8 schematically illustrate the video sequence structure for INTER coding, in so-called "low delay" and "random access" configurations,

respectively. These are the two coding structures comprised in the common test conditions in the HEVC standardization process.

Fig. 7 shows an exemplary low- delay temporal coding structure 70. In this configuration, an input image frame is predicted from several already coded frames. Therefore, only forward temporal prediction, as indicated by arrows 71, is allowed, which ensures a low delay property. The low delay property means that on the decoder side, the decoder is able to display a decoded image straight away once this image is in a decoded format, as represented by arrow 72. Note: the input video sequence is shown as comprised of a base layer 73 and an enhancement layer 74, which are each further comprised of a first image frame I and subsequent image frames B.

In addition to temporal prediction, inter-layer prediction between the base 73 and enhancement layer 74 is also illustrated in Fig. 7 and referenced by arrows, including arrow 75. Indeed, the scalable video coding of the enhancement layer 74 aims to exploit the redundancy that exists between the coded base layer 73 and the enhancement layer 74, in order to provide good coding efficiency in the enhancement layer 74.

As a consequence, several prediction modes can be employed in the coding of enhancement images. This type of standard HEVC coding can be rendered compatible with the texture coding according to embodiments of the present invention as described with reference to Figures 5 and 6.

Fig. 8 illustrates an exemplary random access temporal coding structure 80 e.g. as defined in the HEVC standard. The input sequence is broken down into groups of images (pictures), here indicated by arrows GOP. A random access property means that several access points are enabled in the compressed video stream, i.e. the decoder can start decoding the sequence at a frame which is not necessarily the first frame in the sequence. This takes the form of periodic INTRA image coding in the stream as illustrated by Figure 8.

In addition to INTRA images, the random access coding structure enables INTER prediction, both forward 81 and backward 82 (in relation to the display order as represented by arrow 83) predictions can be effected. This is achieved by the use of B images, as illustrated. The random access configuration also provides temporal scalability feature, which takes the form of the hierarchical B images, B_0 to B_3 as illustrated, the organization of which is shown in the figure.

As for the low delay coding structure of Fig. 7, additional prediction tools are used in the coding of enhancement images: inter-layer prediction tools.

This type of standard HEVC coding can be rendered compatible with the texture coding according to embodiments of the present invention detailed above. The goal is to design a temporal and inter-layer prediction scheme that is compliant with the texture codec of Figures 5 and 6, and which is efficient. By efficient, one means the predictor provided by this prediction scheme provides prediction values which are as close to the original image as possible, in order to favor compression efficiency in the enhancement layer.

To achieve this goal, the prediction process should respect the following property. Embodiments of the invention should have a full residual image to be able to perform a DCT transform and DCT channel modeling over the complete image area, i.e. globally. Therefore, the prediction process should provide a full prediction image of an enhancement image to encode, before starting to transform, quantize and encode this enhancement image. In other words, when predicting a (e.g. rectangular) block of the enhancement image, the prediction should not depend on neighboring pixel values of the block. Indeed, in the opposite case it would be necessary to encode and reconstruct those neighboring blocks before computing the prediction current block, which is not compliant with the decoding process according to embodiments of the invention.

Fig. 9 illustrates a standard video encoding device, of a generic type, conforming to the HEVC or H.264/AVC video compression system. A schematic block diagram 90 of a standard HEVC or H.264/AVC encoder is shown. The input to this non-scalable encoder includes the original sequence of frame images 91 to compress. The encoder successively performs the following steps to encode a standard video bit-stream. A first image or frame to be encoded (compressed) is divided into pixel blocks, called CTB (coded Tree Block) in the HEVC standard. These CTB are then divided into coding units of variable sizes which are the elementary coding elements in HEVC. Coding units are then partitioned into one or several prediction units for prediction as will be described in detail later. In a first part and for the purpose of simplification, we will consider that coding units and prediction units coincide. The first image is thus split into coding units. Each coding unit first undergoes a motion estimation operation 93, which comprises a search, among the reference images stored in a dedicated memory buffer 94, for reference

areas that would provide a good prediction of the coding unit. This motion estimation step provides one or more reference image indexes which identify the reference images containing reference areas, as well as the corresponding motion vectors which identify the reference areas in the reference images. A motion compensation step 95 then applies the estimated motion vectors to the found reference areas and copies the so-obtained areas into a temporal prediction image. Moreover, an Intra prediction step 96 determines the spatial prediction mode that would provide the best performance to predict the current coding unit and encode it in INTRA mode.

Afterwards, a coding mode selection mechanism 97 chooses the coding mode, among the spatial and temporal predictions, which provides the best rate distortion trade-off in the coding of the current coding unit. The difference between the current coding unit 92 (in its original version) and the so-chosen prediction area (not shown) is calculated. This provides a (temporal or spatial) residual to compress. The residual coding unit then undergoes a transform (DCT) and a quantization 98. Entropy coding 99 of the so-quantized coefficients QTC (and associated motion data MD) is performed. The compressed texture data 100 associated with the coded current coding unit 92 is sent for output.

Then, the current coding unit is reconstructed by scaling and inverse transform 101. This comprises inverse quantization and inverse transform, followed by a sum between the inverse transformed residual and the prediction area of the current coding unit. Once the current image is reconstructed and deblocked 102, it is stored in a memory buffer 94 (the DPB, Decoded Image Buffer) so that it is available for use as a reference image to predict any subsequent images to be encoded.

Finally, a last entropy coding step is given the coding mode and, in case of an inter coding unit, the motion data, as well as the quantized DCT coefficients previously calculated. This entropy coder encodes each of these data into their binary form and encapsulates the so-encoded coding unit into a container called NAL unit (Network Abstract Layer). A NAL unit contains all encoded coding units from a given slice. A coded HEVC bit-stream includes a series of NAL units.

Fig. 10 illustrates a block diagram of a scalable video encoder, which comprises a straightforward extension of the standard video coder of Fig. 9, towards a scalable video coder. This video encoder may comprise a number of

subparts or stages, illustrated here are two subparts or stages A10 and B10 producing data corresponding to a base layer 103 and data corresponding to one enhancement layer 104. Each of the subparts A10 and B10 follows the principles of the standard video encoder 90, with the steps of transformation, quantisation and entropy coding being applied in two separate paths, one corresponding to each layer.

The first stage B10 aims at encoding the H.264/AVC or HEVC compliant base layer of the output scalable stream, and hence is identical to the encoder of Fig. 9. Next, the second stage A10 illustrates the coding of an enhancement layer on top of the base layer. This enhancement layer brings a refinement of the spatial resolution to the (down-sampled 107) base layer. As illustrated in Fig. 10, the coding scheme of this enhancement layer is similar to that of the base layer, except that for each coding unit of a current image 91 being compressed or coded, an additional prediction mode can be chosen by the coding mode selection module 105. This new coding mode corresponds to the inter-layer prediction 106. Inter-layer prediction 106 consists in re-using the data coded in a layer lower than current refinement or enhancement layer, as prediction data of the current coding unit. The lower layer used is referred to as the reference layer for the inter-layer prediction of the current enhancement layer. In case the reference layer contains an image that temporally coincides with the current image, then it is called the base image of the current image. The co-located coding unit (at same spatial position) of the current coding unit that has been coded in the reference layer can be used as a reference to predict the current coding unit. More precisely, the prediction data that can be used in the co-located coding unit corresponds to the coding mode, the coding unit partition, the motion data (if present) and the texture data (temporal residual or reconstructed coding unit). In case of a spatial enhancement layer, some up-sampling 108 operations of the texture and prediction data are performed.

As illustrated by the foregoing it will be appreciated that, standard video coding approaches as detailed in association with Figs. 9 and 10, proceed on a coding unit by coding unit basis. In particular, all processing related to the encoding of one coding unit (vocabulary used in HEVC similar to the terminology “block “used in previous standards) are fully completed before starting to process and to code a subsequent coding unit in the image or frame. This means that the spatial prediction, temporal prediction, coding mode decision and actual coding are fully

processed before considering a next coding unit in the image. This is mainly due to the dependency that exists between neighbouring coding units in the images to be coded. Indeed, in standard video coders like HEVC or H.264/AVC, some causal dependencies exist between spatial neighbouring coding units of a coded image. These dependencies mainly arise from:

– Prediction information, i.e. spatial intra prediction parameters (direction used in the spatial prediction) and temporal prediction information (e.g. motion vectors) are encoded in a predictive way from coding unit to coding unit. This creates a dependency between successive coded coding units. (Such predictive coding of prediction information is acceptable and can be maintained in the design of the scalable video codec addressed by embodiments of this invention).

– Spatial prediction in the pixel domain creates a dependency between neighbouring coding units on the texture level. More precisely, a given coding unit needs to be fully available in its decoded (or reconstructed) version in the pixel domain, before the standard coder starts to process the next coding unit in the image. This is necessary so that texture spatial prediction from coding unit to coding unit is done in a perfectly synchronized way on the encoder and on the decoder side. Such spatial dependency between neighbouring coding units in the pixel domain is not compliant with the use of the texture coding process of Fig. 5 to encode INTER images. One of the problems addressed by embodiments of this invention is how to design a scalable video coding framework, so that the texture coding process of Fig. 5 can be used to encode INTER images of a scalable refinement layer.

Fig. 11 is a schematic block diagram of a standard H.264/AVC decoding system 1100 (which is very similar to an HEVC decoding system). This decoding process of a H.264/AVC bit-stream 1110 starts by the entropy decoding 1120 of each coding unit of each coded image in the bit-stream. This entropy decoding provides the coding mode, the motion data (reference images indexes, motion vectors of INTER coded coding units) and residual data. This residual data consists in quantized and transformed DCT coefficients. Next, these quantized DCT coefficients undergo inverse quantization (scaling) and inverse transform operations 1130.

The decoded residual is then added to the temporal 1140 or INTRA 1150 prediction area of current coding unit, to provide the reconstructed coding unit. The choice 1125 between INTRA or INTER prediction depends on the prediction mode information which is provided by the entropy decoding step.

The reconstructed coding unit finally undergoes one or more in-loop post-filtering processes, e.g. deblocking 1160, which aim at reducing the blocking artefact inherent to any block-based video codec, and improve the quality of the decoded image.

The full post-filtered image is then stored in the Decoded Image Buffer (DPB), represented by the frame memory 1170, which stores images that will serve as references to predict future images to decode. The decoded images 1180 are also ready to be displayed on screen.

Fig. 12 presents a block diagram of a scalable decoder 1200 which would apply on a scalable bit-stream made of two scalability layers, e.g. comprising a base layer and an enhancement layer. This decoding process is thus the reciprocal processing of the scalable coding process of Fig. 10. The scalable stream being decoded 1210, as shown in Fig. 12, is made of one base layer and one spatial enhancement layer on top of the base layer, which are demultiplexed 1220 into their respective layers

The first stage of Fig. 12 concerns the base layer decoding process B12. As previously explained for the non-scalable case, this decoding process starts by entropy decoding 1120 each coding unit of each coded image in the base layer. This entropy decoding 1120 provides the coding mode, the motion data (reference images indexes, motion vectors of INTER coded coding units) and residual data. This residual data consists of quantized and transformed DCT coefficients. Next, these quantized DCT coefficients undergo inverse quantization and inverse transform operations 1130. Motion compensation 1140 or Intra prediction 1150 data can be added 12C. Deblocking 1160 is effected. The so-reconstructed residual data is then stored in the frame buffer 1170.

Next, the decoded motion and temporal residual for INTER coding units, and the reconstructed coding units are stored into a frame buffer in the first of the scalable decoder of Fig. 12. Such frames contain the data that can be used as reference data to predict an upper scalability layer.

Next, the second stage of Fig. 12 performs the decoding of a spatial enhancement layer A12 on top of the base layer decoded by the first stage. This spatial enhancement layer decoding involves the entropy decoding of the second layer 1210, which provides the coding modes, motion information as well as the transformed and quantized residual information of coding units of the second layer.

Next step consists in predicting coding units in the enhancement image. The choice 1215 between different types of coding unit prediction (INTRA, INTER or inter-layer) depends on the prediction mode obtained from the entropy decoding step 1210.

Concerning INTRA coding units, their treatment depends on the type of INTRA coding unit.

– In case of inter-layer predicted INTRA coding unit (Intra-BL coding mode), the result of the entropy decoding 1210 undergoes inverse quantization and inverse transform 1211, and then is added 12D to the co-located coding unit of current coding unit in base image, in its decoded, post-filtered and up-sampled (in case of spatial scalability) version.

– In case of a non-Intra-BL INTRA coding unit, such a coding unit is fully reconstructed, through inverse quantization, inverse transform to obtain the residual data in the spatial domain, and then INTRA prediction 1230 to obtain the fully reconstructed coding unit 1250.

Concerning INTER coding units, their reconstruction involves their motion compensated 1240 temporal prediction, the residual data decoding and then the addition of their decoded residual information to their temporal predictor. In this INTER coding unit decoding process, inter-layer prediction can be used in two ways. First, the motion vectors associated to the considered coding unit can be decoded in a predictive way, as a refinement of the motion vector of the co-located coding unit in the base image. Second, the temporal residual can also be inter-layer predicted form the temporal residual of the co-sited coding unit in the base layer.

It may be noted that in a particular scalable coding mode of a coding unit , all the prediction information of the coding unit (e.g. coding mode, motion vector) may be fully inferred from the co-located coding unit in the base image. Such coding unit coding mode is known as so-called “base mode” in the state of the art.

Fig. 13 illustrates encoder architecture 1300 according to an embodiment of the present invention. The goal of this scalable codec design is to exploit inter-layer redundancy in an efficient way through inter-layer prediction, while enabling the use of the low-complexity texture encoder of Fig. 5.

The diagram of Fig. 13 illustrates the base layer coding, and the enhancement layer coding process for a given image of a scalable video, as proposed by embodiments of the invention.

The first stage of the process corresponds to the processing of the base layer, and is illustrated on the bottom part of the figure 1300A.

First, the input image to code 1310 is down-sampled 13A to the spatial resolution of the base layer, to obtain a raw base layer 1320. Then it is encoded 13B in an HEVC compliant way, which leads to the "encoded base layer" 1330 and associated base layer bit-stream 1340.

In the next step, some information is extracted from the coded base layer that will be useful afterwards in the inter-layer prediction of the enhancement image. The extracted information comprises at least.

– The reconstructed (decoded) base image 1350 which is later used for inter-layer texture prediction.

– The prediction information 1370 of the base image which is used in several inter-layer prediction tools in the enhancement image. It comprises, among others, coding unit information, prediction unit partitioning information, prediction modes, motion vectors, reference image indices, etc.

– Temporal residual data 1360, used for temporal prediction in the base layer, is also extracted from the base layer, and is used next in the prediction of the enhancement image.

Once all these information have been extracted from the coded base image, it undergoes an up-sampling process, which aims at adapting this information to the spatial resolution of the enhancement layer. The up-sampling of the extracted base information is affected as described below, for the three types of data listed above.

– With respect to the reconstructed base image 1350, it is being up-sampled to the spatial resolution of the enhancement layer 1380A. In the same way

as for the INTRA LCC coder of figure 3, an interpolation filter corresponding to the DCTIF 8-tap filter used for motion compensation in HEVC is employed.

– The base prediction information 1370 is being transformed, so as to obtain a coding unit representation that is adapted to the spatial resolution of the enhancement layer. The prediction information up-sampling mechanism is introduced below.

– The temporal residual information 1360 associated with INTER predicted blocks in the base layer is collected into an image buffer, and is up-sampled to 1380C by means of a 2-tap bi-linear interpolation filter. This bi-linear interpolation of residual data is identical to that used in the former H.264/SVC scalable video coding standard.

Once all the information extracted from the base layer is available in its up-sampled form, then the encoder is ready to predict 13C the enhancement image. The prediction process used in the enhancement layer is at the core of the invention and is executed in a strictly identical way on the encoder side and on the decoder side.

The prediction process involves selecting the enhancement image organization in a rate distortion optimal way in terms of coding unit (CU) representation, prediction unit (PU) partitioning and prediction mode selection. (These concepts are further defined later in connection with Fig. 14, and form part of the HEVC standard).

Fig. 14 depicts the coding units and prediction unit concepts specified in the HEVC standard.. A coding unit of an HEVC image corresponds to a square block of that image, and can have a size in a pixel range from 8x8 to 64x64. A coding unit which has the highest size authorized for the considered image is also called a Largest Coding Unit (LCU) or CTB (coded tree block) 1410. As already mentioned above, for each coding unit of the enhancement image, the encoder decides how to partition it into one or several prediction units (PU) 1420. Each prediction unit can have a square or rectangular shape and is given a prediction mode (INTRA or INTER) and some prediction information. With respect to INTRA prediction, the associated prediction parameters consist in the angular direction used in the spatial prediction of the considered prediction unit, associated with corresponding spatial residual data. In case of INTER prediction, the prediction information comprises the reference image indices and the motion vector(s) used to

predict the considered prediction unit, and the associated temporal residual texture data. Illustrations 14A to 14H show some of the possible arrangements of partitioning which are available.

Referring again to Fig. 13, the prediction process 13C attempts to construct a whole prediction image 1391 of current enhancement image to code. To do so, it determines the best rate distortion trade-off between the quality of that prediction image and the rate cost of the prediction information to encode.

The outputs of this prediction process are as follows:

– A set of coding units with associated size, which covers the whole prediction image.

– For each coding unit, a partitioning of this coding unit into one or several prediction units. Each prediction unit is selected among all the prediction unit shapes allowed by the HEVC standard, which are illustrated on the bottom of figure 14.

– for each prediction unit, a prediction mode decided for that prediction unit, together with the prediction parameters associated with that prediction unit.

Therefore, for each candidate coding unit in the enhancement image, the prediction process of Fig. 13 determines the best prediction unit partitioning and prediction unit parameters in that candidate CU.

In particular, for a given prediction unit partitioning of the CU, the prediction process searches the best prediction type for that prediction unit. In HEVC, each prediction unit is given the INTRA or INTER prediction mode. For each mode, prediction parameters are determined. INTER prediction mode consists in the motion compensated temporal prediction of the prediction unit. This uses two lists of past and future reference images depending on the temporal coding structure used (see Fig. 7 and Fig. 8). This temporal prediction process as specified by HEVC is re-used here. This corresponds to the prediction mode called "HEVC temporal predictor" 1390 on Fig. 13. Note that in the temporal predictor search, the prediction process searches the best one or two (respectively for uni- and bi-directional prediction) reference areas to predict a current prediction unit of current image.

INTRA prediction in HEVC involves predicting a prediction unit with the help of neighboring PUs of current prediction unit that are already coded and

reconstructed. Such spatial prediction process cannot be used in the proposed system, because it is not compliant with the use of the texture coding process of Fig. 5.

As a consequence, the spatial prediction process of HEVC has been replaced in the coder of Fig.13 by two prediction types, called "Intra BL" and "Base Mode". The Intra BL prediction type comprises predicting a prediction unit of the enhancement image with the spatially corresponding area in the up-sampled decoded base image. The "Base Mode" prediction mode comprises predicting an enhancement prediction unit from the spatially corresponding area in a so-called "Base Mode prediction image". This Base Mode prediction image is constructed with the help of inter-layer prediction tools. The construction of this base mode prediction image is explained in detail below, with reference to Fig. 18. Briefly, it is constructed by predicting current enhancement image by means of the up-sampled prediction information and temporal residual data that has previously been extracted from the base layer and re-sampled to the enhancement spatial resolution.

It may be noted that the "Intra BL" and "Base Mode" prediction modes try to exploit the redundancy that exists between the underlying base image and current enhancement image. They correspond to so-called inter-layer prediction tools that we have introduced into the HEVC coding system.

The "rate distortion optimal mode decision" of Figure 13 results in the following elements.

– a set of coding unit representations with associated prediction information for current image. This is called prediction information 1392 on Fig. 13. All this information then undergoes a prediction information coding step, which constitutes a part of the coded video bit-stream. Note that in this prediction information coding, the two inter-layer prediction modes, i.e. Intra BL and Base Mode, are signaled as particular INTRA prediction modes. As a result, in terms of prediction information coding, the spatial prediction modes of HEVC are all removed and two INTRA prediction modes are used instead.

It may be noted that according to another embodiment, the "Intra BL" and/or "Base Mode" prediction images of Fig. 13 can be inserted into the list of reference images used in the temporal prediction of current enhancement image.

– an image 1391, which represents the final prediction image of current enhancement image to code. This image is then used to encode the texture data part of current enhancement image.

The next encoding step illustrated in Fig. 13 comprises computing the difference 1393 between the original image and the obtained prediction image. This difference comprises the residual data of current enhancement image 1394, which is then processed by the texture coding process 13D, as described above. The process provides encoded DCT X values 1395 which comprise enhancement coded texture for output and decoder information such as parameters of the channel model 1397 for output. A further available output is the enhancement coded prediction information 1398 derived from the prediction information 1392.

Fig. 15 summarizes all the prediction modes that can be used in the scalable codec architecture, according to embodiments of the invention, used to predict a current enhancement image. Schematic 1510 corresponds to the current enhancement to predict. The base image 1520 corresponds to the base layer decoded image that temporally coincides with current enhancement image. Schematic 1530 corresponds to an example reference image in the enhancement layer used for the temporal prediction of current image 1510. Finally, schematic 1540 corresponds to the Base Mode prediction image introduced above in association with Fig. 13.

As illustrated by Fig. 15, and as explained above, the prediction of current enhancement image 1510 comprises determining, for each block 1550 in current enhancement image 1510, the best available prediction mode for that block 1550, considering temporal prediction, Intra BL prediction and Base Mode prediction.

Fig. 15 also illustrates the fact that the prediction information contained in the base layer is extracted, and then is used in two different ways.

First, the prediction information of the base layer is used to construct 1560 the "Base Mode" prediction image 1540. This construction is discussed below with reference to Fig. 18.

Second, the base layer prediction information is used in the predictive coding 1570 of motion vectors in the enhancement layer. Therefore, the INTER prediction mode illustrated on Fig. 15 makes use of the prediction information contained in the base image 1520. This allows inter-layer prediction of the motion

vectors of the enhancement layer, hence increases the coding efficiency of the scalable video coding system.

Fig. 16 schematically illustrates an architecture of the scalable video decoder 1600 according to an embodiment of the invention. This decoder architecture performs the reciprocal process of the encoding process of Fig. 13.

Inputs to the decoder illustrated in Fig. 16 include:

- coded base layer bit-stream 1601
- coded enhancement layer bit-stream 1602.

The first stage of the decoding process 16A corresponds to the base layer, starting with the decoding 16A' of the base layer encoded base image 1610. This decoding is then followed by the preparation of all data useful for the inter-layer prediction of the enhancement layer. The data extracted from the base layer decoding step is of three types:

- the decoded base image 1611 undergoes a spatial up-sampling step 16C, in order to form the "Intra BL" prediction image 1612. The up-sampling process 16C used here is identical to that of the encoder (Fig. 13).
- the prediction information contained in the base layer (base motion information 1613) is extracted and re-sampled 16D towards the spatial resolution of the enhancement layer. The prediction info up-sampling process is the same as that used on the encoder side.
- the temporal residual texture data contained in the base layer (base residual 1615) is extracted and up-sampled 16E, in the same way as on the encoder side, to give up-sampled residual information .

Once all the base layer texture and prediction information has been up-sampled, then it is used to construct the "Base Mode" prediction image 1616, exactly in the same way as on the encoder side.

Next, the processing of the enhancement layer 16B is effected as illustrated in the upper part of Fig. 16. This begins with the entropy decoding 16F of the prediction information contained in the enhancement layer bit-stream to provide decoded prediction information 1630. This, in particular, provides the coding unit organization of the enhancement image, as well as their partitioning into prediction units, and the prediction mode (coding mode 1631) associated to each prediction unit.

Once the prediction mode of each prediction unit of the enhancement image is obtained, the decoder 1600 is able to construct the final complete prediction image 1650 that was used in the encoding of current enhancement image.

The next decoder steps then comprises decoding 16G the texture data (encoded DCT X 1632) associated to current enhancement image. This LCC texture decoding process follows the same process as explained above with reference to Fig. 6 and produces decoded residual data X_{deQ} 1633. The channel model parameters 1634 are also entropy decoded and are used as part of the texture coding 16G.

Once the entire residual image 1633 is obtained from the texture decoding process, it is added 16H to the prediction image 1650 previously constructed. This leads to the decoded current enhancement image 1635 which, optionally, undergoes some in-loop post-filtering process 16I. Such processing may comprise the HEVC deblocking filter, Sample Adaptive Offset (specified by HEVC) and Adaptive Loop Filtering (also specified by the HEVC standard).

The decoded image 1660 is ready for display and the individual frames can each be stored as a decoded reference image 1661, which may be useful for motion compensation 16J in association with the HEVC temporal predictor 1670, as applied for subsequent frames.

Fig. 17 schematically illustrates the prediction information up-sampling process, executed both by the encoder and the decoder in at least one embodiment of the invention in order to construct the "Base Mode" prediction image e.g. 1540. The prediction information up-sampling step is a useful mean to perform inter-layer prediction.

The left side of Fig. 17 1710 illustrates a part of the base layer image. In particular, the Coding Unit representation that has been used to encode the base image is illustrated, for the two first CTB (Coded tree block) of the image 1711 and 1712. The CTBs have a height and width, represented by arrows 1713 and 1714, respectively, and an identification number 1715, here shown running from zero to two. The Coding Unit quad-tree representation of the second CBT 1712 is illustrated, as well as prediction unit (PU) partitions e.g. partition 1716. Moreover, the motion vector associated with each prediction unit, e.g. vector 1717 associated with prediction unit 1716, is shown.

On the right side of Fig. 17, the result 1750 of the prediction information up-sampling process applied on base layer 1710 is illustrated. On this figure, the CTB size (height and width indicated by arrows 1751 and 1752, respectively) is the same in the enhancement image and in the base image, i.e. the base image CTB has been magnified. As can be seen, the up-sampled version of base CTB 1712 results in the enhancement CTBs 2, 3, 6 and 7 (references 1753, 1754, 1755 and 1756, respectively). The individual prediction units exist in a scaling relationship known as a quad-tree. It may be noted that the coding unit quad-tree structure of coding unit 1712 has been re-sampled in 1750 as a function of the scaling ratio that exists between the enhancement image and the base image. The prediction unit partitioning is of the same type (i.e. the corresponding prediction units have the same shape) in the enhancement layer and in the base layer. Finally, motion vector coordinates e.g. 1757 have been re-scaled as a function of the spatial ratio between the two layers.

In other words, three main steps are involved in the prediction information up-sampling process:

– the coding unit quad-tree representation is first up-sampled. To do so, a depth parameter of the base coding unit is decreased by one in the enhancement layer.

– the coding unit partitioning mode is kept the same in the enhancement layer, compared to the base layer. This leads to prediction units with an up-scaled size in the enhancement layer, which have the same shape as their corresponding prediction unit in the base layer.

– the motion vector is re-sampled to the enhancement layer resolution, simply by multiplying associated *x* and *y* coordinates by the appropriate scaling ratio.

As a result of the prediction information up-sampling process, some prediction information is available on the encoder and on the decoder side, and can be used in various inter-layer prediction mechanisms in the enhancement layer.

In the current scalable encoder and decoder architectures, this up-scaled prediction information is used in two ways.

– they are used in the construction of the "Base Mode" prediction image of current enhancement image, as already discussed with reference to Fig. 13 and Fig. 16.

– the up-sampled prediction information is also used for the inter-layer prediction of motion vectors in the coding of the enhancement image. Therefore one additional predictor is used compared to HEVC, in the predictive coding of motion vectors.

Fig. 18 illustrates the construction of a Base Mode prediction image 1800 in the context of at least one embodiment of the invention. This image is referred to as a Base Mode image, because it is predicted by means of the prediction information issued from the base layer 1801. The figure also indicates the magnification 1802 of the base layer 1801 to the dimensions of an associated enhancement layer. The inputs to this process are as follows:

– lists of reference images e.g. 1803 useful in the temporal prediction of current enhancement image i.e. the Base Mode prediction image 1800.

– prediction information e.g. temporal prediction 18A extracted from the base layer 1801 and re-sampled e.g. temporal prediction 18B to the enhancement layer 1802 resolution. This corresponds to the prediction information resulting from the process described in association with Fig. 17.

– temporal residual data issued from the base layer decoding, and re-sampled to the enhancement layer resolution e.g. inter-layer temporal residual prediction 18C.

– base layer reconstructed image 1804.

The Base Mode image construction process comprises predicting each coding unit e.g. CTB1805 of the enhancement image, conforming to the prediction modes and parameters inherited from the base layer.

It proceeds as follows.

– For each largest coding unit in current enhancement image 1805
  - Obtain the up-sampled Coding Unit representation issued from the base layer (algorithm of figure 17)
  - For each CU contained in current CTB
    - For each PU in current CU

- Predict current PU with its prediction information inherited from the base layer

The prediction unit prediction step proceeds as follows. In the case where the corresponding base prediction unit was Intra-coded e.g. base layer intra coded block 1806, then current prediction unit is predicted by the reconstructed base prediction unit, re-sampled to the enhancement layer resolution 1807. This prediction is associated with an inter-layer spatial prediction 1811. In case of an INTER coded base prediction unit 1808, then the corresponding prediction unit in the enhancement layer 1809 is also temporally predicted, by using the motion information 18B inherited from the base layer 18A. This means that the reference image(s) in the enhancement layer that correspond to the same temporal position of the reference image(s) of the base prediction unit are used. A motion compensation step 18B is applied by applying the motion vector inherited 1810 from the base onto these reference images. Finally, the up-sampled temporal residual data 18C of the co-located base prediction unit is applied onto the motion compensated enhancement prediction unit, which provides the predicted prediction unit in its final state.

Once this process has been applied on each prediction unit in the enhancement image, a full "Base Mode" prediction image is available.

Fig. 19 illustrates an algorithm according to at least one embodiment of the invention used to encode an INTER image. The input to the algorithm comprises the original image to encode, respectively re-sampled to the spatial resolution of each scalability layer to encode.

In what follows the term "base layer" can be used to designate a reference layer used for inter layer prediction. This terminology is adapted to the case where a scalable coder generates 2 layers. However, it is well known that for a coder generating more than 2 layers, any layer lower than the layer to be encoded can be used for inter layer prediction. It may be noted that in general, the layer immediately below the layer to encode is used. "

The overall algorithm includes a loop over each scalability layer to encode. The current INTER image is being encoded with each scalability layer being successively or sequentially processed through the algorithm. The layers are indexed 1902. For each scalability layer in succession, the algorithm tests 1903 if current layer corresponds to the base layer, the base layer being indexed as layer 0

(zero). If so, then a standard image encoding process is applied on the current image. For the case illustrated in Fig. 19, the base image is HEVC-encoded 1904.

When a current layer is not the base layer (e.g. is a first enhancement layer), the algorithm switches to preparing all the prediction data useful to predict current enhancement image to code, according to embodiments of the proposed invention. This data includes three main parts:

– the decoded base image of current image is obtained 1905 and up-sampled 1906 in the pixel domain towards the spatial resolution of current enhancement layer. This provides one prediction image, called the "Intra BL" prediction image.

– all the prediction information contained in the coded base layer is extracted from the base image 1907, and then is up-sampled 1908 towards current enhancement layer, as previously explained with reference to figure 17. Next, this up-sampled prediction info is used in the construction of the "Base Mode" prediction image 1909 of current enhancement image, as previously explained with reference to Fig. 18.

– temporal residual data contained in the base image is extracted from the base layer 1910, and then is up-sampled 1911 towards the spatial resolution of current enhancement layer.

Next, the up-sampled prediction info, together with this up-sampled temporal residual data, are used in the construction of the "Base Mode" prediction image of current enhancement image, as previously explained with reference to Fig. 18.

The next step of the algorithm includes searching the best way to predict the current enhancement image, given the available set of prediction data previously prepared. The algorithm performs the best prediction search 1912 based on the obtained three sets of prediction images: temporal reference(s), Intra BL, Base Mode. This prediction search step computes the following data.

– for each CTBin current image, the search step decides how to divide the CTB into smaller Coding Units (CUs).

– for each Coding Unit, the search step decides how to partition the coding unit into one or more prediction unit(s), and how to predict each prediction unit.

– the prediction parameters decided for each prediction unit include the prediction mode (INTRA or INTER) together with the prediction parameters associated to this prediction mode. With respect to INTER prediction, the same temporal prediction system as in HEVC is employed. Therefore, the prediction parameters include the indexes of the reference image(s) used to predict current prediction unit, as well as the associated motion vector(s). Concerning INTRA prediction, two types of INTRA prediction are allowed in embodiments of the invention: Intra BL prediction and Base Mode prediction. The best INTRA prediction between these two modes is determined.

The best prediction for current prediction unit, among the best INTER prediction for that prediction unit and the best INTRA prediction for that prediction unit, is determined.

Next, for a candidate coding unit, the best candidate prediction unit for that coding unit is selected. Finally the best coding unit splitting configuration (see Fig. 14) for the considered LCU is selected.

It may be noted that the prediction modes that are evaluated in this prediction search step are such that no texture prediction from a block to another block in the same image is involved. Therefore a whole prediction image can be computed before that texture coding process starts processing current image.

Once the prediction search for current image is done, then a set of prediction information is available for current image. This prediction information is able to fully describe how current enhancement image is predicted. Therefore, this prediction information is encoded 1913 and written to the output enhancement bit-stream, in order to indicate the decoder how to predict current image.

In addition to the prediction information, the prediction step also provides a full prediction image for current image. In the next step of the algorithm of Fig. 19, the so-obtained prediction image is then subtracted from the original image to code in current enhancement layer i.e. the residual image for the current image is obtained 1914.

The next step then comprises applying the texture coding of Fig. 5 on the residual image 1915 issued from previous step. Then texture coding process is performed as described previously with reference to Figure 5.

Once the current image is encoded at the current scalability level, then the algorithm checks whether current layer is the last scalability layer to encode

1916. If yes, then the algorithm ends 1917. If no, the algorithm moves to process the next scalability layer, i.e. it increments current layer index 1918, and returns to the testing step 1903 described above.

Fig. 20 schematically illustrates the overall algorithm used to decode an INTER image, according to at least one embodiment of the proposed invention. The input to this algorithm includes the compressed representations of the input image, comprising a plurality of scalability layers to be decoded, indexed as 2002.

Similar to the coding algorithm of Fig. 19, this decoding algorithm comprises a main loop on the scalability layers that constitutes the scalable input bit-stream to process.

Each layer is considered sequentially, the following is applied. The algorithm tests 2003 if a current layer corresponds to the lowest layer of the stream, the base layer normally being assigned a value 0 (zero). If so, then a standard, e.g. HEVC, decoding process is applied 2004 on current image.

If not, then the algorithm prepares all the prediction data useful to construct the prediction image of current enhancement image. Thus the same base layer data extraction and processing as on the encoder side is performed (1905 to 1911). This leads to restoration of the set of three prediction data schemes used to construct the prediction image of current enhancement image. This is facilitated by computation of the same Intra BL and Base Mode prediction images.

The next step of the algorithm comprises decoding the prediction information for the current image from the input bit-stream 2005. This provides information on how to construct the current prediction image 2006, given the Intra BL, Base Mode and temporal reference images available.

The decoded prediction data thus indicates how each CTB is decomposed into coding units (CU) and prediction units (PU), and how each prediction unit is predicted. The decoder is then able to construct the full prediction image of current enhancement image being decoded. At this stage of the decoder, exactly the same prediction image as on the encoder side is available.

The next step comprises the texture decoding of the input coded texture data on the current residual image 2007, for the entire enhancement image. The same decoding algorithm is applied as described with reference to Fig. 6.

Once the decoded residual image is available, the obtained residual image is added to the prediction image previously computed 2008, which provides the reconstructed version of current enhancement image.

Additionally it is possible to follow this with post-processing of current image (not shown), i.e. a deblocking filter, sample adaptive offset and adaptive loop filtering.

Finally, the algorithm tests if current scalability layer is the last layer to decode 2009. If so, the algorithm of Fig. 20 ends 2010. If not, the algorithm increments the layer 2011 and returns to the testing step 2003, which checks if the current layer is the base layer.

Coding unit and prediction unit concepts of HEVC have been illustrated in Figure 14. As explained with reference to Figure 14, coding units of HEVC have a maximum size equal to 64x64, and are organized in a quad-tree manner to represent a coded HEVC image.

An up-sampling process applied to base layer prediction information is schematically illustrated in Figure 17. Such a method is applied for inter-layer prediction processes by the overall codec schematically illustrated in Figure 13 and 16.

Embodiments of the invention as described in what follows include a particular method for up-sampling prediction information from the base layer to the enhancement layer, in the case of dyadic spatial scalability. Figure 17 illustrates a goal of this up-sampling process. In the following description, an up-sampling method used according to at least one embodiment of the invention is presented. In particular, one of the main features of the described embodiments relates to the up-sampling of HEVC prediction unit partitions.

The overall prediction up-sampling process illustrated in Figure 17 comprises firstly up-sampling the coding unit structure, and then in up-sampling the prediction unit partitions. Figure 21 illustrates an example of how HEVC coding units are up-sampled from a base layer towards a spatial enhancement layer, in the context of embodiments of the invention, in the case of dyadic spatial scalability.

As already explained, the coding units in HEVC are organized in a quad-tree fashion. To do so, each coding unit has an associated depth level in the quad-tree. The up-sampling of coding units involves the following process. For a given CTB in the enhancement layer, the coding units that spatially correspond to

that enhancement CTB are searched for in the base image. The enhancement CTB is then given coding unit depth values that are equal to the depth values contained in the corresponding base coding units, decreased by 1. These decreased depth values then provide a quad-tree that corresponds to the quad-tree in the base image, up-sampled by 2 in width and height.

Figure 22 schematically illustrates an exemplary method that may be applied in the context of embodiments of the invention order to up-sample prediction unit partitions. The exemplary mechanism up-samples base prediction units differently, according to the prediction unit partition type in the base layer.

For each up-sampled coding unit in the enhancement layer, the following applies.

- if the co-located CU in the base layer has a depth higher than zero, then the prediction unit type in the enhancement CU is set equal to the prediction unit type in the base layer;
- otherwise if the base CU has a depth value equal to 0, i.e. the CU has the maximum allowed size, then the following applies:

If the co-located base CU 2201 has a symmetric prediction unit partition, then the enhancement CU 2202 is given the 2Nx2N prediction unit type. Such a case is illustrated in Figure 22(a). The reason for this is that the base CU 2201, when up-sampled to the enhancement layer resolution, spatially covers two largest coding units in the enhancement picture.

If the co-located base CU 2212 has an asymmetric prediction unit type, then the enhancement CU 2212 is assigned a prediction unit type that depends on its spatial position in the enhancement layer, and on the base asymmetric prediction unit. This is illustrated in Figure 22(b). Indeed, in the example of Figure 22 (b), the base CU 2211 has a PU type equal to nLx2N. The spatial up-sampling of the base CU 2212 covers 4 LCUs in the enhancement layer. Since the goal of the PU up-sampling is to preserve the spatial geometry of the base layer, the enhancement PU assignment is carried out as follows:

Enhancement LCUs that have an even x index (2212a and 2212c), are given a symmetric PU type equal to Nx2N. Enhancement LCUs that have an odd x index (2212b and 2212d) are given the 2Nx2N PU type. Therefore, the noticeable technical effect of this embodiment is that for base CUs of a maximum allowed size, asymmetric PU partitions are transformed into symmetric PU partitions in the

enhancement layer. This technical point makes the proposed method of the embodiment of the invention different from the prior art.

A method used in the prior art is illustrated in Figure 22(c). This prior art method consists in systematically providing the enhancement coding units with the same prediction unit partitions as in the base layer, whatever the enhancement coding unit configuration.

The technical result obtained between the prior art as illustrated in Figure 22(c)) and the embodiment of the invention as illustrated in Figure 22(b) can be seen in Figure 22.

The up-sampled prediction information in the embodiment of the invention as illustrated in Figures 22(a) and 22(b) provide a better representation of the motion contained in the considered video sequence.

Figure 23 illustrates an example of an up-sampling method in with the context of an embodiment of the invention, when applied to a base coding unit that has the maximum allowed coding unit size. A similar process as explained with reference to Figure 22 (b) is performed, for a base prediction unit 2301 of type 2NxnD for a coding unit 2302 of an enhancement layer.

In the example of Figure 23, the base CU 2301 has a PU type equal to 2NxnD. The spatial up-sampling of the base CU 2301 covers 4 CTBs in the enhancement layer. Since the goal of the PU up-sampling is to preserve the spatial geometry of the base layer, the enhancement PU assignment is carried out as follows:

The PU type assignment in the enhancement CTB depends on the parity of the y- coordinates of the enhancement CU *U*. If the parity is odd, then the enhancement CTB is assigned the PU type 2NxN. Otherwise, it is given the PU type 2Nx2N. Consequently enhancement LCUs on the lower row (2302c and 2212d), are given a symmetric PU type equal to 2NxN. Enhancement CTBs on the upper row (2302a and 2302b) are given the 2Nx2N PU type.

Figure 24 illustrates an example of a global prediction information up-sampled algorithm in the context of an embodiment of the invention employed in the scalable video codec of Figures 13 and 16. The inputs to this algorithm include the enhancement image currently being processed by the scalable coder of Figure 13 or the decoder of Figure 16.

The algorithm performs a processing loop on the CTBs contained in a current enhancement image. For each CTB noted *currCTB*, the algorithm sets out to compute the prediction information of that CTB, by transforming the prediction information of the coding units of the base image coding units that spatially coincide with the current enhancement CTB *currCTB.* To do so, the following steps are applied for current enhancement CTB *currCTB.*

Firstly in step S2400, the process is initialised with the current CTB *currCTB* of the enhancement image. Then in step S2401 the spatial area in the base image that spatially corresponds to the current enhancement CTB *currCTB* is determined. This spatial area corresponds to a portion of a CTB in the base image. The CTB size in the base layer is ¼ of the CTB size in the enhancement layer.

Then, in step S2402 a loop on coding units *baseCU*s that are contained in the corresponding spatial area of the base image is initialised. Several cases are possible:

- the corresponding spatial area *baseCU*s of the base layer may be contained entirely within a CU of the base layer, i.e. the CTB of the enhancement layer at that spatial location is not divided among CUs of the base layer.
- the corresponding spatial area *baseCU*s of the base layer contains exactly one CU in the base layer.
- the corresponding spatial area baseCUs of the contains several CUs in the base layer.

Therefore, the loop on each base image CU of the area processes at least one CU of the base layer. Each CU of the *baseCU*s processed by this loop is noted *subCU*. For each considered coding unit *subCU* of the base image, a corresponding CU in the enhancement image, noted *enhCU*, is created by the up-sampling process.

For each successively considered coding unit *subCU* of the base layer and corresponding *enhCU,* of the enhancement layer, the algorithm of Figure 24 assigns, in step S2403, a depth value to the enhancement CU *enhCU* being created. This depth is set equal to the depth value of the corresponding *base CU subCU*, decreased by 1. If the base CU *subCU* has a depth equal to 0, then the depth of the corresponding enhancement CU *enhCU* is also set to 0.

The next step S2404 of the algorithm involves deriving the prediction unit partition of the base CU *subCU*, in order to assign an appropriate prediction

unit partition of the enhancement CU *enhCU*. This step will be described in more detail with reference to Figure 25. The prediction unit type of the base CU *subCU* is transformed and the resulting PU type is given to the enhancement coding unit *enhCU*

In step S2405 it is determined if the current base coding unit *subCU* being processed is the last coding unit of the *baseCU*s of the corresponding spatial area of the base layer. In the case where at least one further base coding unit subCU is to be processed steps, S2403 and S2404 are repeated for deriving the depth and the prediction unit partition, respectively, of the corresponding enhancement unit *enhCU* of the next CU contained in the baseCU area in the base image.

Once all base coding units concerned by the *baseCU*s in the base image are processed, then the algorithm of Figure 24 ends in step S2408.

Figure 25 is a flow chart illustrating steps of an exemplary method for deriving the prediction unit partitioning information from the base layer to the enhancement layer, in the context of an embodiment of the invention. The method of Figure 25 is invoked by the algorithm of Figure 24. The input to the algorithm of Figure 25 includes:

- the current coding unit *enhCU* being considered in the current enhancement image in the algorithm of Figure 24.
- the current coding unit *subCU* being considered in the current base image by the algorithm of Figure 24.

An aim of the algorithm figure 25 is to provide the enhancement coding unit *enhCU* with the prediction unit partition that best reflects the motion information contained in the base layer.

The first step S2500 of the algorithm tests if the depth value of current base CU *subCU* is greater than zero. If this is the case, then in step S2501 the same prediction unit type is assigned to the enhancement CU *enhCU* as that of the corresponding base coding unit *subCU*. Otherwise, the particular prediction unit derivation process proceeds to step S2502 in the case where the depth value of the base CU is equal to zero. This signifies that the size of the base CU *subCU* is equal to the highest CU size allowed. This covers the practical examples previously presented with reference to Figures 22 and 23.

The following then applies:

- The algorithm tests in step S2502 if the prediction unit type of the base CU *subCU* is a symmetric PU type. If this test is positive, then the corresponding enhancement coding unit *enhCU* is given the PU type 2Nx2N in step S2512. An example of such configuration is illustrated in Figure 22(a). Then the algorithm of figure 25 ends in step S2507. It may be noted that in this case, it is also possible to provide the enhancement CU *enhCU* with the NxN prediction unit type in step S2512. This may be advantageous if the enhancement layer coding process includes a refinement step of the up-sampled motion information.

- If the previous test of step S2502 is negative, then the next step S2503 of the algorithm tests if the base coding unit *subCU* has an associated PU type equal to nLx2N. This corresponds to the asymmetric PU type illustrated in Figure 22(b), in the base layer. If this test is positive, then the PU type assigned to the enhancement CU in step S2513 depends on the parity of the x- position of the corresponding enhancement CU *enhCU*. If the x- coordinates of the corresponding enhancement CU *enhCU* is even, then the enhancement CU is assigned the PU type Nx2N. This corresponds to the left side enhancement CU illustrated in Figure 22(b). Otherwise the enhancement CU is assigned the PU type 2Nx2N. Once the PU type is assigned to the enhancement CU, the algorithm of Figure 25 ends in step S2507.

- If the previous test of step S2503 is negative, then the next step S2504 of the algorithm tests if the PU type of the base CU *subCU* is equal to nRx2N. If so, in step S2514 the PU type assignment in the corresponding enhancement CU *enhCU* again depends on the parity of the x- coordinates of the enhancement CU *enhCU*. If the parity is odd, then the enhancement CU is assigned the PU type Nx2N. Otherwise, it is given the PU type 2Nx2N. The algorithm of Figure 25 then ends in step S2507.

- If the previous test of step S2504 is negative, then the next step S2505 of the algorithm tests if the PU type of the base CU *subCU* is equal to 2NxnU. If so, then in step S2515 the PU type assignment in the enhancement CU *enhCU* depends on the parity of the y- coordinates of the enhancement CU *enhCU*. If the parity is even, then the enhancement CU is assigned the PU type 2NxN. Otherwise, it is given the PU type 2Nx2N. The algorithm of Figure 25 then ends in step S2507.

- If the previous test of step S2505 is negative, then the next step S2506 of the algorithm tests if the PU type of the base CU *subCU* is equal to 2NxnD. If so, in step S2515 then the PU type assignment in the enhancement CU depends on the parity of the y- coordinates of the enhancement CU *enhCU*. If the parity is odd, then the enhancement CU *enhCU* is assigned the PU type 2NxN. Otherwise, it is given the PU type 2Nx2N. The algorithm of Figure 25 then ends in step S2507. This last case corresponds to the exemplary case illustrated in Figure 23.

Once the PU type, as determined by the algorithm of Figure 25, is assigned to the current enhancement CU *enhCU*, the algorithm of Figure 25 ends and the enhancement image coding process returns to step S2405 of the algorithm of Figure 24.

Figure 26 schematically illustrates a mechanism for deriving prediction information from a base layer image to an enhancement layer image according to an embodiment of the invention, in the case of spatial scalability, when a scaling ratio equal to 1.5 links the base layer and the enhancement layer.

Figure 26(a) illustrates an example of a HEVC coded image organization, in a coded base image of a scalable HEVC bit-stream. The organization of the coded base image, in terms of CTB, coding units (CUs) and prediction units (PUs) is schematically illustrated.

Figure 26(b) schematically illustrates the enhancement image organization in terms of CTBs, CUs and PUs, resulting from a prediction information up-sampling process applied to the base image prediction information. By prediction information, in this example is meant a coded image structure in terms of CTBs, CUs and PUs.

Figure 26 illustrates a case where the CTB size in the enhancement layer is identical to the CTB size in the base layer. As can be seen with reference to Figure 26(b), the prediction information that corresponds to one CTB in the base image spatially overlaps several CTBs in the enhancement image. Indeed there is no longer any correspondence between a CTB in the enhancement layer and a co-located part of a CTB in the base layer, as was the case in the dyadic spatial scalability case (as illustrated in Figure 17).

Embodiments of the invention help to provide an algorithm for deriving prediction information of the base layer towards the spatial resolution of the

enhancement layer, while capturing as much prediction information as possible from the base layer. This last point is of interest, since the inter-layer prediction process involved in the encoding of the enhancement image (as previously explained with reference to Figure 13) becomes all the more efficient as this inter-layer prediction information derivation process is efficient. As a result, the more efficient the inter-layer prediction process, the higher the compression efficiency of the enhancement layer.

A general method according to at least one embodiment of the invention includes a first step which comprises generating a first quad-tree of the enhancement image coding units and prediction units with the highest possible depth with regard to splitting CTBs into CUs and PUs. Then a 4x4 block based inter-layer derivation process is applied. It will be appreciated that other block sizes may be used, for example smaller block sizes for more accuracy. A second step involves a bottom-to-top prediction unit and coding unit merging process, which aims at providing a synthetized quad-tree representation of enhancement CTBs, i.e. with PUs and CUs that are enlarged as much as possible.

A particular embodiment of the first step is described with reference to Figures 26 and 27 Next, a particular embodiment of the second main step will be described with reference to Figure 28. Finally, Figures 29 to 32 illustrate steps of methods associated with the execution of one or more embodiments of the invention.

Figure 27 illustrates a method of deriving prediction information, from a coded base image, for a 4x4 block in the enhancement layer of video sequence data.

In an initial step, an initial representation of enhancement CTBsis constructed. This involves splitting each enhancement CTB 2720 as much as is allowed by the HEVC specification leading to a series of coding units 2721 having an image area size of 8x8 pixels. Each 8x8 coding unit 2721 is given the prediction unit partition type NxN, which means that PUs initially constructed in the enhancement image have all a size equal to 4x4.

For each of these initial 4x4 prediction units, an inter-layer derivation of prediction information is invoked. The interlayer derivation process comprises determining the image area in the base layer that spatially corresponds to the considered enhancement 4x4 PU. Several cases arise with respect to the co-

located spatial area. As specified by the HEVC standard, the minimum size for prediction units is equal to 4x4, and in general a prediction unit size is a multiple of 4, in terms of both width and height. Consequently, a prediction unit PU can be divided into one or several 4x4 blocks in the base coded image 2710 of Figure 27. Figure 27 shows a possible organization of the base image 2710 in terms of prediction units. It also illustrates, for each prediction unit in the base layer, the one or several 4x4 blocks of the enhancement picture that are spatially contained in that prediction unit.

As can be seen, a prediction unit area of the base layer, when spatially up-scaled towards the enhancement layer resolution, overlaps several 4x4 blocks in the enhancement image. Therefore, some of the 4x4 enhancement blocks are fully overlapped by the up-scaled PU area, while some others are partially covered by the up-scaled PU area.

- a fully overlapped 4x4 enhancement block thus may be given the prediction information of the base prediction unit, transformed towards the higher resolution.
- in the case of a partially covered enhancement 4x4 block, several prediction units of the base image area spatially correspond to this 4x4 block. Hence, a partially covered 4x4 blocks may be given prediction information that depends on at least one of the spatially corresponding base prediction units.

Figure 28 schematically illustrates the correspondence between each 4x4 enhancement block being considered, and the respective corresponding co-located spatial area in the base image. As can be seen, the corresponding co-located area in the base image may (1) be fully contained within a prediction unit of the base layer, or may (2) overlap at least two different base prediction units of the base layer.

In the first case (A), the prediction information derivation for the considered 4x4 enhancement block is simplified. It comprises obtaining the prediction information values of the corresponding base prediction unit within which the enhancement block is fully contained, transforming the obtained prediction information values towards the resolution of the enhancement layer, and providing the considered 4x4 enhancement block with the so-transformed prediction information.

In the second case where the co-located base area of current 4x4 enhancement block overlaps two (enhancement block Y) or four 4x4 (enhancement block B), (enhancement block Z), blocks that constitute the prediction units of the base image. The overlapped two or four base blocks may have equal or different prediction information values.

- If the overlapped prediction units of the base image have equal prediction information (the case of enhancement block Z in Figure 28), then the enhancement 4x4 block Z is given that common prediction information, in its up-scaled form.

- Otherwise if the prediction information of the overlapping prediction units differs between the overlapping prediction units (the case of block B or Y in Figure 28), a choice has to be made on the base prediction information to up-scale. In this particular embodiment of the invention, the prediction information of the overlapped base PU that has the highest address, in terms of raster-scan ordering of 4x4 PUs in the base image, is selected and upscaled. i.e. in the case of block Y the prediction information of the right PU covered by the base image area that spatially corresponds to current 4x4 block of the enhancement image is selected and in the case of block B the prediction information of the right-bottom 4x4 PU covered by the base image area that spatially corresponds to current 4x4 block of the enhancement image. The reason for that choice is as follows:

The inter-layer derived prediction information is used later in the coding and decoding of the enhancement image (see Figures 13 and 16). In particular, the inter-layer derived prediction information is used to predict the motion information of the enhancement layer. A predictive coding of enhancement motion vector is performed, which employs in particular the up-sampled motion vectors as reference values to predict the motion vectors of the enhancement image.

More generally the predictive coding of motion vectors in HEVC involves a list of motion vector predictors. These predictors correspond to the motion vectors of already coded PUs, among the spatial and temporal neighbouring PUs of a current PU. In the case of scalable coding, the list of motion vector predictors is enriched: the inter-layer derived motion vector for each enhancement PU is appended to the list of motion vector predictors for that PU.

To emphasize the efficiency of motion vector prediction, it is advantageous to have a list of motion vector predictor which is diversified in terms

of motion vector predictor values. Therefore, one way to favour the diversity of motion vectors contained in such a list in the prediction of enhancement layer's motion vectors is to employ the motion vector of the right-bottom co-located PU in the base layer, when dealing with the prediction of an enhancement PU's motion vector(s).

Once a strategy has been decided on how to derive the prediction information of the enhancement 4x4 PU that spatially overlaps one or several 4x4 base PUs, each 4x4 enhancement PU may be provided with some inter-layer derived prediction information.

Note that the 4x4 base PUs do not necessarily corresponds to encoded PUs. When a base PU is larger than 4x4, it is considered to be constituted of several 4x4 elementary PU, each elementary PU having the same prediction information than the original larger PU.

Figure 29 schematically illustrates a method of grouping the prediction information derived for each 4x4 PU (prediction unit) of the enhancement image according to an embodiment of the invention. The illustrated process can be used to group the prediction information resulting from the method of deriving prediction information described with reference to Figures 27 and 28.

The grouping process according to this embodiment of the invention, also referred to as a merging process in what follows, comprises successively grouping PUs together into one or more coding units (CU), and then in grouping coding units together in order to form one or more CTBs, where possible. Neighbouring prediction units can be merged together if they have identical prediction information derived from the base layer. In the same way, neighbouring coding units can be merged together if all the prediction units that it contains have been merged, i.e. they all have the prediction unit type 2Nx2N, and also have equal prediction information.

These successive elementary PU and then CU merging steps are first applied on the CU and PU organization of enhancement CTBs issued from the process of generating a first quad tree of the enhancement image coding units and prediction units with the highest possible depth as previously described. These quad-tree representations are initially made of the CU and PU with the smallest allowed size, i.e. 8x8 for a CU and 4x4 for a PU. This initial state is illustrated in Figure 29 (a).

The first elementary merging step S2901, involving the merging of prediction units, is first applied to each 8x8 CU. This step comprises testing, for each 8x8 CU, if all PUs contained in the considered CU have equal prediction information values. If it is the case than all the PUS of that CU have equal prediction information values, then the PU can be merged together. This merging consists in providing the concerned 8x8 CU with the prediction unit type 2Nx2N, which means that the CU now comprises only one PU, and may be considered as one PU. The prediction formation associated with that 2Nx2N PU is then equal to the common prediction information of the four merged PUs. The result of this PU merging step applied on initial CUs is illustrated in Figure 29(b). It can be seen that the CUs that initially contained four identical PUs now contain only one PU. On the opposite, CU that contained different PUs in terms of prediction information remain in their initial state.

The next merging step S2902 is illustrated in Figure 29(c) and proceeds on the coding unit level. This step sets out to merge neighbouring CUs together, where this is possible. To do so, for each CTB, it considers 4 neighbouring CUs that form leafs of the quad-tree representation of the considered CTB. It tests if the four neighbouring CUs have all a prediction unit type equal to 2Nx2N. If so, this means that these CUs are homogeneous in terms of prediction information. If the test is positive, then the process tests if the four considered CUs have identical prediction information. If that is the case, it is possible to merge these 4 CUs together, in order to form a single merged CU that spatially covers the four initial CUs. If this is not the case, then the 4 CUs are left in their initial state. An exemplary result of this CU merging process is illustrated in Figure 29(c).

Figure 30 is a flow chart illustrating steps of a global algorithm according to an embodiment of the invention successively applied to CTBs of an enhancement image.

The inputs to this algorithm include the following:

- a current Coding Unit *currCU* of the enhancement image (enhancement CU) for processing.
- the prediction information contained in the coded base image, for each CTB of the base image.

The output of the algorithm of Figure 30 includes a set of derived prediction information for the considered enhancement CTB.

The first part of the algorithm consists in recursively constructing the quad-tree splitting of the current enhancement CTB into coding units with the smallest allowed size 8x8. This takes the following form. The current enhancement CTB is initialized before division into CUs or PUs and thus has a depth value equal to 0. The algorithm first tests in step S3001 if the depth of the input Coding Unit *currCU* is strictly less than the maximum depth allowed for a coding unit. In practice, this maximum depth is typically equal to 5 for a CTB of size 64x64. If the test is positive, then this signifies that the current CU has a size which is greater than the minimum size allowed for a CU. In that case, the next step S3002 of the algorithm comprises splitting the current CU *currCU* into 4 sub coding units, noted as *subCU*. This splitting leads to a set S of four coding units {$subCU_0$, subCU3} each *subCU* having a depth value incremented by 1 compared to the input CU *currCU*. Then, in the next step S3003, the algorithm begins performing a loop on the set of d sub coding units {$subCU_0$, $subCU_3$}. For each *subCU*, the current algorithm is called in a recursive way, with the new sub coding unit as the input parameter.

In this way, the initial input coding units *currCU* are recursively divided into sub coding units *subCU*, until all the coding units resulting from this splitting step have the highest allowable depth value, i.e. the smallest allowable size (8x8).

If it is indicated in initial step S3001 that the current input coding unit *currCU* has the maximum allowed depth value, then the algorithm proceeds to step S3008 to perform a prediction information inter-layer derivation process for the smallest coding unit.

To do so, the following is applied. First, in step S3008 the input coding unit is given the prediction unit type NxN, which signifies that the input CU is initially divided into 4 PUs, each with a size equal to 4x4. Then from step S3009 a loop is performed on the four 4x4 image blocks corresponding to the so-obtained PUs. For each successively considered 4x4 block b_{4X4}, the inter-layer derivation of the prediction information for that 4x4 block b_{4X4} from the base layer is performed in step S3010. This derivation process is described in more detail with reference to Figure 31. Once this is done, current 4x4 block is assigned prediction information derived from the base layer. Once the loop of the four 4x4 enhancement blocks b_{4X4} of current CU *currCU* is terminated i.e. it is determined in step S3012 that the last b_{4X4} block has been processed, the process proceeds to step 3013 which comprises merging the obtained 4x4 prediction units, if possible, as previously

described with reference to Figure 29. For step S3013 the algorithm of Figure 32 is executed to merge the derived 4x4 prediction units contained in the current *currCU*.

Once this PU merging step is done, the algorithm of Figure 30 comes to an end. This means that its returns to the process that called the current algorithm with the current 8x8 coding unit as the input parameter and returns to step S3004 that performed a recursive call to the algorithm, during the loop on the sub coding unit *subCU* contained in the set S. Once this loop is done, 4 sub coding units *subCU* are available, in their inter-layer derived state. This means the PUs contained in these sub CU all contain some prediction information derived from the base layer. Step S3007 then comprises merging the four sub CUs $\{subCU_0, \ subCU_3\}$ if possible, as previously described with reference to Figure 29(c). The CU merging step will be described in more detail with reference to Figure 33.

In this way, a bottom to top CU merging process is performed, until a synthetized quad-tree representation of the considered enhancement CTB is obtained.

Figure 31 is a flow chart illustrating a method for performing inter-layer derivation of prediction information at the 4x4 enhancement block level, according to an embodiment of the invention.

The inputs to this algorithm include the current 4x4 enhancement block 3100 being processed by the algorithm of Figure 30, as well as the prediction information contained in the base image.

An initial step of the algorithm includes testing if the spatial area of the base image corresponding to current 4x4 enhancement block is entirely located within a 4x4 PU of the base image. If the scaling ratio between the base layer resolution and the enhancement layer resolution is equal to 1.5, then in this embodiment the mathematical corresponding to this test is in step 3101 of Figure 31 as " $b_{4x4}.x$ mod 3≠ 1 and $b_{4x4}.y$ mod 3≠ 1 ? ".

If the test is positive, then the next step S3102 comprises determining the 4x4 PU in the base image that contains the co-located area of current enhancement block b_{4x4}. Next, the prediction information of the 4x4 base PU so found is derived towards the spatial resolution of the enhancement layer. This derivation takes into account the spatial scaling ratio that links the base and the enhancement layer. Once the so-derived prediction information is assigned to the

enhancement 4x4 PU, in step S3103 and the algorithm of Figure 31 ends in step S3120.

Else, if the first test S3101 of the algorithm is negative then this signifies that the input enhancement 4x4 block of the enhancement layer overlaps several 4x4 base PUs of the base layer. Then in step S3104 the algorithm tests if the enhancement 4x4 block overlaps several 4x4 base PUs in the horizontal direction. With respect to the 1.5 scaling ratio for example, this corresponds to the expression: "$b_{4x4}.x$ mod 3=1?". If this is not the case, then this means the enhancement block overlaps exactly two 4x4 base PUs in the vertical direction. In that case, the algorithm proceeds to step S3105 and obtains the two co-located 4x4 PUs and considers the bottom PU, among the two concerned base PUs. In step S3106 the prediction information of this bottom PU *bottom_base.predinfo* is up-sampled towards the resolution of the enhancement layer and then is assigned to the current 4x4 block in the enhancement image.

If the preceding test of step S3104 is positive, then this signifies that the enhancement 4x4 block overlaps at least two base 4x4 PUs in the horizontal direction. Then a test of step S3107 determines if the enhancement block also overlaps two base 4x4 PUs in the vertical direction. This is given by the expression "$b_{4x4}.y$ *mod 3=0 OR* $b_{4x4}.y$ *mod 3=2*" in the case of the 1.5 scaling ratio. If this test is positive, then this means the co-located spatial area of the enhancement 4x4 block overlaps 4 base 4x4 PUs in the base image and are obtained in step S3108. In that case, the right bottom PU *right_bottom_base* among those four overlapped 4x4 base PUs is selected in step S3110. Its prediction information *right_bottom_base.predinfo* is up-sampled, and is given to the current 4x4 block in the enhancement image.

Finally, if the test of step S3107 is negative, then this means the co-located spatial area of current enhancement 4x4 block spatially intersects exactly two 4x4 PU of the base image, in the horizontal direction. The two co-located PUs are obtained in step S3109 and then the right side PU among the two concerned base PU is selected in step S3111. Its prediction information is up-sampled and given to the current 4x4 block in the enhancement image.

When the current enhancement 4x4 block b_{4x4} has been given some up-sampled prediction information values, then the algorithm of Figure 31 ends in step S3120.

It will be appreciated that in other embodiments of the invention other choices of the block supplying the prediction information may be envisaged in steps S3106 and S3120. In one particular embodiment of the invention the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected. The objective of maximising the diversity is mainly to improve the compression performance of the motion information coding. For example one diversity criterion could be to privilege the elementary prediction unit providing the motion information enabling the greatest minimisation of the coding cost of the motion information of said processing block. A solution to determine this elementary prediction unit could be to select the elementary prediction unit providing the motion information differing the most from motion information values contained in the list of motion information predictors associated with said processing block. This determination could be based on the computation of the variance of the motion vectors of the list of motion information predictors obtained when adding the motion information of the determined elementary prediction unit to the list of motion information predictors associated with said processing block.

Figure 32 is a flow chart illustrating a method of merging prediction units according to an embodiment of the invention. This process may be implemented by step S3013 of Figure 30 in order to merge 4 neighbouring prediction units of the enhancement layer, once they have been assigned inter-layer inherited prediction information. The input to the algorithm of Figure 32 includes the current coding unit *currCU* of the enhancement image, which contains the prediction units to be merged together.

An initial step S3201 of the algorithm includes assigning the prediction unit type NxN to the current CU *currCU*, as its initial prediction unit type. Next, in step S3202 the algorithm tests if the four PUs contained in *currCU* have identical prediction information values. If this is not the case, then the algorithm of Figure 32 ends in step S3204. Otherwise, if the four PUs contained in *currCu* have identical prediction values, the algorithm assigns the prediction unit type 2Nx2N to the current coding unit *currCU* in step S3203. This means the coding unit is now made of only one prediction unit, whose prediction information is equal to the common prediction information of the four initial prediction units. Once the considered

enhancement CU is in merged state (2Nx2N PU type), then the algorithm of Figure 32 ends in step S3204.

Figure 33 is a flow chart illustrating a method of merging derived coding units according to an embodiment of the invention. This method may be performed by step S3007 of Figure 30, in order to merge 4 neighbouring coding units of the enhancement layer, if this is possible. The input 3300 to this algorithm includes the set of coding units S={*subCU0,...subCU3*}, as well as the coding unit *currCU* that embeds these four CUs, previously introduced with reference to Figure 30. The first step S3301 of the algorithm of Figure 33 tests if all the coding units contained in the set of coding units S have a prediction unit type equal to 2Nx2N. If this is not the case, then the 4 neighboring coding units are not all equal, and hence cannot be merged together. Consequently the algorithm of Figure 33 ends in step S3305.

Otherwise if it is determined in step S3301 that all the coding units of set S have prediction type 2Nx2N then the algorithm proceeds to the next testing step S3302, which determines if the coding units contained in set S have equal prediction information. If this is not the case, then the 4 neighboring coding units are not all equal, and hence cannot be merged together, and so the algorithm of Figure 33 ends in step S3305. Otherwise, it is determined that the 4 considered coding units are strictly identical in terms of PU type and prediction information, then the algorithm proceeds to step S3304 to process the merging of the 4 considered coding units. This simply comprises decrementing the depth value of each coding unit contained in the set S, and then in providing the embedding coding unit *currCU* with the prediction unit type 2Nx2N. Once this is done, the algorithm ends in step S3305.

Embodiments of the invention thus provide ways of enabling motion information to be inherited from the base layer, during the coding of the enhancement layer, in the scope of an HEVC enhancement layer for example, while capturing as much as motion information as possible from the base layer. Some embodiments may be applied in particular in the case of spatial scalability with a scaling ratio equal to 1.5 between base and enhancement layer.

Although the present invention has been described hereinabove with reference to specific embodiments, the present invention is not limited to the specific embodiments, and modifications will be apparent to a skilled person in the

art which lie within the scope of the present invention. In particular the different features from different embodiments may be interchanged or combined, where appropriate.

For example, although some embodiments of the invention have been described with respect to their application in HEVC coding it will be appreciated that embodiments of the invention may be applied to other coding methods and standards.

Many further modifications and variations will suggest themselves to those versed in the art upon making reference to the foregoing illustrative embodiments, which are given by way of example only and which are not intended to limit the scope of the invention, that being determined solely by the appended claims.

In the claims, the word "comprising" does not exclude other elements or steps, and the indefinite article "a" or "an" does not exclude a plurality. The mere fact that different features are recited in mutually different dependent claims does not indicate that a combination of these features cannot be advantageously used.

CLAIMS

1. A method of determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the step of deriving enhancement layer prediction information comprising for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, deriving the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, deriving the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected according to a predetermined criterion;

wherein the predetermined criterion is based on at least one of the relative location of said one of said plurality of elementary prediction units with respect to the other elementary prediction units of said plurality of elementary prediction units and the base layer prediction information of the said elementary prediction units.

2. A method according to claim 1, wherein the predetermined criterion is based on the raster scan ordering of the plurality of the elementary prediction units.

3. A method according to claim 2 wherein the predetermined criterion determines that the prediction information of the last elementary prediction unit in raster scan order is selected.

4. A method according to any preceding claim, wherein if the plurality of elementary prediction units are distributed in a horizontal direction the predetermined criterion determines that the prediction information of the elementary prediction unit located most right with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

5. A method according to any one of claims 1 to 3, wherein if the plurality of elementary prediction units are distributed in a vertical direction, the predetermined criterion determines that the prediction information of the bottom elementary prediction unit with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

6. A method according to any one of claims 1 to 3, wherein if the plurality of elementary prediction units are distributed in both vertical and horizontal directions the predetermined criterion determines that the prediction information of the elementary prediction unit located at the right bottom part with respect to the other elementary prediction units of said plurality of elementary prediction units..

7. A method according to any preceding claim where if the base layer prediction information of the elementary prediction units corresponding to said region is the same for said elementary prediction units, the enhancement prediction information is derived from said common base layer prediction information.

8. A method according to any preceding claim, wherein the predetermined criterion determines that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected

9. A method according to any preceding claim wherein the spatial scaling between an image of the enhancement layer and a corresponding image of the base layer is a non-integer ratio

10. A method according to claim 9 wherein the non-integer ratio is 1.5.

11. A method according to any preceding claim wherein the corresponding base layer image is a base layer image temporally coincident with the enhancement layer image.

12. A method according to any preceding claim, wherein the processing block has a 2NX2N pixel size, N being an integer.

13. A method according to claim 12, wherein the processing block has a 4X4 pixel size.

14. A method according to any preceding claim further comprising grouping together, into a common processing unit for encoding, a plurality of neighbouring processing units having the same enhancement layer prediction information.

15. A method according to the preceding claim wherein the common processing unit is considered as a processing unit of 2Nx2N size

16. A method according to claim 14 or 15, further comprising grouping together, into a larger common processing unit for encoding, a plurality of neighbouring common processing units having the same enhancement layer prediction information.

17. A method according to the preceding claim wherein the larger common processing unit is considered as a processing unit of 2Nx2N size.

18. A method according to any one of claims 14 to 17, wherein the depth value of each common processing unit of the larger common processing unit is decremented.

19. A method according to any preceding claim comprising a prior step of partitioning the enhancement layer into coding units such that each coding unit has the highest possible depth value; and partitioning each coding unit into said processing blocks.

20. A device for determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information,

the device comprising prediction information derivation means for deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the prediction information derivation means being operable to derive for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected according to a predetermined criterion;

wherein the predetermined criterion is based on at least one of the relative location of said one of said plurality of elementary prediction units with respect to the other elementary prediction units of said plurality of elementary prediction units and the base layer prediction information of the said elementary prediction units.

21. A device according to claim 20, wherein the predetermined criterion is based on the raster scan ordering of the plurality of the elementary prediction units.

22. A device according to claim 21 wherein the predetermined criterion determines that the prediction information of the last elementary prediction unit in raster scan order is selected.

23. A device according to one of claims 20 to 22, wherein if the plurality of elementary prediction units are distributed in a horizontal direction the predetermined criterion is such that the prediction information of the elementary prediction unit located most right with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

24. A device according to any one of claims 20 to 22, wherein if the plurality of elementary prediction units are distributed in a vertical direction, the predetermined criterion is such that the prediction information of the bottom elementary prediction unit with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

25. A device according to any one of claims 20 to 22, wherein if the plurality of elementary prediction units are distributed in both vertical and horizontal directions the predetermined criterion is such that the prediction information of the elementary prediction unit located at the right bottom with respect to the other elementary prediction units of said plurality of elementary prediction units is selected.

26. A device according to any one of claims 20 to 25 where if the base layer prediction information of the elementary prediction units corresponding to said region is the same for said elementary prediction units, the prediction information derivation means is operable to derive the enhancement prediction information from said common base layer prediction information.

27. A device according to any one of claims 20 to 26, wherein the predetermined criterion is such that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected

28. A device according to any one of claims 20 to 27 wherein the spatial scaling between an image of the enhancement layer and a corresponding image of the base layer is a non-integer ratio

29. A device according to claim 28 wherein the non-integer ratio is 1.5.

30. A device according to any one of claims 20 to 29 wherein the corresponding base layer image is a base layer image temporally coincident with the enhancement layer image.

31. A device according to any preceding claim, wherein the processing block has a 2NX2N pixel size, N being an integer.

32. A device according to claim 31, wherein the processing block has a 4X4 pixel size.

33. A device according to any one of claims 20 to 32 further comprising grouping means for grouping together, into a common processing unit for encoding, a plurality of neighbouring processing units having the same enhancement layer prediction information.

34. A device according to the preceding claim wherein the common processing unit is considered as a processing unit of 2Nx2N size.

35. A device according to claim 33 or 34, wherein the grouping means is further configured to group together, into a larger common processing unit for encoding, a plurality of neighbouring common processing units having the same enhancement layer prediction information.

36. A device according to the preceding claim wherein the larger common processing unit is considered as a processing unit of 2Nx2N size.

37. A device according to claim 35 or 36, wherein the depth value of each common processing unit of the larger common processing unit is decremented.

38. A device according to any one of claims 20 to 37 comprising means for partitioning the enhancement layer into elementary prediction units such that each elementary prediction unit has the highest possible depth value; and partitioning each elementary prediction unit into said processing blocks.

39. A method of determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the step of deriving enhancement layer prediction information comprising for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, deriving the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, deriving the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected so that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected.

40. A method according to claim 39 wherein the spatial scaling between an image of the enhancement layer and a corresponding image of the base layer is a non-integer ratio

41. A method according to claim 40 wherein the non-integer ratio is 1.5.

42. A method according to any one of claims 39 to 41 wherein the corresponding base layer image is a base layer image temporally coincident with the enhancement layer image.

43. A method according to any one of claims 39 to 42, wherein the processing block has a 2NX2N pixel size, N being an integer.

44. A method according to claim 43, wherein the processing block has a 4X4 pixel size.

45. A method according to any one of claims 39 to 44 further comprising grouping together, into a common processing unit for encoding, a plurality of neighbouring processing units having the same enhancement layer prediction information.

46. A method according to the preceding claim wherein the common processing unit is considered as a processing unit of 2Nx2N size

47. A method according to claim 45 or 46, further comprising grouping together, into a larger common processing unit for encoding, a plurality of neighbouring common processing units having the same enhancement layer prediction information.

48. A method according to the preceding claim wherein the larger common processing unit is considered as a processing unit of 2Nx2N size.

49. A method according to claim 47 or 48, wherein the depth value of each common processing unit of the larger common processing unit is decremented.

50. A method according to any one of claims 39 to 49 comprising a prior step of partitioning the enhancement layer into elementary prediction units such that each elementary prediction unit has the highest possible depth value; and partitioning each elementary prediction unit into said processing blocks.

51. A device for determining prediction information for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information,

the device comprising prediction information derivation means for deriving from base layer prediction information, enhancement layer prediction information for one or more processing blocks of the enhancement layer; the prediction information derivation means being operable to derive for a processing block:

in the case where a region of the base layer, spatially corresponding to the processing unit, is wholly located within one elementary prediction unit, the enhancement prediction information from the base layer prediction information of said one elementary prediction unit;

otherwise, in the case where a plurality of elementary prediction units are at least partially located in the region of the base layer spatially corresponding to the processing block, the enhancement prediction information from the base layer prediction information of one of said elementary prediction units, selected so that the prediction information of the elementary prediction unit providing the best diversity among motion information values associated with the said processing block is selected.

52. A method of encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the method comprising

determining enhancement layer prediction information for a processing block of the enhancement layer according to the method of any one of claims 1 to 19 or any one of claims 39 to 50; and

encoding the processing unit into an encoded video bitstream using said enhancement layer prediction information.

53. A device for encoding at least part of an image of an enhancement layer of video data from a corresponding base layer image of lower spatial resolution of the

video data, the enhancement layer being composed of processing blocks and the base layer being composed of elementary prediction units each having associated base layer prediction information, the device comprising

a device for determining enhancement layer prediction information for a processing block of the enhancement layer according to the method of any one of claims 20 to 38 or claim 51; and

an encoder for encoding the processing unit into an encoded video bitstream using said enhancement layer prediction information.

54. A computer program product for a programmable apparatus, the computer program product comprising a sequence of instructions for implementing a method according to any one of claims 1 to 19; or 39 to 50; when loaded into and executed by the programmable apparatus.

55. A computer-readable storage medium storing instructions of a computer program for implementing a method, according to any one of claims 1 to 19; or 39 to 50.

56. A method of encoding at least part of image portion substantially as hereinbefore described with reference to, and as shown in Figures 26 to 33.

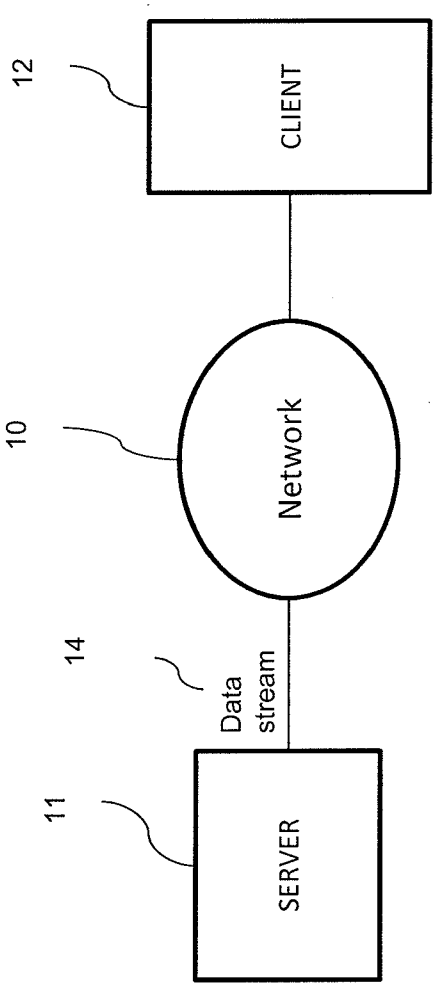


FIG. 1A

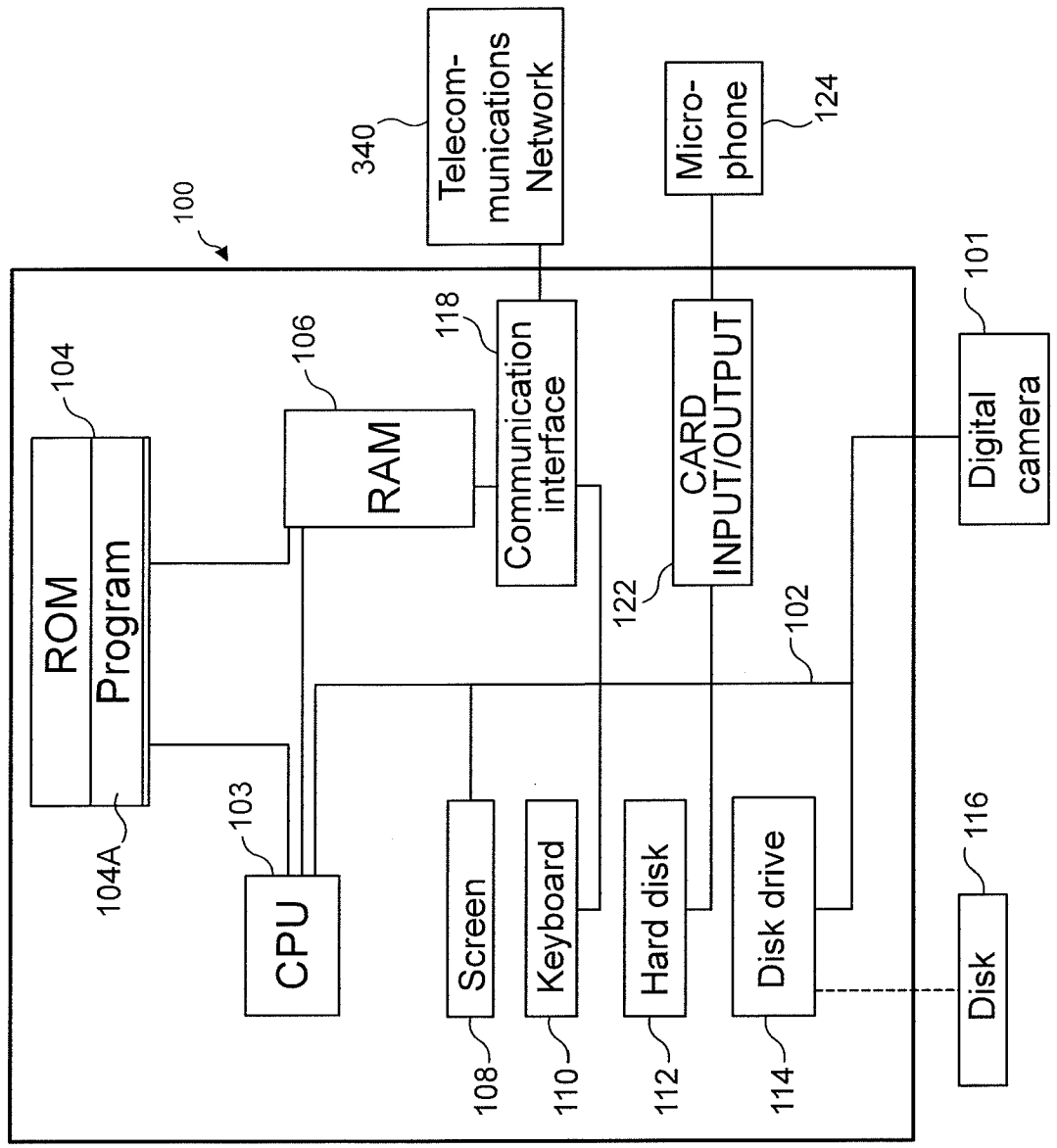

100
340
Telecom-munications Network
Micro-phone
124
101
Digital camera
118
Communication interface
106
RAM
104
ROM
Program
CARD INPUT/OUTPUT
122
102
104A
103
CPU
Screen
Keyboard
Hard disk
Disk drive
Disk
116
108
110
112
114

Fig. 1B

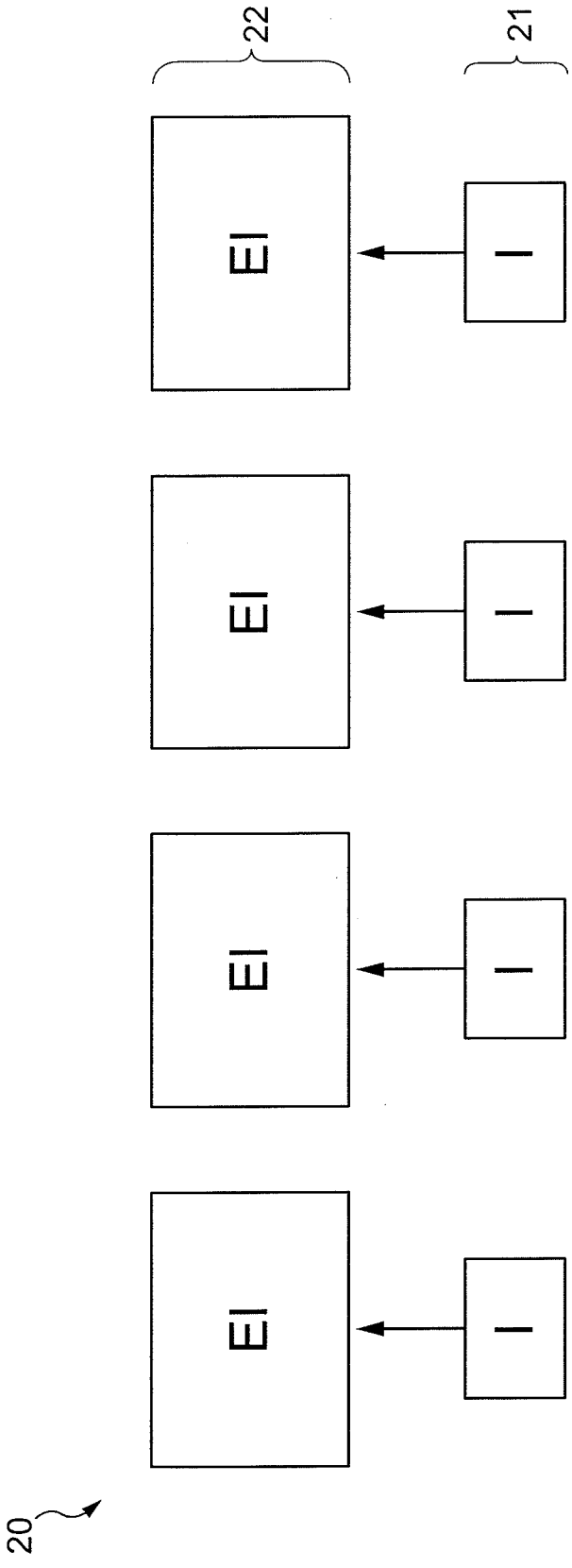


Fig. 2

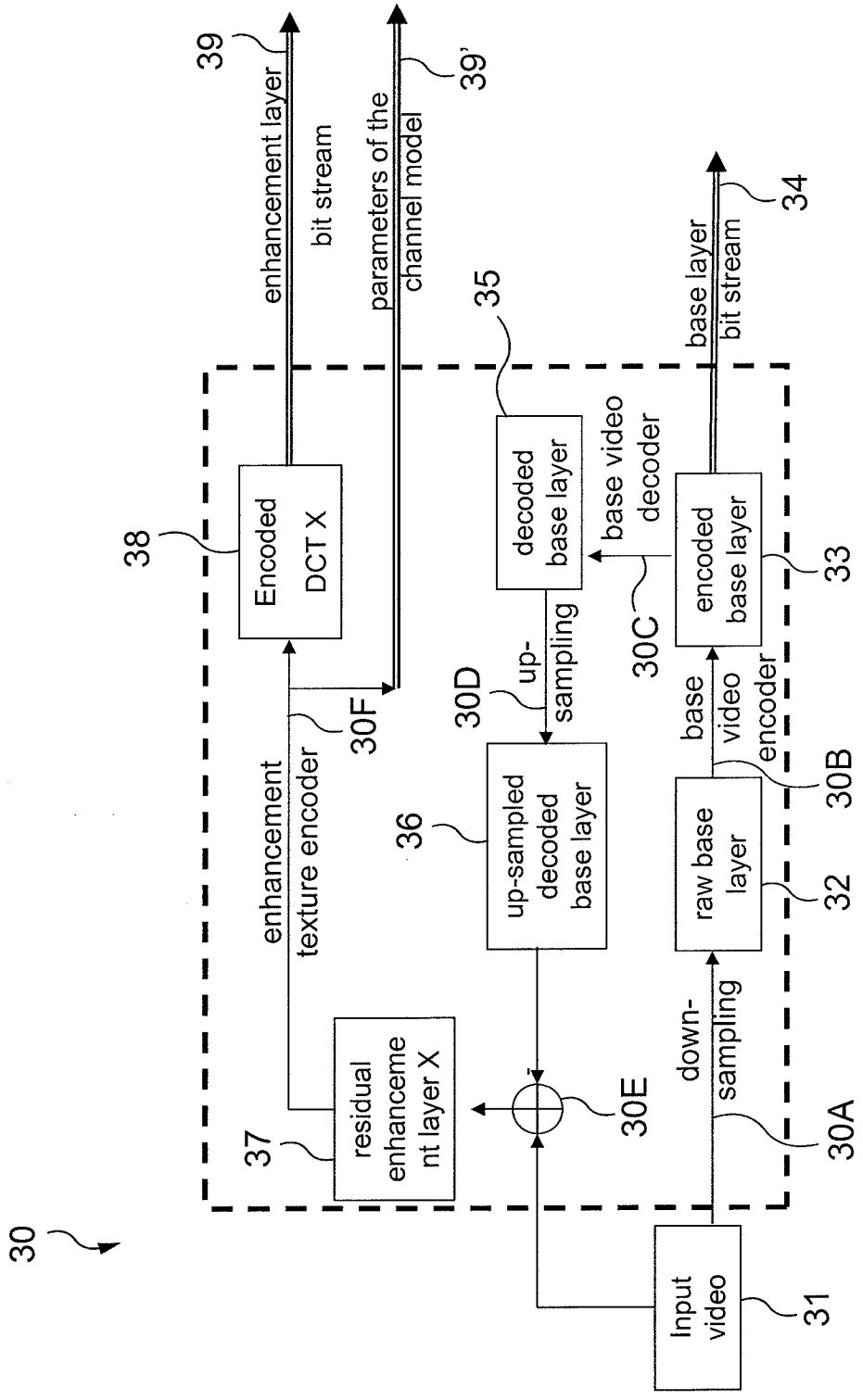

30
Input video
31
residual enhancement layer X
37
enhancement texture encoder
Encoded DCT X
38
39
enhancement layer bit stream
30F
30E
up-sampled decoded base layer
36
30D
up-sampling
decoded base layer
35
base video decoder
30C
raw base layer
32
base video encoder
30B
down-sampling
30A
encoded base layer
33
base layer bit stream
34
parameters of the channel model
39'

Fig. 3

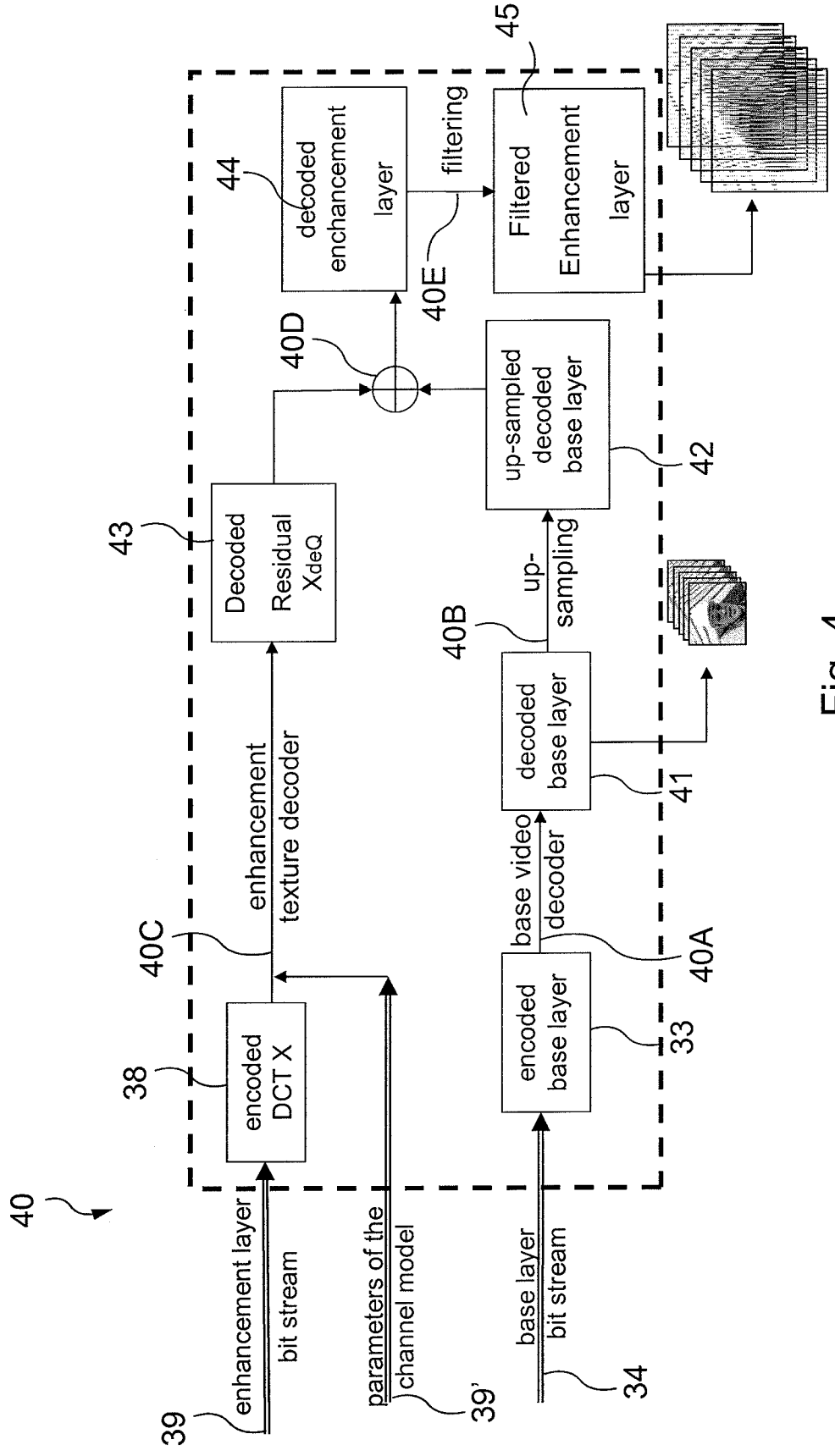

40
39
enhancement layer
bit stream
38
encoded DCT X
40C
enhancement texture decoder
43
Decoded Residual XdeQ
40D
44
decoded enchancement layer
filtering
40E
45
Filtered Enhancement layer
parameters of the channel model
39'
base layer
bit stream
34
encoded base layer
33
base video decoder
40A
decoded base layer
41
40B
up-sampling
up-sampled decoded base layer
42

Fig. 4

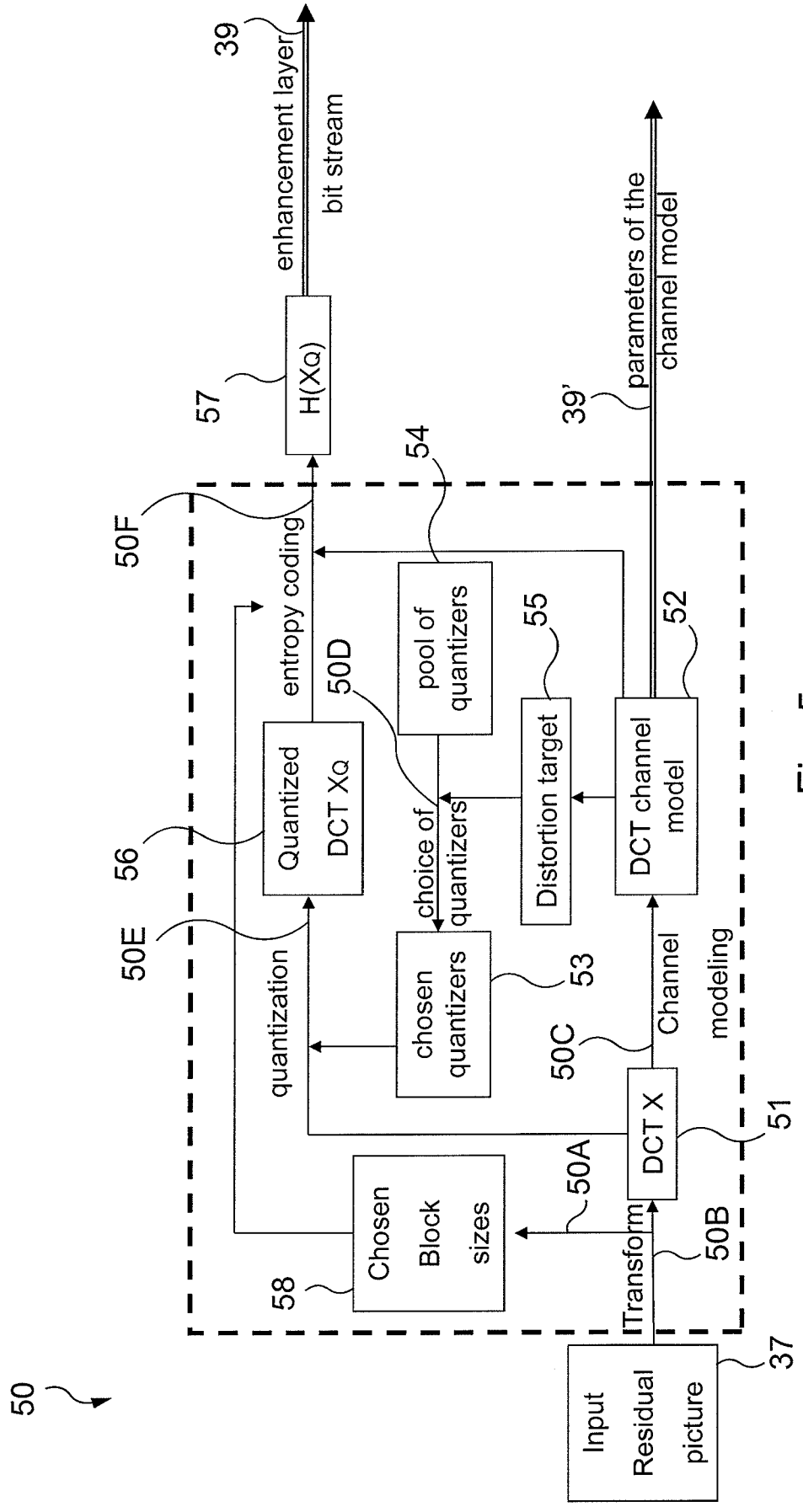

50
Input Residual picture
37
Transform
50B
DCT X
51
50A
Chosen Block sizes
58
quantization
50E
chosen quantizers
53
choice of quantizers
50C
Channel modeling
Quantized DCT Xq
56
50D
pool of quantizers
54
Distortion target
55
DCT channel model
52
entropy coding
50F
H(Xq)
57
enhancement layer bit stream
39
parameters of the channel model
39'

Fig. 5

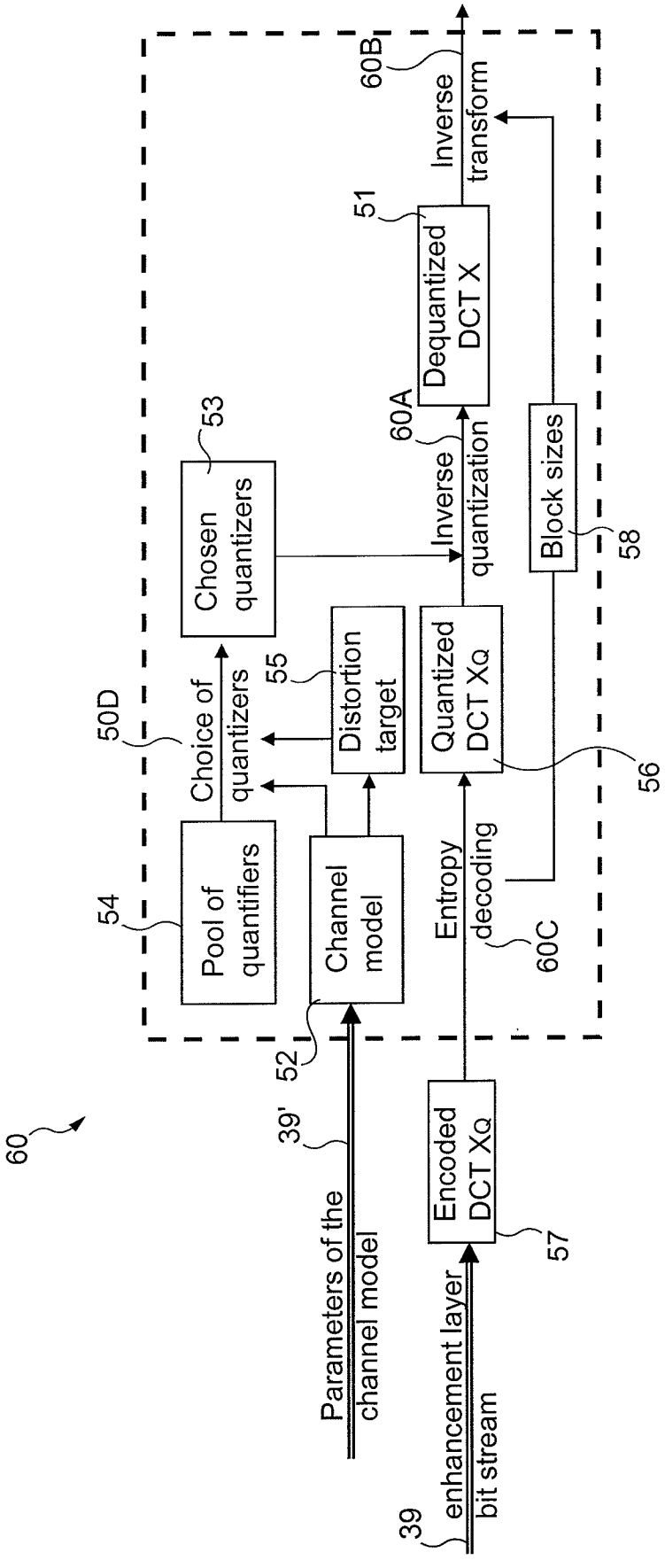

60
39
enhancement layer
bit stream
39'
Parameters of the
channel model
Encoded
DCT Xq
57
52
Pool of
quantifiers
54
Channel
model
Choice of
quantizers
50D
Chosen
quantizers
53
Distortion
target
55
Entropy
decoding
60C
Quantized
DCT Xq
56
Inverse
quantization
60A
Dequantized
DCT X
51
Block sizes
58
Inverse
transform
60B

Fig. 6

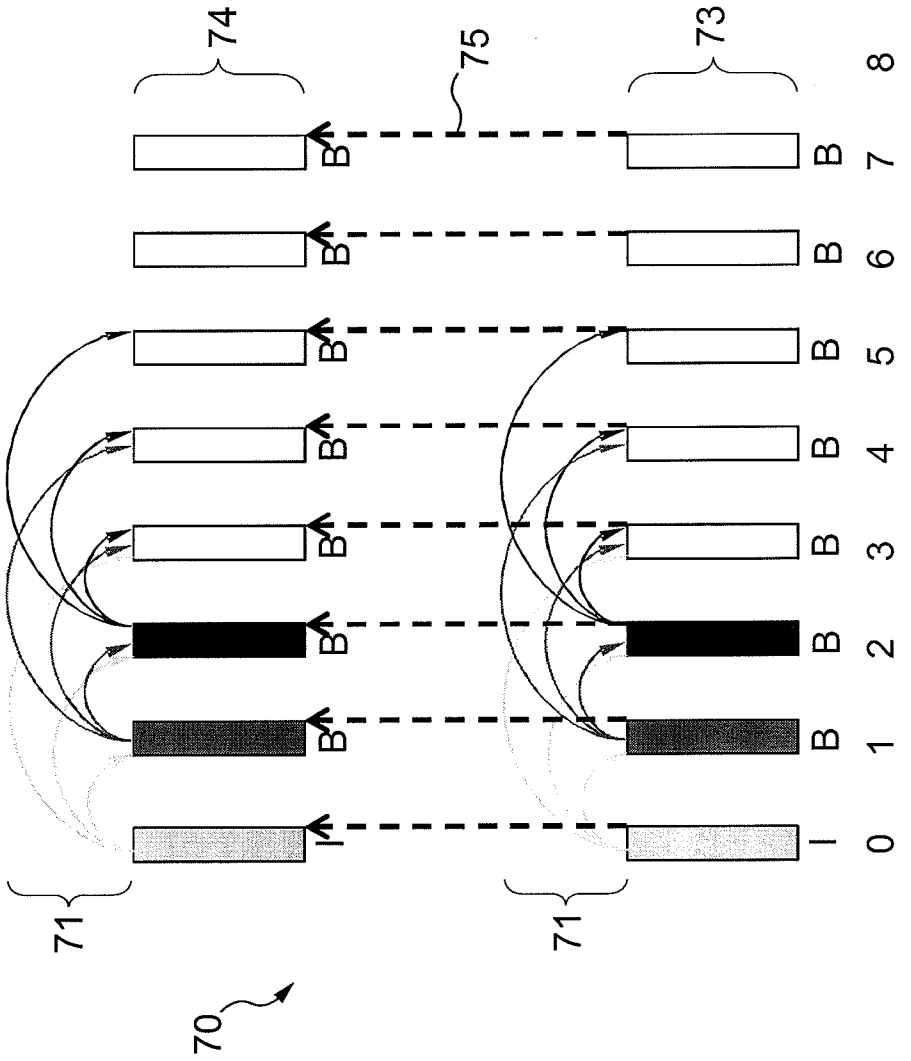
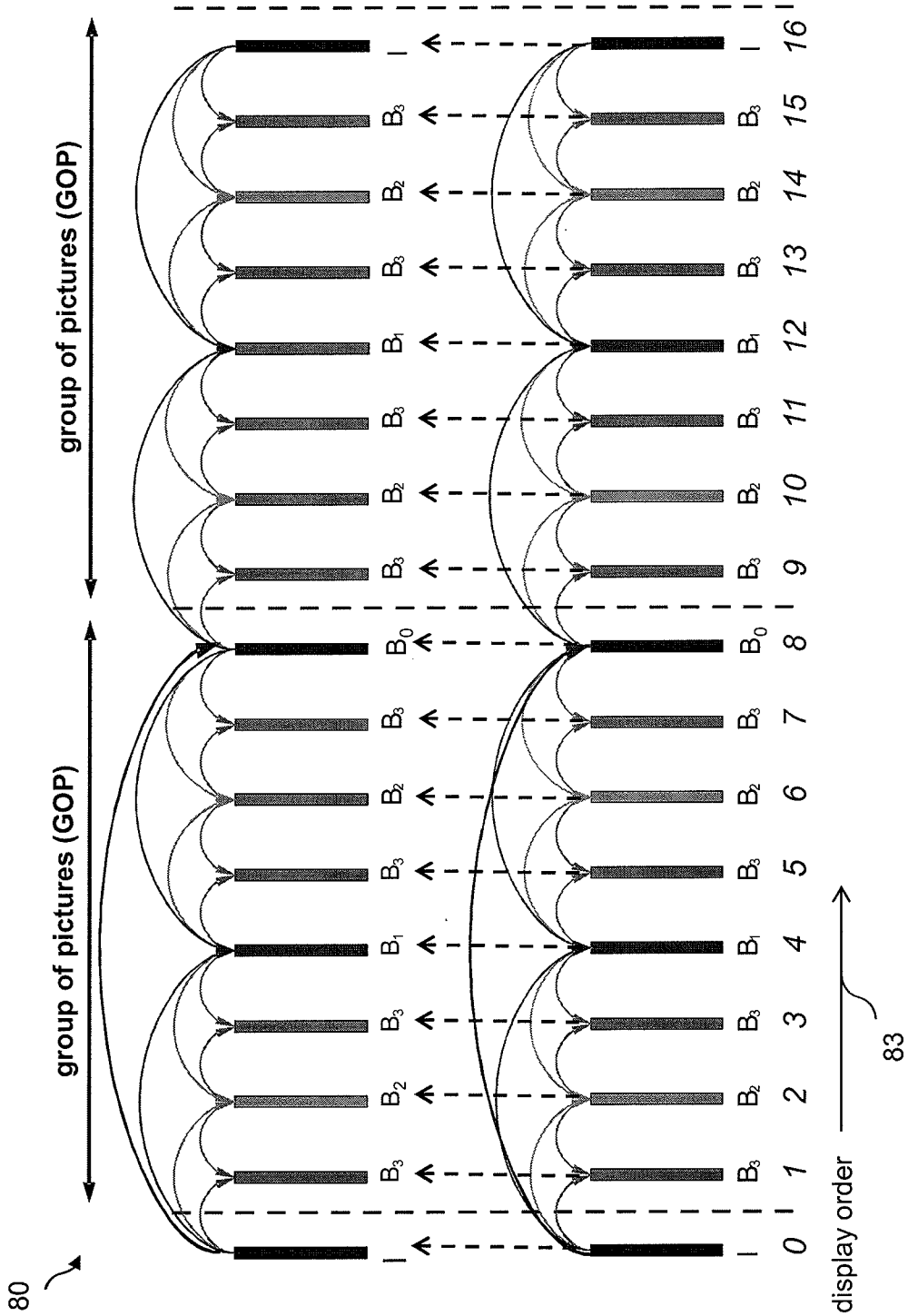


Fig. 7


group of pictures (GOP)
group of pictures (GOP)
80
I
B3
B2
B3
B1
B3
B2
B3
B0
B3
B2
B3
B1
B3
B2
B3
I
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
display order
83

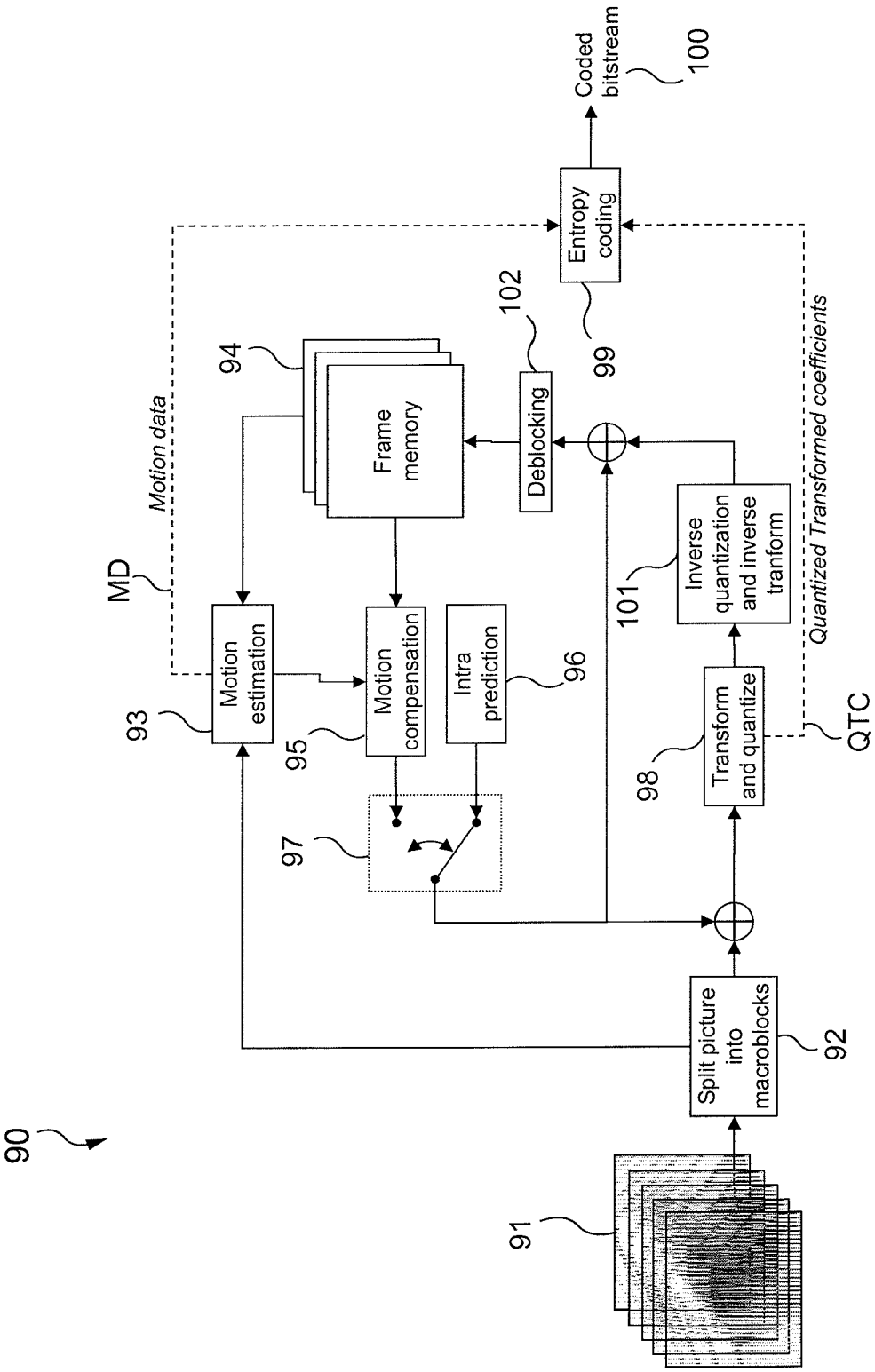

90
91
Split picture into macroblocks
92
Motion estimation
93
94
Frame memory
Motion data
MD
Motion compensation
95
Intra prediction
96
97
Transform and quantize
98
QTC
Inverse quantization and inverse tranform
101
Quantized Transformed coefficients
Deblocking
102
Entropy coding
99
Coded bitstream
100

Fig. 9

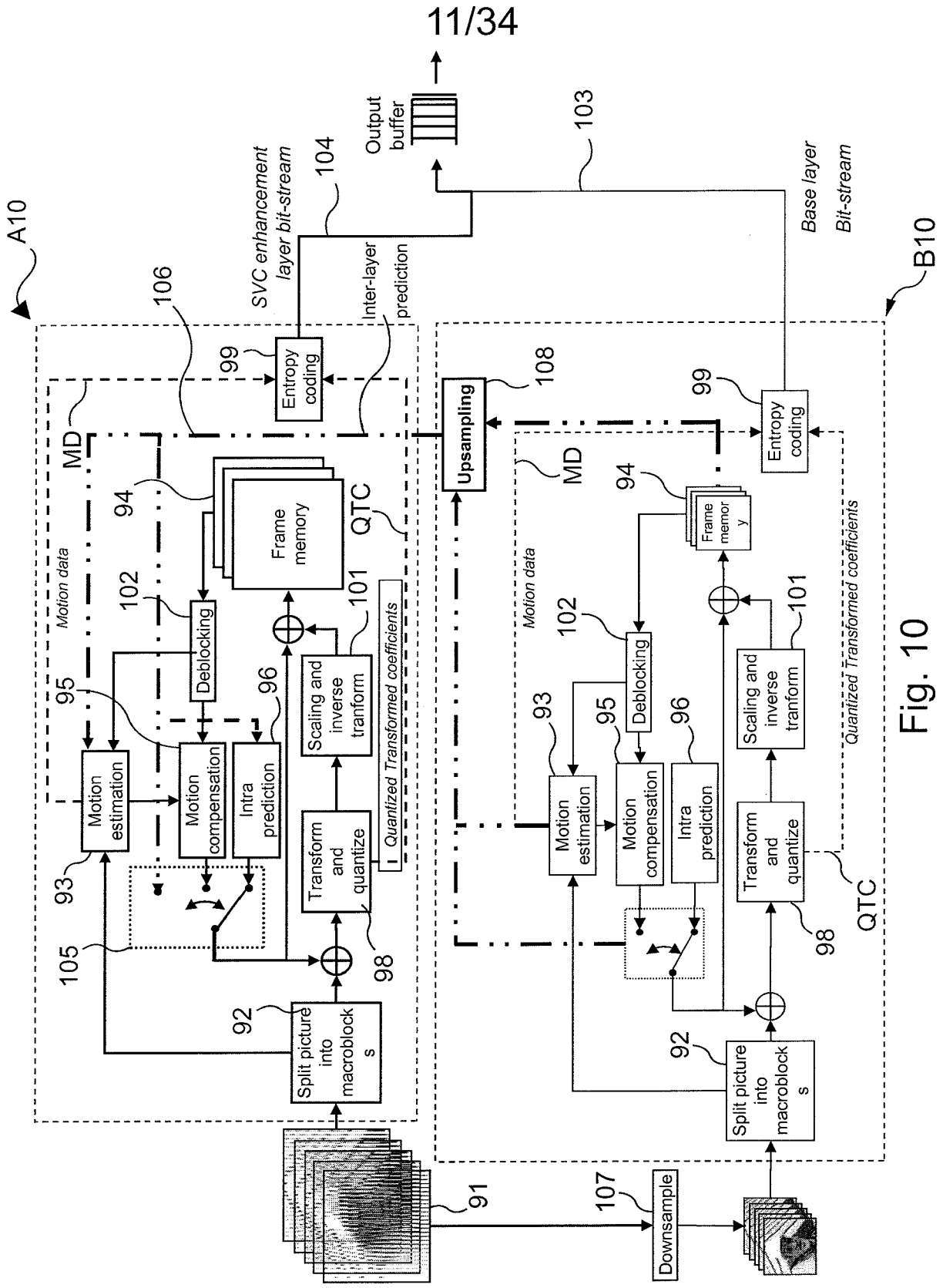

A10
B10
SVC enhancement layer bit-stream
Base layer Bit-stream
Output buffer
103
104
106
Inter-layer prediction
99
Entropy coding
MD
Motion data
94
Frame memory
QTC
Quantized Transformed coefficients
102
Deblocking
95
96
Motion estimation
Motion compensation
Intra prediction
Scaling and inverse tranform
101
93
105
92
Split picture into macroblock s
Transform and quantize
98
Upsampling
108
107
Downsample
91

Fig. 10

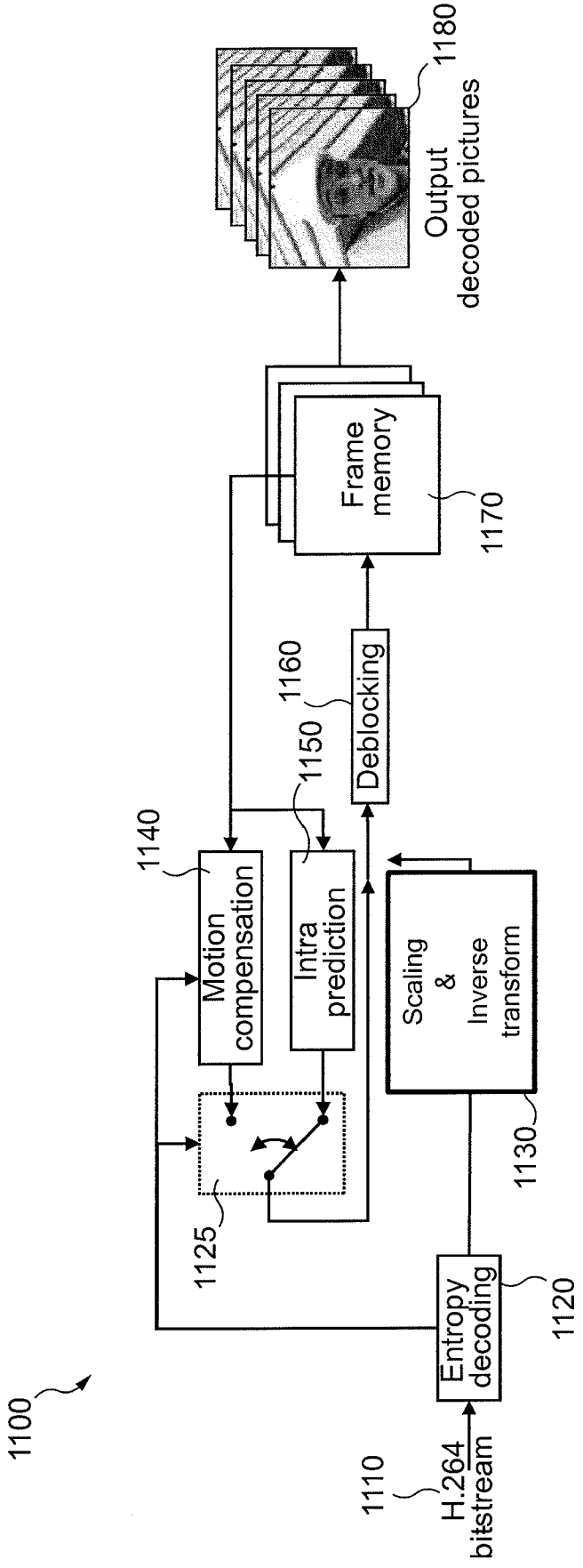

1100
1110
H.264 bitstream
Entropy decoding
1120
1125
1130
Scaling & Inverse transform
1140
Motion compensation
Intra prediction
1150
1160
Deblocking
Frame memory
1170
1180
Output decoded pictures

Fig. 11

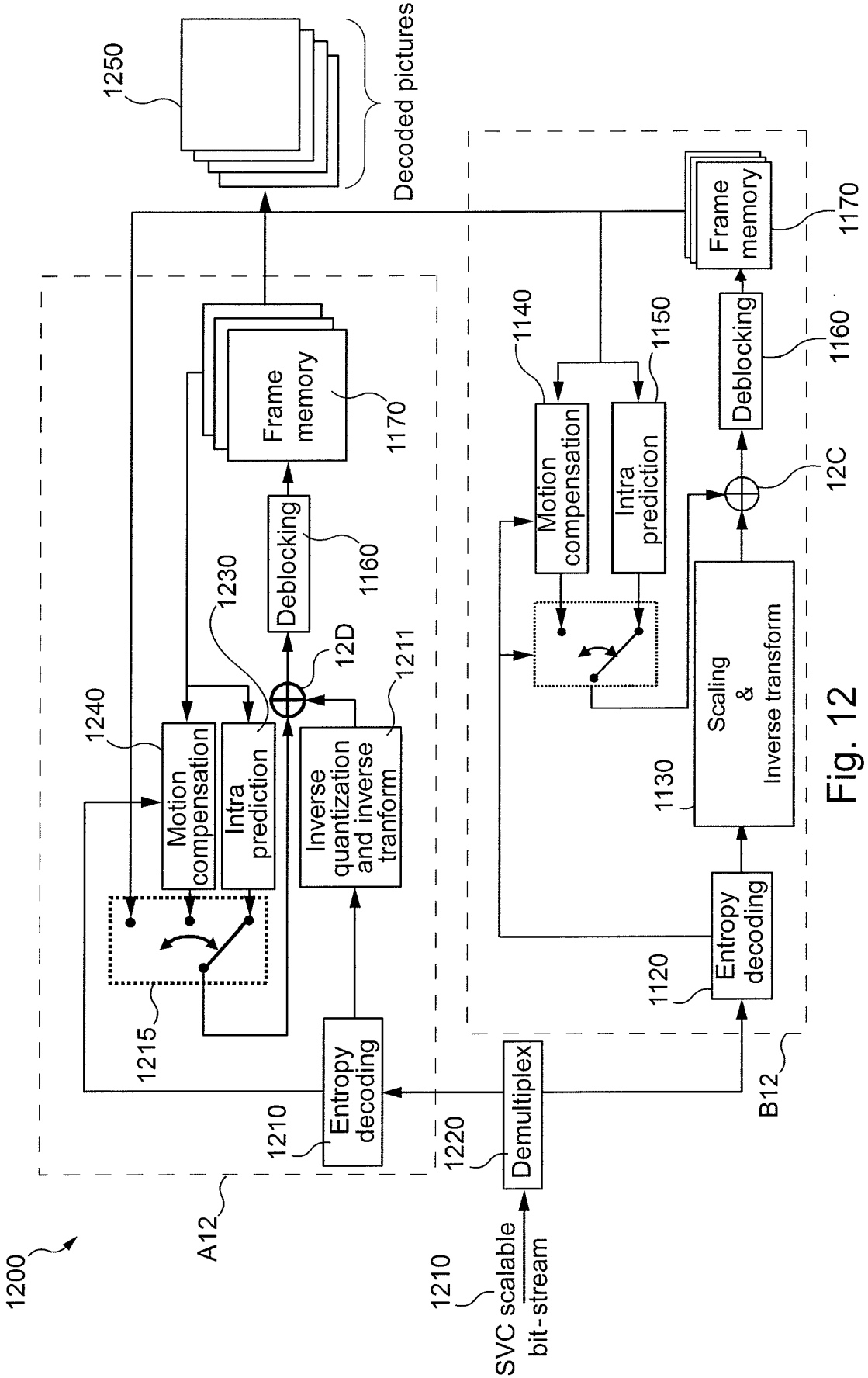

1200
A12
1210
SVC scalable bit-stream
1220
Demultiplex
1210
Entropy decoding
1215
1240
Motion compensation
Intra prediction
Inverse quantization and inverse tranform
1211
12D
1230
Deblocking
1160
Frame memory
1170
1250
Decoded pictures
B12
1120
Entropy decoding
1130
Scaling & Inverse transform
12C
Deblocking
1160
Frame memory
1170
1140
Motion compensation
1150
Intra prediction

Fig. 12

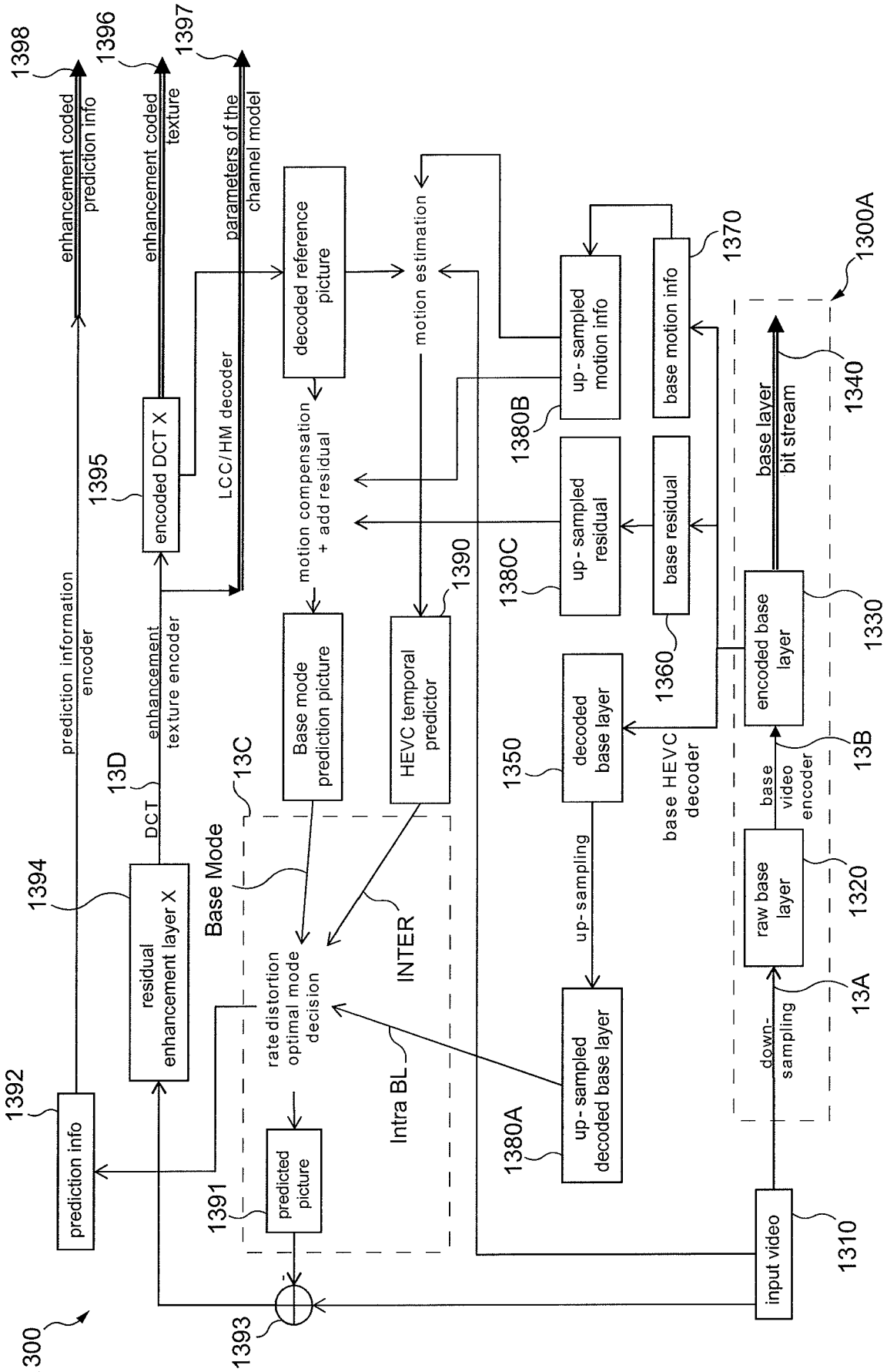

1398
1396
1397
enhancement coded prediction info
enhancement coded texture
parameters of the channel model
decoded reference picture
motion estimation
1395
encoded DCT X
LCC/HM decoder
motion compensation + add residual
1380B
up-sampled motion info
base motion info
1370
1300A
1340
base layer bit stream
1390
1380C
up-sampled residual
base residual
prediction information encoder
enhancement texture encoder
13D
DCT
13C
Base Mode
Base mode prediction picture
HEVC temporal predictor
1360
encoded base layer
1330
1394
residual enhancement layer X
INTER
1350
decoded base layer
base HEVC decoder
base video encoder
13B
rate distortion optimal mode decision
up-sampling
raw base layer
1320
1392
prediction info
Intra BL
1380A
up-sampled decoded base layer
down-sampling
13A
1391
predicted picture
input video
1310
1393
1300

Fig. 13

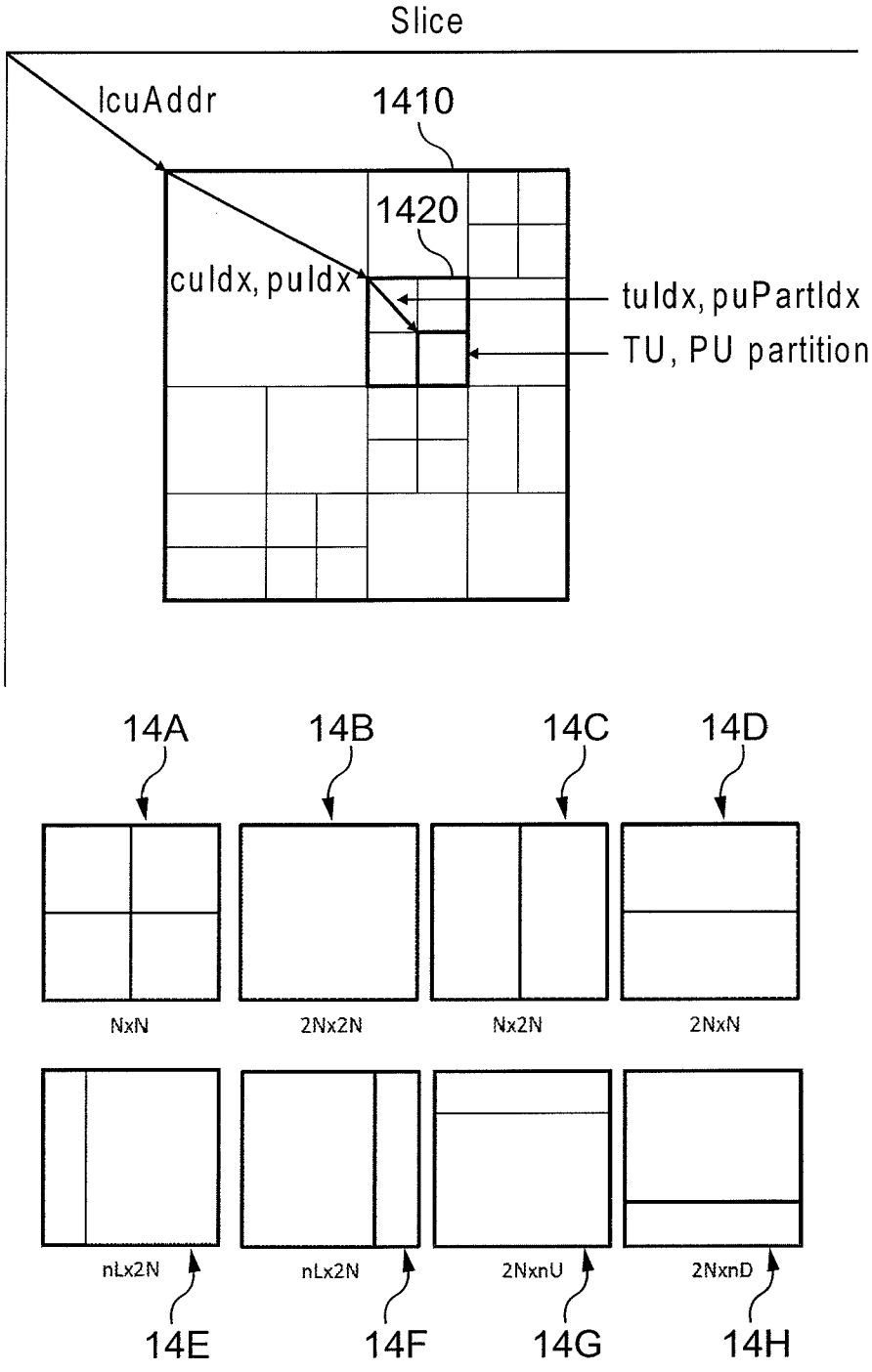

Slice
lcuAddr
1410
1420
culdx, puldx
tuIdx, puPartIdx
TU, PU partition
14A
14B
14C
14D
NxN
2Nx2N
Nx2N
2NxN
nLx2N
nLx2N
2NxnU
2NxnD
14E
14F
14G
14H

Fig. 14

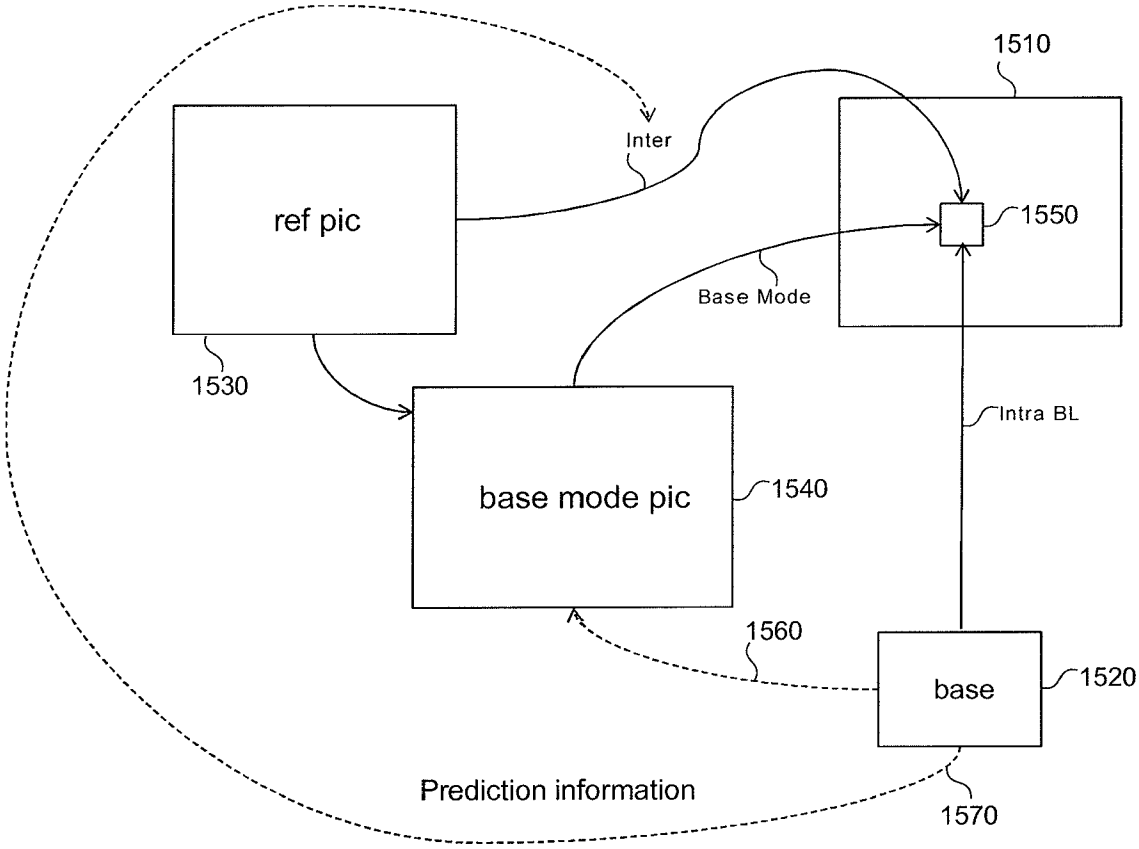

1510
Inter
ref pic
1550
Base Mode
1530
Intra BL
base mode pic
1540
1560
base
1520
Prediction information
1570

Fig. 15

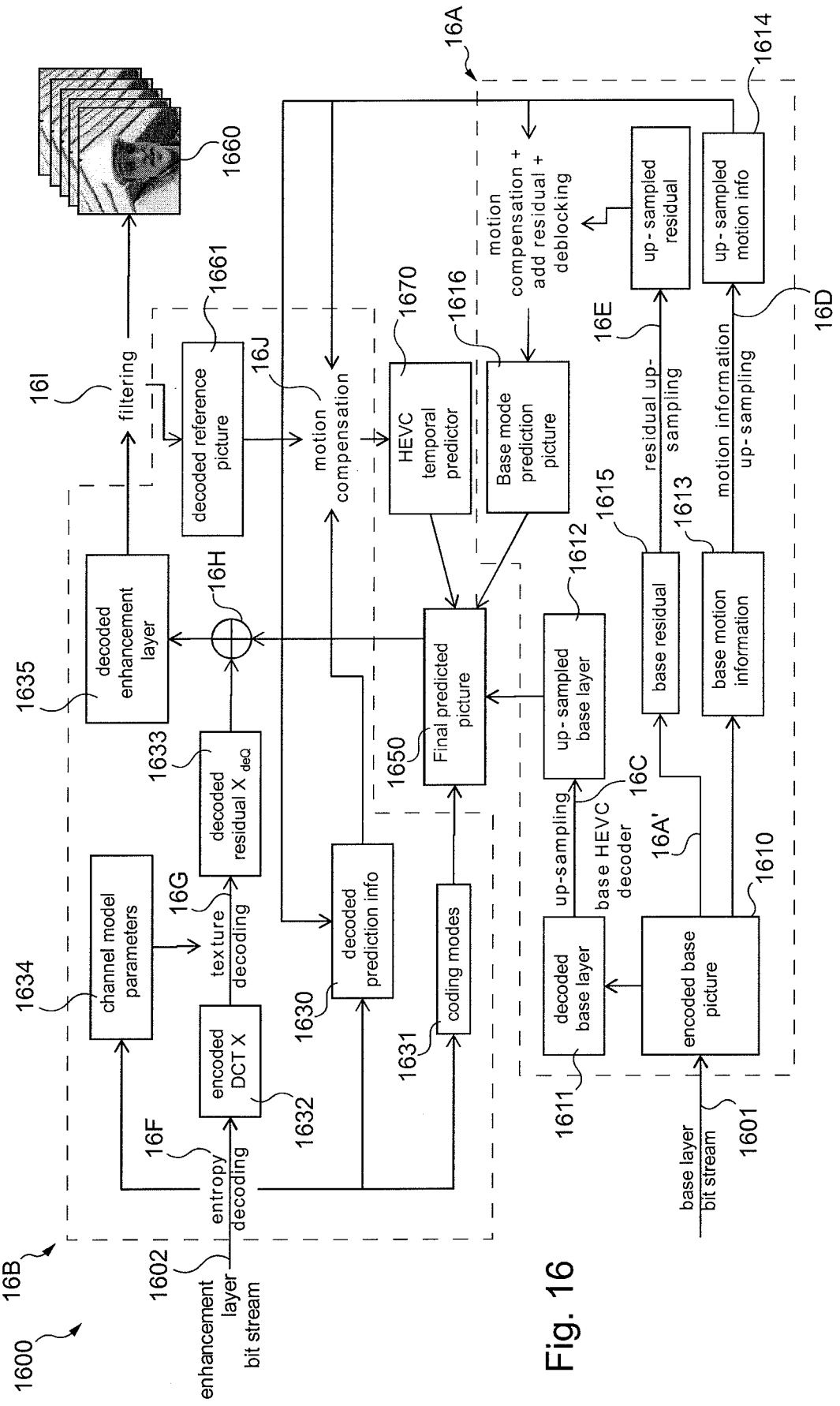

1600
16B
1602
enhancement layer bit stream
entropy decoding
16F
1634
channel model parameters
texture decoding
16G
encoded DCT X
1632
1633
decoded residual X deQ
16H
1635
decoded enhancement layer
filtering
16I
1660
1661
decoded reference picture
16J
motion compensation
1630
decoded prediction info
1631
coding modes
1650
Final predicted picture
HEVC temporal predictor
1670
Base mode prediction picture
1616
motion compensation + add residual + deblocking
16A
1614
up-sampled motion info
16D
motion information up-sampling
1610
base motion information
1612
up-sampled base layer
1615
base residual
1613
residual up-sampling
16E
up-sampled residual
16A'
16C
base HEVC decoder
up-sampling
1611
decoded base layer
encoded base picture
1601
base layer bit stream

Fig. 16

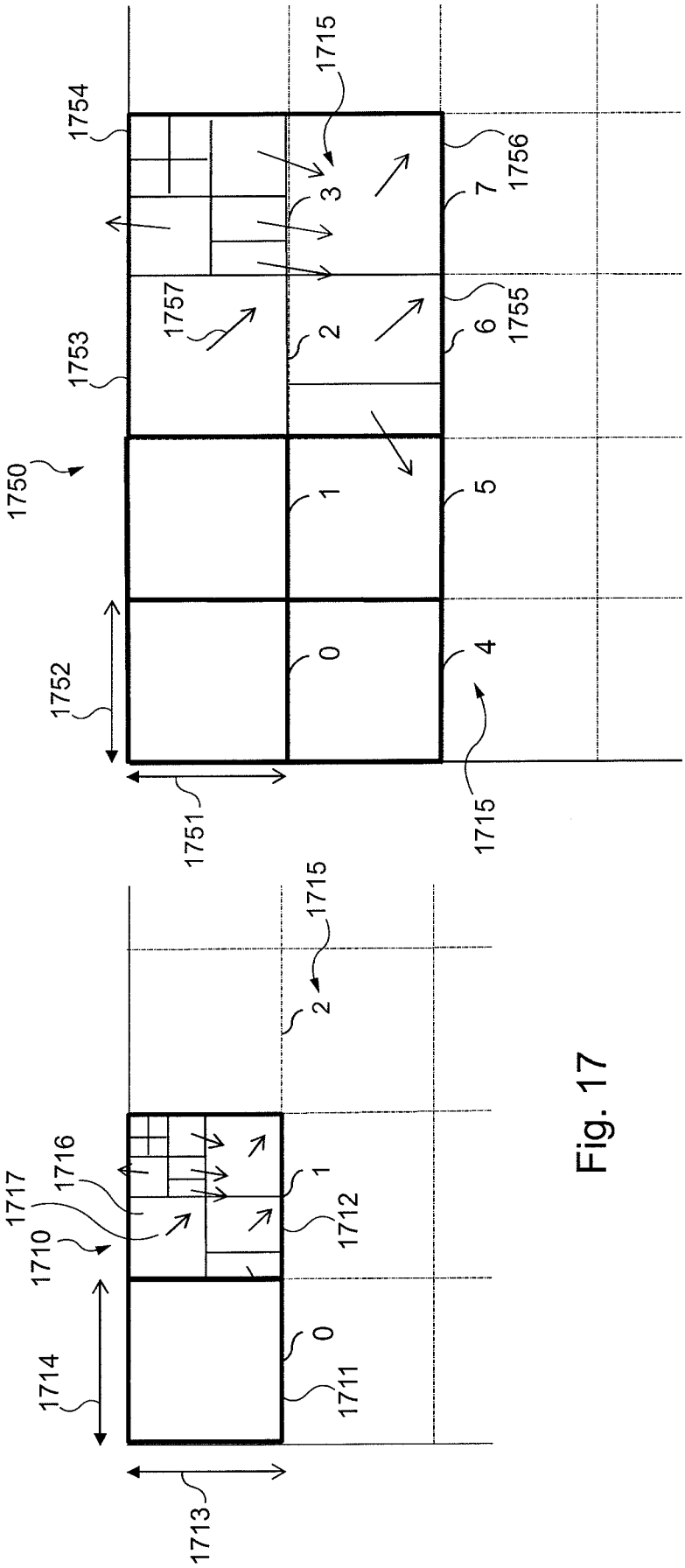

1715
1754
1756
1757
1755
1753
1750
1752
1751
1715
1716
1717
1710
1712
1714
1711
1713

Fig. 17

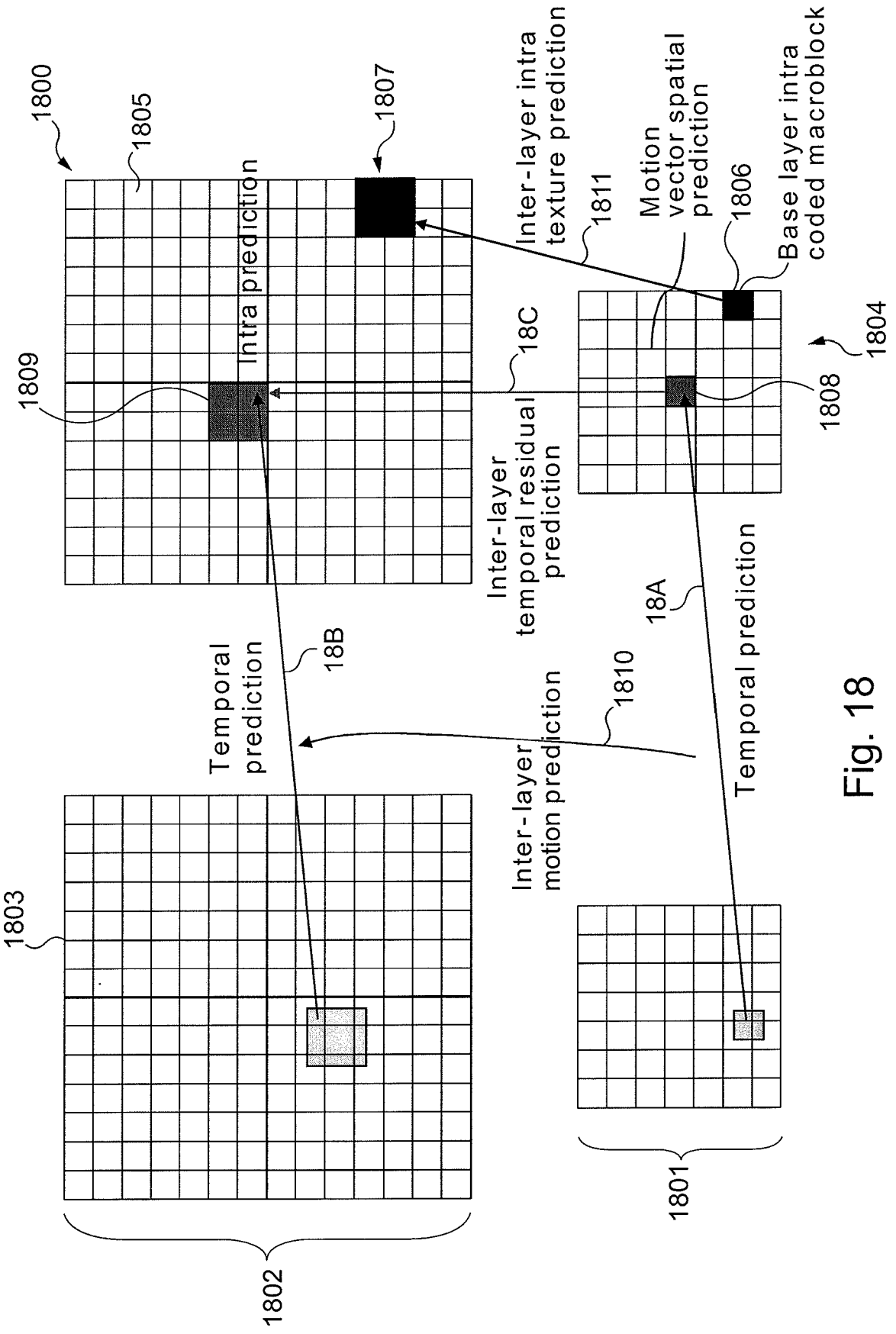

1800
1805
1807
Intra prediction
Inter-layer intra texture prediction
1811
Motion vector spatial prediction
1806
Base layer intra coded macroblock
1809
18C
1804
1808
Inter-layer temporal residual prediction
18A
Temporal prediction
18B
Temporal prediction
1810
Inter-layer motion prediction
1803
1801
1802

Fig. 18

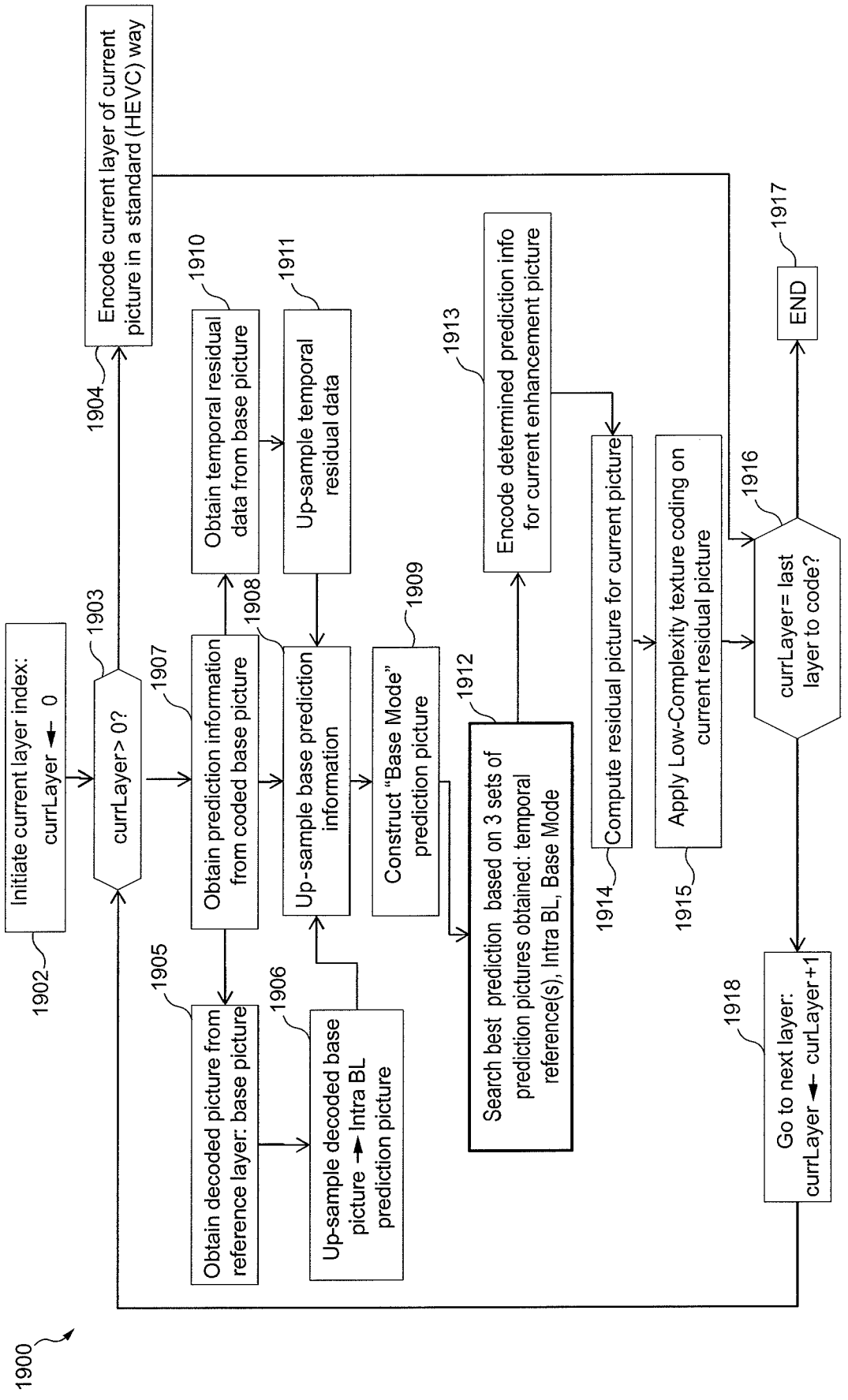

1900
Initiate current layer index: currLayer ← 0
1902
currLayer> 0?
1903
Encode current layer of current picture in a standard (HEVC) way
1904
Obtain decoded picture from reference layer: base picture
1905
Up-sample decoded base picture → Intra BL prediction picture
1906
Obtain prediction information from coded base picture
1907
Up-sample base prediction information
1908
Construct "Base Mode" prediction picture
1909
Obtain temporal residual data from base picture
1910
Up-sample temporal residual data
1911
Search best prediction based on 3 sets of prediction pictures obtained: temporal reference(s), Intra BL, Base Mode
1912
Encode determined prediction info for current enhancement picture
1913
Compute residual picture for current picture
1914
Apply Low-Complexity texture coding on current residual picture
1915
currLayer= last layer to code?
1916
END
1917
Go to next layer: currLayer ← currLayer+1
1918

Fig. 19

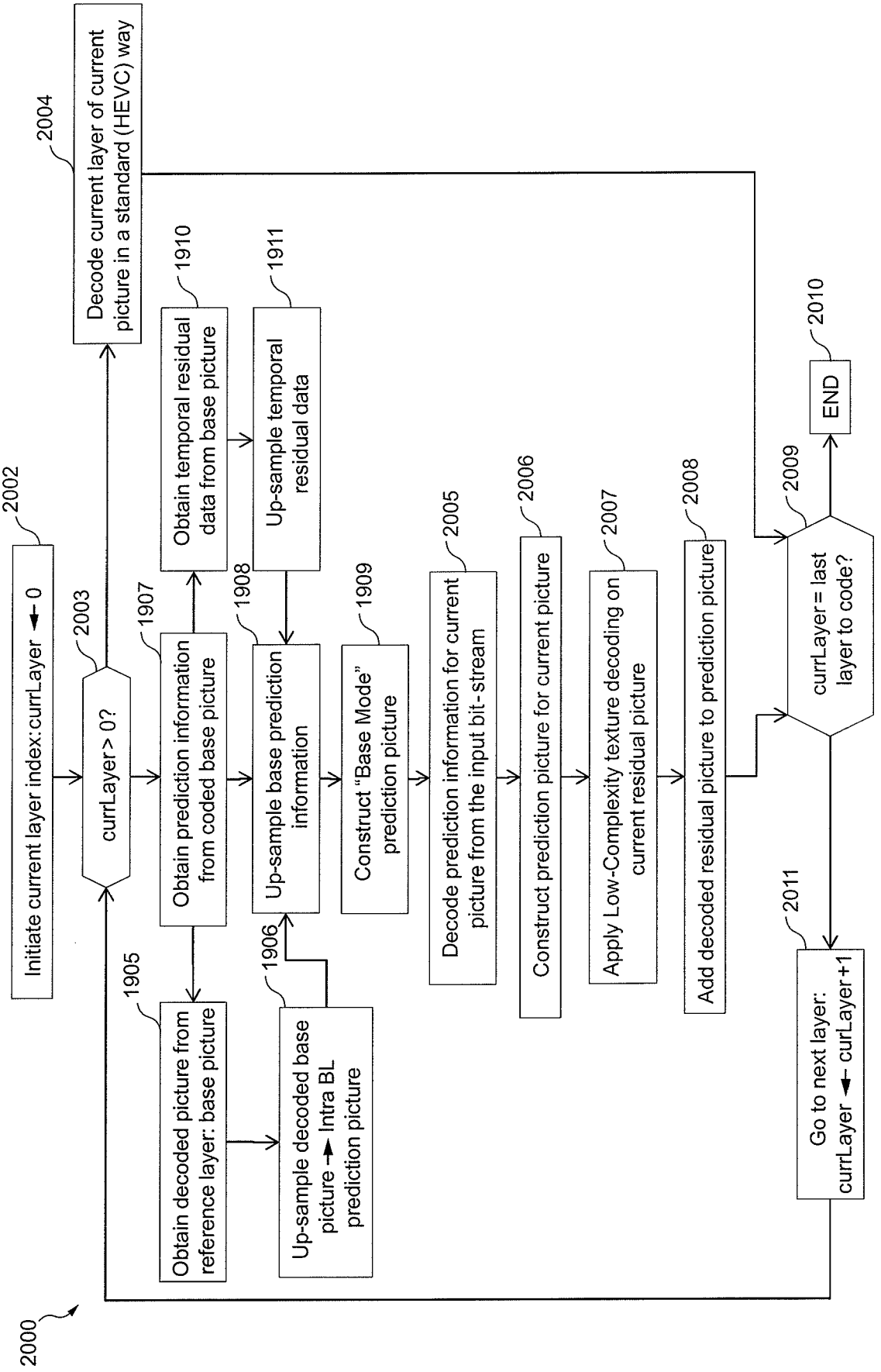

2000
Initiate current layer index:currLayer ← 0
2002
currLayer > 0?
2003
Decode current layer of current picture in a standard (HEVC) way
2004
Obtain prediction information from coded base picture
1907
Obtain decoded picture from reference layer: base picture
1905
Up-sample decoded base picture → Intra BL prediction picture
1906
Up-sample base prediction information
1908
Obtain temporal residual data from base picture
1910
Up-sample temporal residual data
1911
Construct "Base Mode" prediction picture
1909
Decode prediction information for current picture from the input bit-stream
2005
Construct prediction picture for current picture
2006
Apply Low-Complexity texture decoding on current residual picture
2007
Add decoded residual picture to prediction picture
2008
currLayer = last layer to code?
2009
END
2010
Go to next layer: currLayer ← curLayer +1
2011

Fig. 20

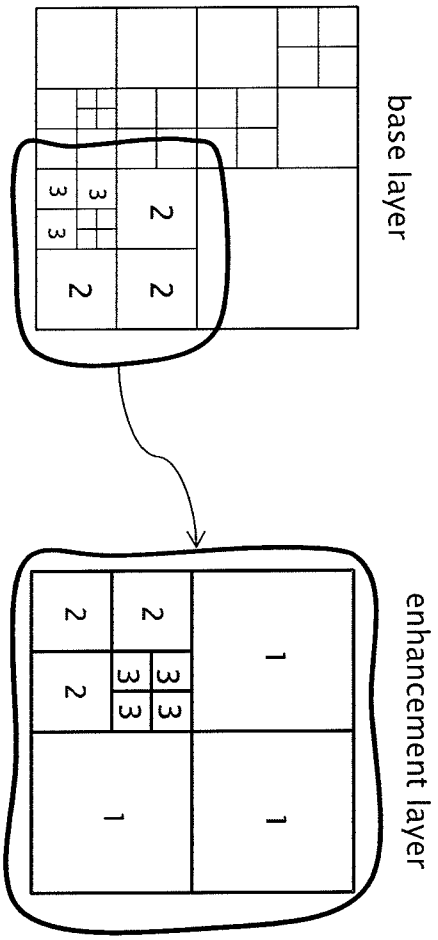


FIG.21

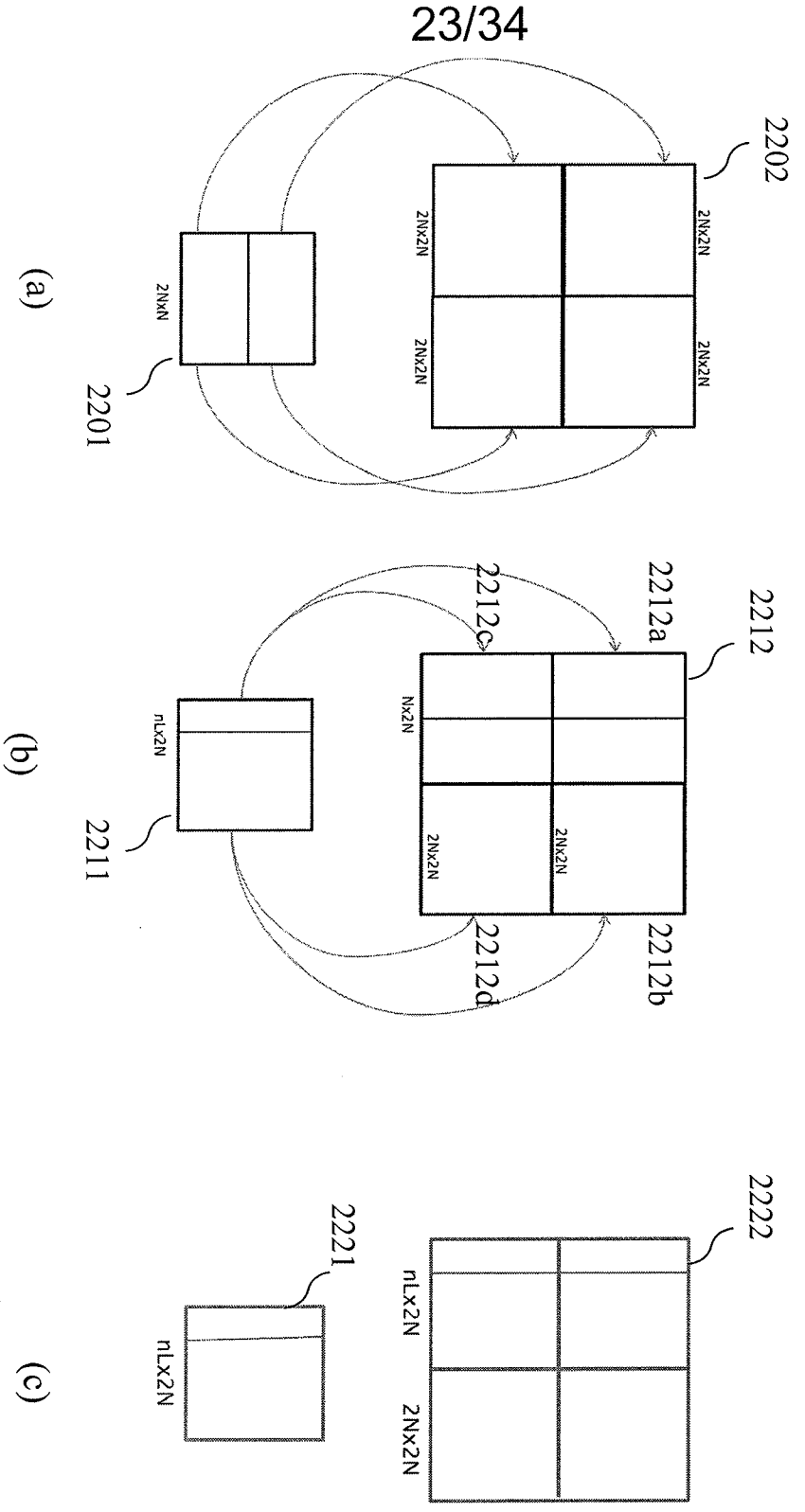

(a)
2202
2NxN
2201
2Nx2N
2Nx2N
2Nx2N
2Nx2N
(b)
2212
2212a
2212c
2211
nLx2N
Nx2N
2Nx2N
2Nx2N
2212d
2212b
(c)
2222
2221
nLx2N
nLx2N
2Nx2N

FIG.22

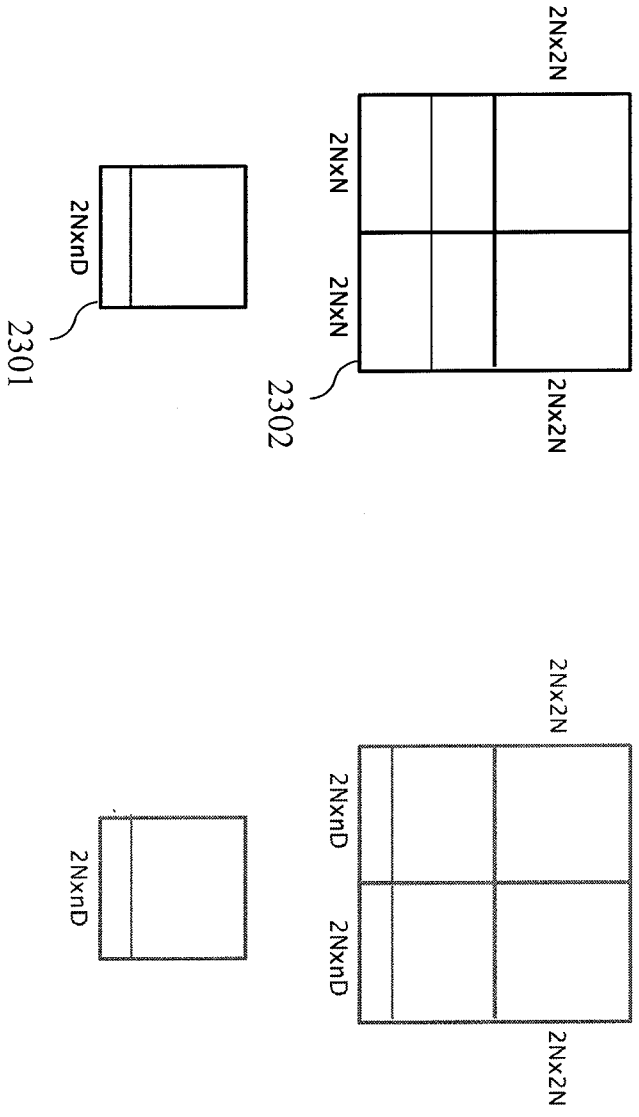

2Nx2N
2NxN
2NxN
2Nx2N
2302
2NxnD
2301
2Nx2N
2NxnD
2NxnD
2Nx2N
2NxnD

FIG.23

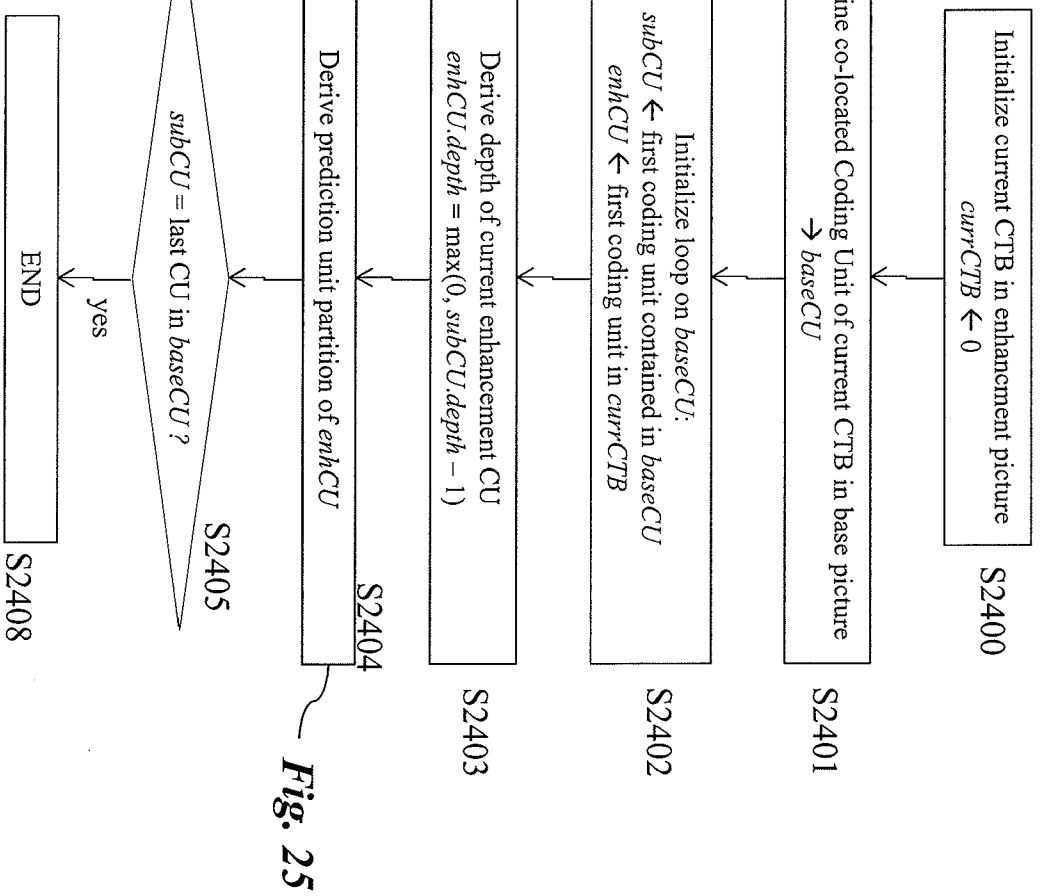

Initialize current CTB in enhancement picture
currCTB ← 0
S2400
Determine co-located Coding Unit of current CTB in base picture
→ baseCU
S2401
Initialize loop on baseCU:
subCU ← first coding unit contained in baseCU
enhCU ← first coding unit in currCTB
S2402
Derive depth of current enhancement CU
enhCU.depth = max(0, subCU.depth – 1)
S2403
Derive prediction unit partition of enhCU
S2404
subCU = last CU in baseCU ?
S2405
no
yes
subCU ← next CU within baseCU
coding tree representation
S2406
enhCU ← next CU within currCTB
S2407
END
S2408

Fig. 25

FIG.24

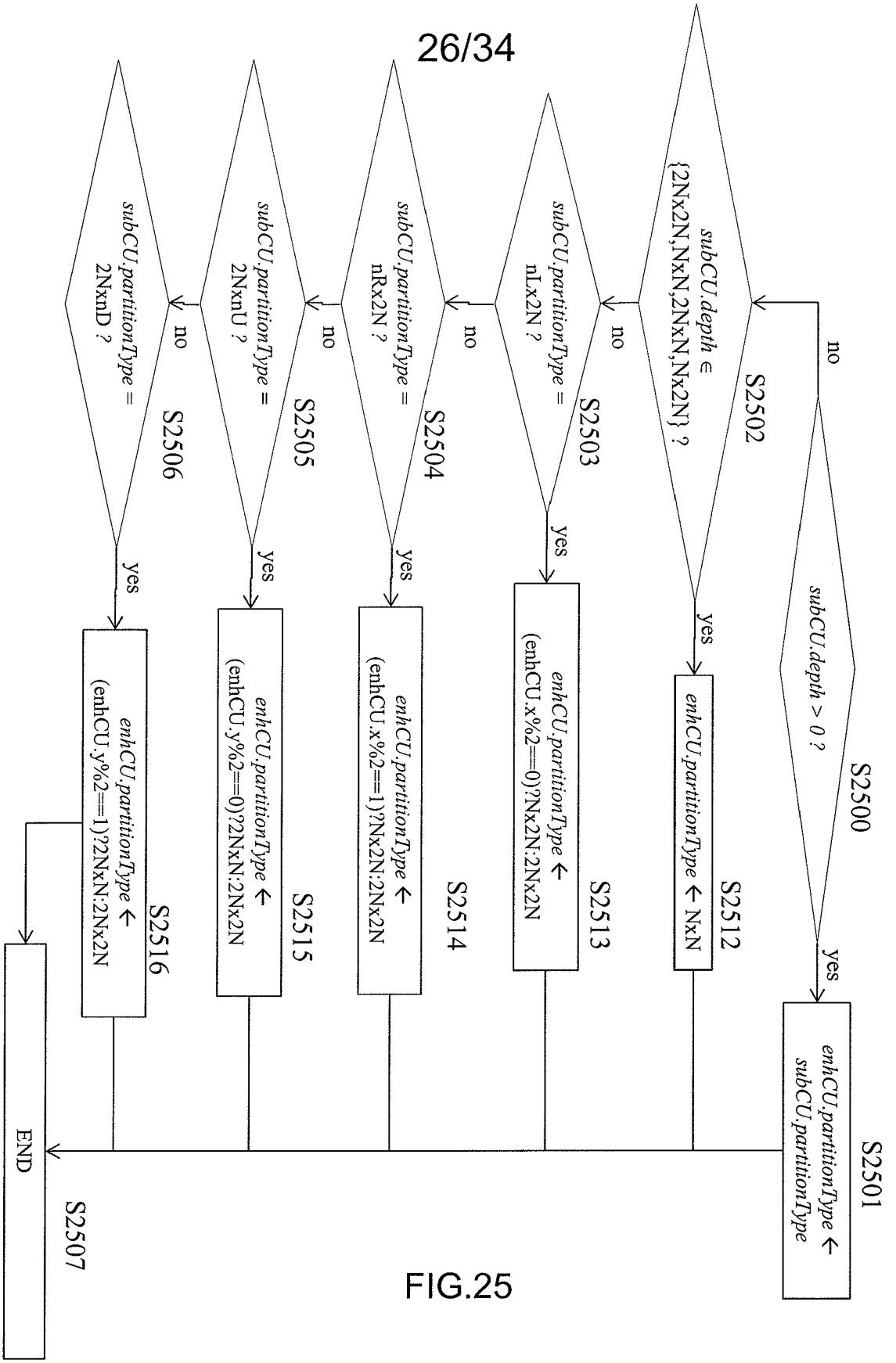

S2500
subCU.depth > 0 ?
no
yes
S2501
enhCU.partitionType ← subCU.partitionType
S2502
subCU.depth ∈ {2Nx2N,NxN,2NxN,Nx2N} ?
yes
S2512
enhCU.partitionType ← NxN
no
S2503
subCU.partitionType = nLx2N ?
yes
S2513
enhCU.partitionType ← (enhCU.x%2==0)?Nx2N:2Nx2N
no
S2504
subCU.partitionType = nRx2N ?
yes
S2514
enhCU.partitionType ← (enhCU.x%2==1)?Nx2N:2Nx2N
no
S2505
subCU.partitionType = 2NxnU ?
yes
S2515
enhCU.partitionType ← (enhCU.y%2==0)?2NxN:2Nx2N
no
S2506
subCU.partitionType = 2NxnD ?
yes
S2516
enhCU.partitionType ← (enhCU.y%2==1)?2NxN:2Nx2N
END
S2507

FIG.25

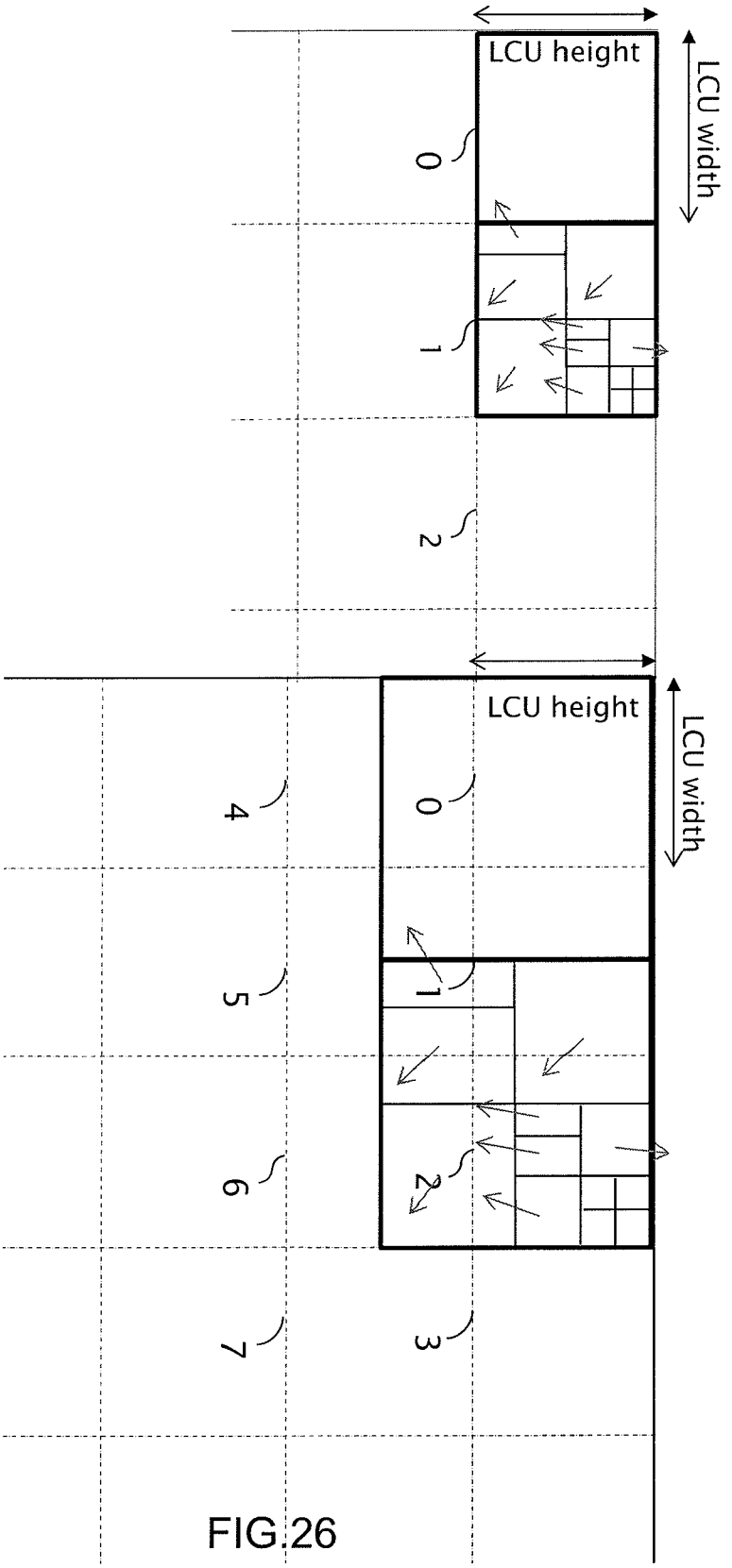


FIG.26

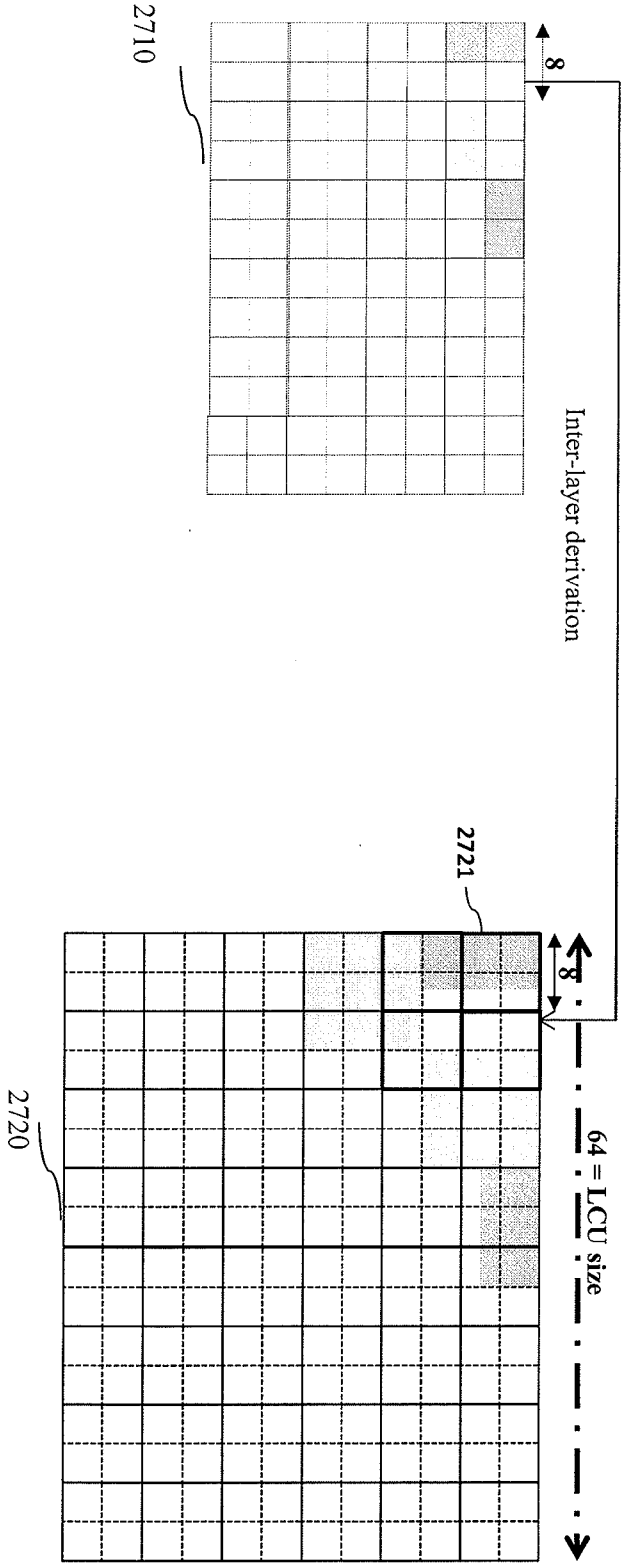


FIG.27

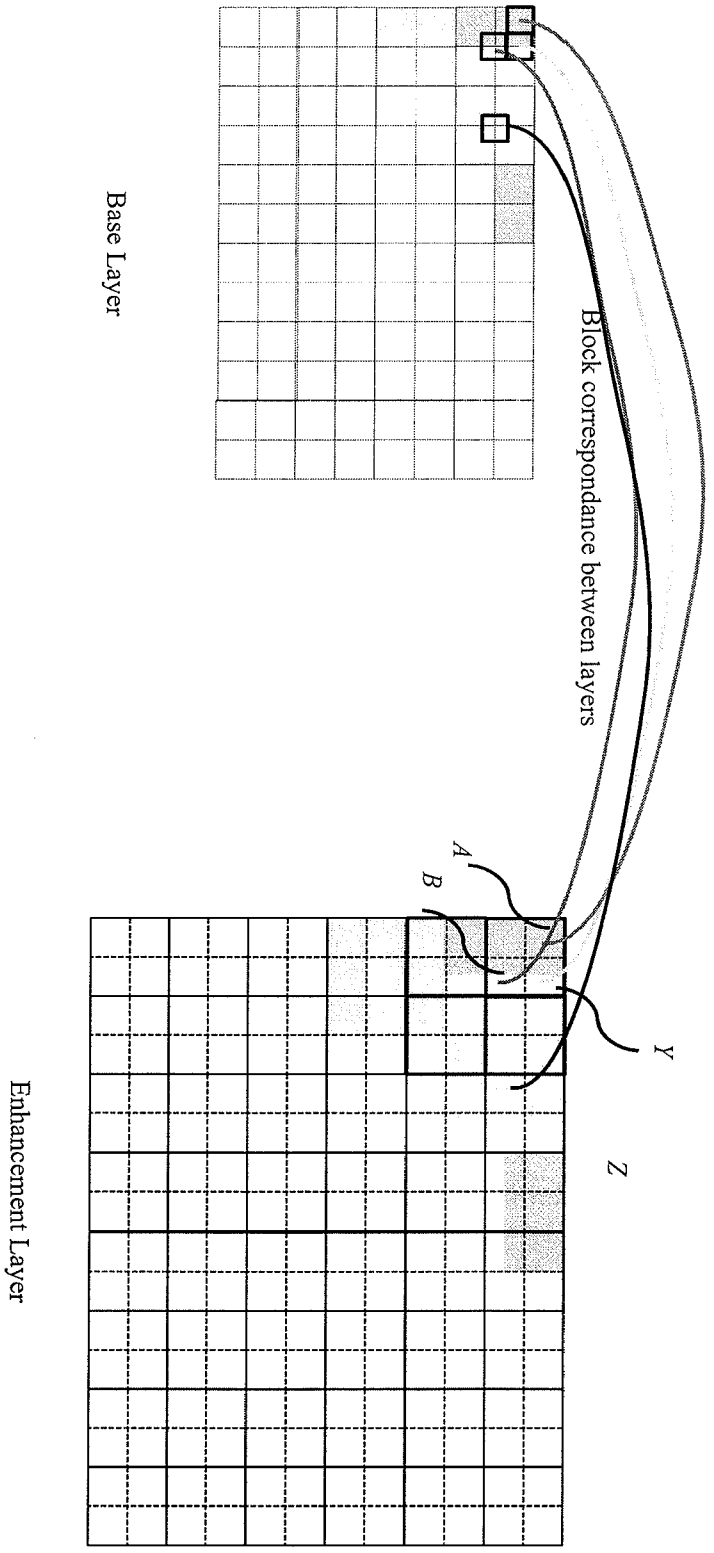

Base Layer
Block correspondance between layers
A
B
Y
Z
Enhancement Layer

FIG.28

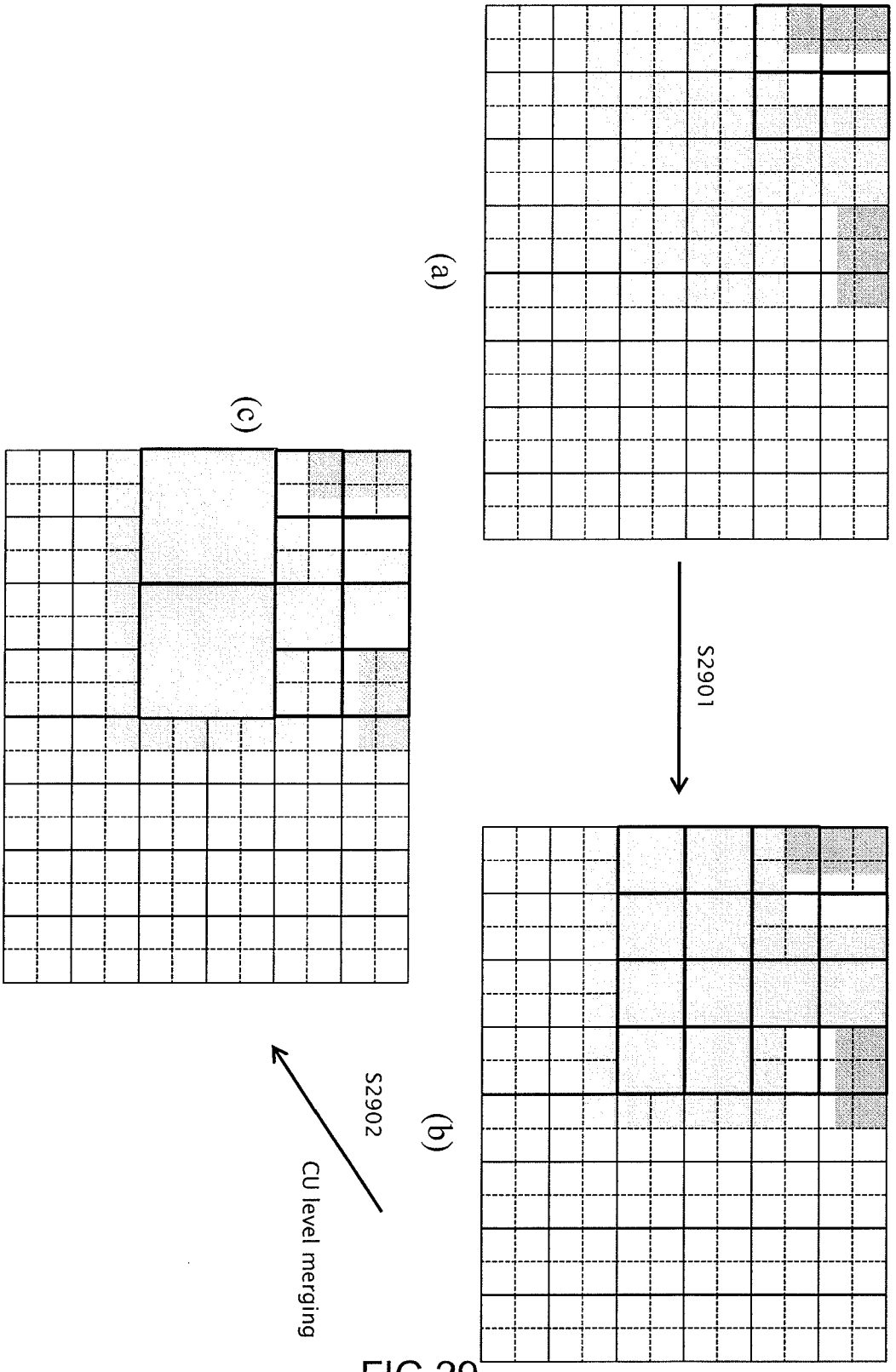


FIG.29

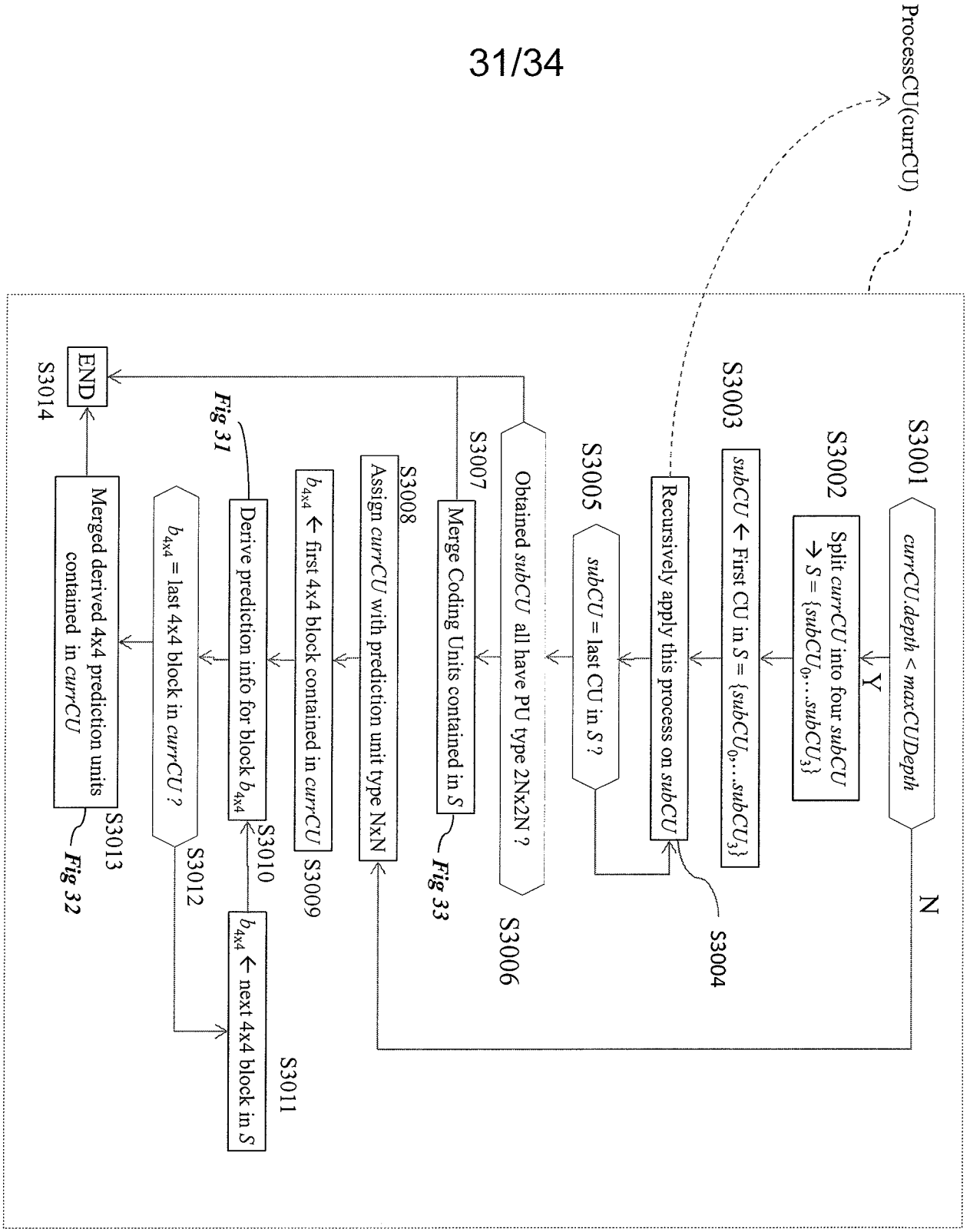

ProcessCU(currCU)
S3001
currCU.depth < maxCUDepth
Y
N
S3002
Split currCU into four subCU
→ S = {subCU0,...subCU3}
S3003
subCU ← First CU in S = {subCU0,...subCU3}
S3004
Recursively apply this process on subCU
S3005
subCU = last CU in S ?
S3006
Obtained subCU all have PU type 2Nx2N ?
S3007
Merge Coding Units contained in S
Fig 33
S3008
Assign currCU with prediction unit type NxN
S3009
b4x4 ← first 4x4 block contained in currCU
S3010
Derive prediction info for block b4x4
Fig 31
S3011
b4x4 ← next 4x4 block in S
S3012
b4x4 = last 4x4 block in currCU ?
S3013
Merged derived 4x4 prediction units contained in currCU
Fig 32
END
S3014

FIG.30

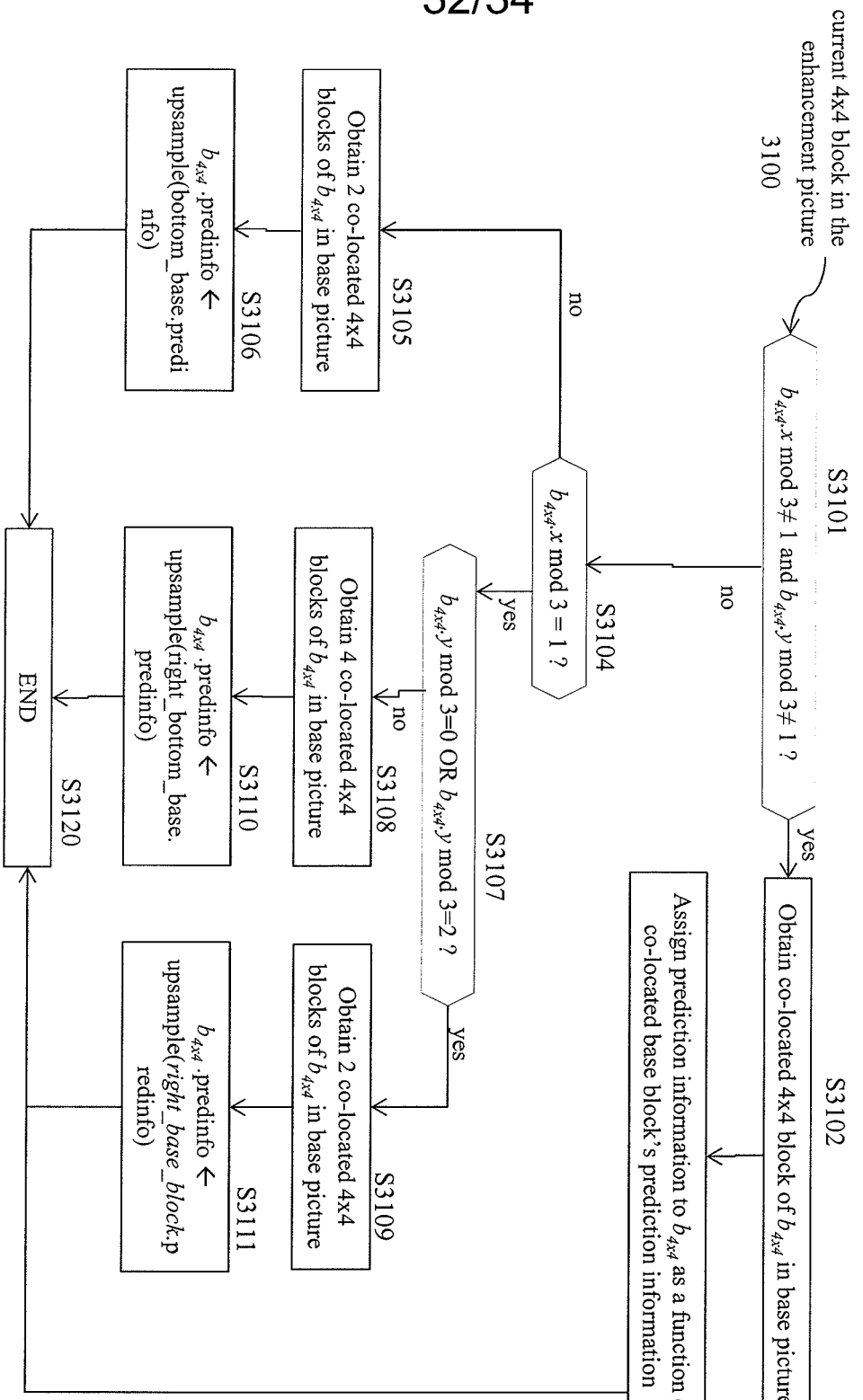

current 4x4 block in the enhancement picture
3100
S3101
$b_{4x4}.x$ mod 3≠ 1 and $b_{4x4}.y$ mod 3≠ 1 ?
yes
S3102
Obtain co-located 4x4 block of b_{4x4} in base picture
Assign prediction information to b_{4x4} as a function of co-located base block's prediction information
no
S3104
$b_{4x4}.x$ mod 3 = 1 ?
yes
S3107
$b_{4x4}.y$ mod 3=0 OR $b_{4x4}.y$ mod 3=2 ?
yes
S3109
Obtain 2 co-located 4x4 blocks of b_{4x4} in base picture
S3111
b_{4x4} .predinfo ← upsample(*right_base_block*.predinfo)
no
S3108
Obtain 4 co-located 4x4 blocks of b_{4x4} in base picture
S3110
b_{4x4} .predinfo ← upsample(right_bottom_base.predinfo)
no
S3105
Obtain 2 co-located 4x4 blocks of b_{4x4} in base picture
S3106
b_{4x4} .predinfo ← upsample(bottom_base.predinfo)
END
S3120

FIG 31

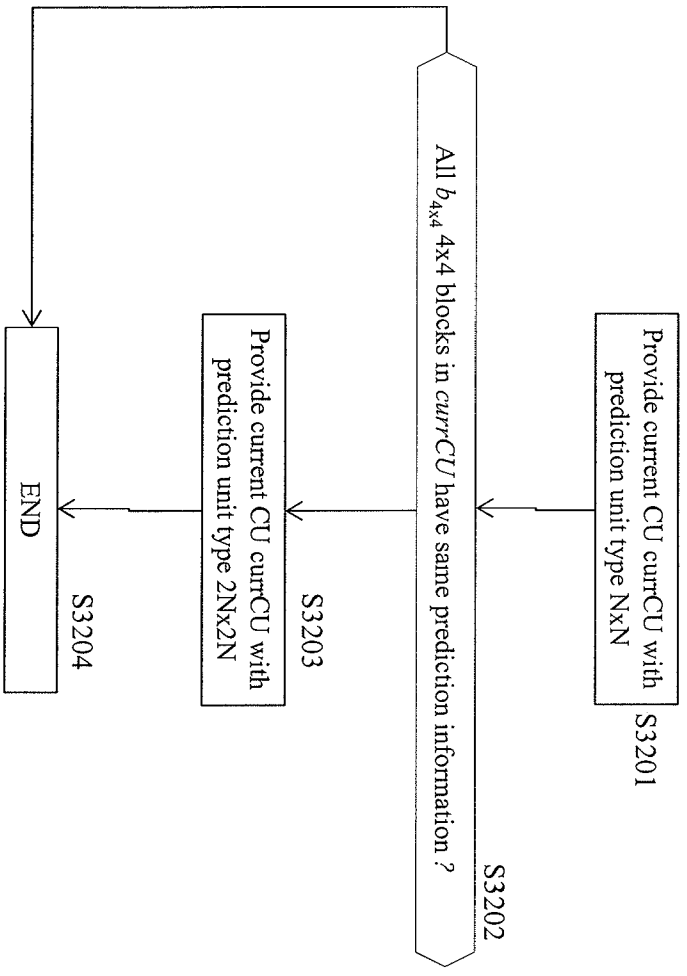

Provide current CU currCU with prediction unit type NxN
S3201
All b_{4x4} 4x4 blocks in *currCU* have same prediction information ?
S3202
Provide current CU currCU with prediction unit type 2Nx2N
S3203
END
S3204

FIG 32

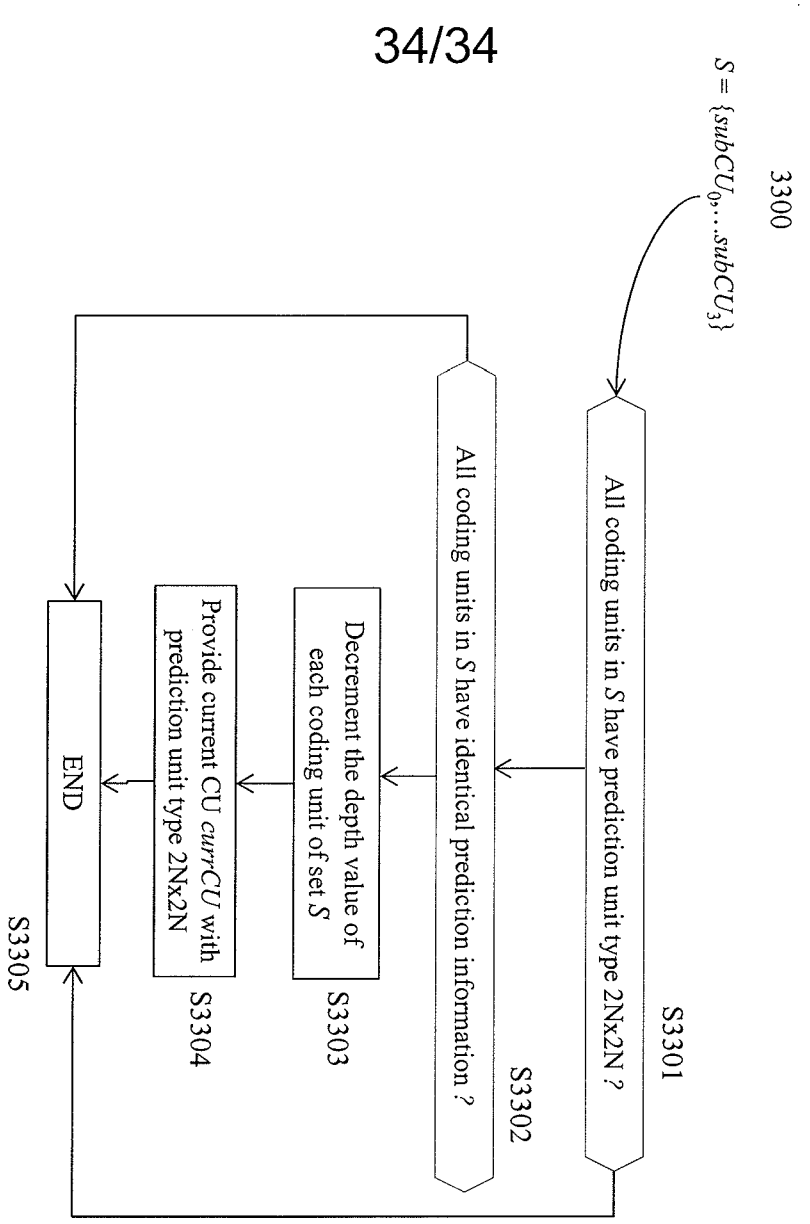

3300
$S = \{subCU_0, ...subCU_3\}$
All coding units in S have prediction unit type 2Nx2N ?
S3301
All coding units in S have identical prediction information ?
S3302
Decrement the depth value of each coding unit of set S
S3303
Provide current CU currCU with prediction unit type 2Nx2N
S3304
END
S3305

FIG.33

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2013/067983

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.: 56
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:

see FURTHER INFORMATION sheet PCT/ISA/210

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:

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.:

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.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/067983

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/26 H04N7/36
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04N

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HONG D ET AL: "Scalability Support in HEVC", 97. MPEG MEETING; 18-7-2011 - 22-7-2011; TORINO; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11),, no. m20710, 13 July 2011 (2011-07-13), XP030049273, the whole document ----- -/--	1-55

[X] 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	Date of mailing of the international search report
11 October 2013	21/10/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Schoeyer, Marnix

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	LASSERRE S ET AL: "Description of the scalable video coding technology proposal by Canon Research Centre France", 11. JCT-VC MEETING; 102. MPEG MEETING; 10-10-2012 - 19-10-2012; SHANGHAI; (JOINT COLLABORATIVE TEAM ON VIDEO CODING OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16); URL: HTTP://WFTP3.ITU.INT/AV-ARCH/JCTVC-SITE/,, no. JCTVC-K0041, 1 October 2012 (2012-10-01), XP030112973, the whole document	1-55
A	GUILLAUME LAROCHE ET AL: "Robust solution for the AMVP parsing issue", 96. MPEG MEETING; 21-3-2011 - 25-3-2011; GENEVA; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11),, no. m19740, 18 March 2011 (2011-03-18), XP030048307, the whole document	1-55
A	JILL BOYCE ET AL: "Information for HEVC scalability extension", 98. MPEG MEETING; 28-11-2011 - 2-12-2011; GENEVA; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11),, no. m21658, 8 November 2011 (2011-11-08), XP030050221, the whole document	1-55
A	LASSERRE S ET AL: "Low Complexity Scalable Extension of HEVC intra pictures based on content statistics", 100. MPEG MEETING; 30-4-2012 - 4-5-2012; GENEVA; (MOTION PICTURE EXPERT GROUP OR ISO/IEC JTC1/SC29/WG11),, no. m24429, 26 April 2012 (2012-04-26), XP030052774, the whole document	1-55

1

Continuation of Box II.2

Claims Nos.: 56

The wording of claim 56 is so obscure that it is not possible to determine the scope of protection sought by the applicants.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.