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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 January 2017 (05.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/001344 A1

(51) **International Patent Classification**:
H04N 19/11 (2014.01) *H04N 19/176* (2014.01)
H04N 19/157 (2014.01) *H04N 19/593* (2014.01)

(21) **International Application Number**: PCT/EP2016/064868

(22) **International Filing Date**: 27 June 2016 (27.06.2016)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
15306049.6 30 June 2015 (30.06.2015) EP

(71) **Applicant**: **THOMSON LICENSING** [FR/FR]; 1-5 rue Jeanne d'Arc, 92130 Issy-les-Moulineaux (FR).

(72) **Inventors**: **THOREAU, Dominique**; c/o Technicolor R&D France, 975 avenue des Champs Blancs, CS17616, 35576 Cesson-Sevigne (FR). **LE PENDU, Mikael**; c/o Technicolor R&D France, 975 avenue des Champs Blancs, CS17616, 35576 Cesson-Sevigne (FR). **BOITARD, Ronan**; c/o Technicolor R&D France, 975 avenue des Champs Blancs, CS17616, 35576 Cesson-Sevigne (FR). **ALAIN, Martin**; c/o Technicolor R&D France, 975 avenue des Champs Blancs, CS17616, 35576 Cesson-Sevigne (FR).

(74) **Agents**: **HUCHET, Anne** et al.; 1-5 rue Jeanne d'Arc, 92130 Issy-Les-Moulineaux (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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

[Continued on next page]

(54) **Title**: METHOD AND APPARATUS FOR DETERMINING PREDICTION OF CURRENT BLOCK OF ENHANCEMENT LAYER

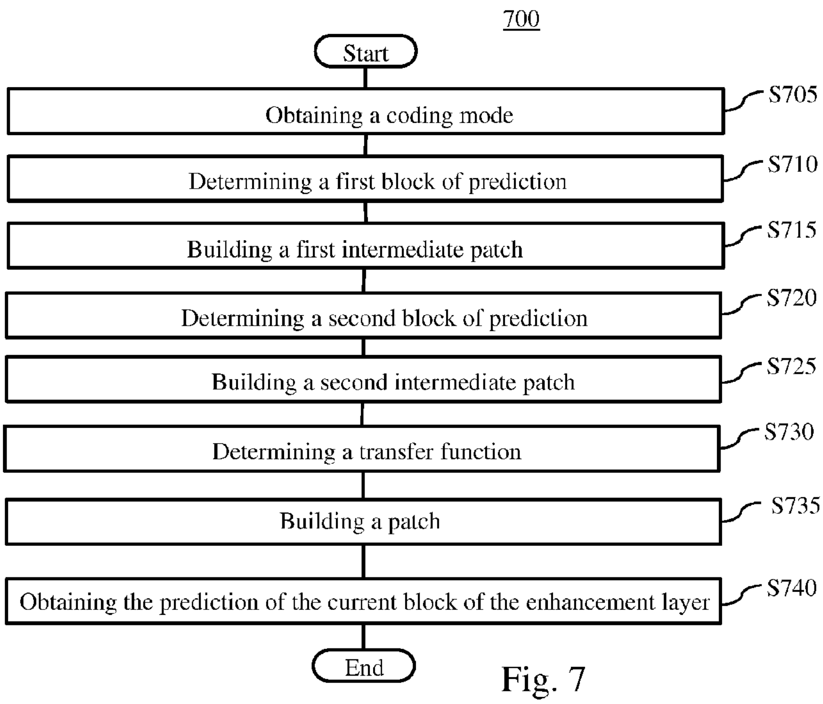


Fig. 7

(57) **Abstract**: A method comprises, building (S715) a first intermediate patch of a low dynamic range; building (S725) a second intermediate patch of a high dynamic range; building (S735) a patch by applying a transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain; predicting (S740) a prediction of the current block of the enhancement layer by extracting a block from the patch; and encoding a residual error between the current block of the enhancement layer and the prediction of the current block of the enhancement layer.

WO 2017/001344 A1

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

METHOD AND APPARATUS FOR DETERMINING PREDICTION OF CURRENT BLOCK OF ENHANCEMENT LAYER

FIELD OF THE INVENTION

The present disclosure relates to a method and an apparatus for determining a prediction of a current block of an enhancement layer.

BACKGROUND OF THE INVENTION

In a field of image processing, Tone Mapping Operators (which may be hereinafter called "TMO") are known. In imaging actual objects in a natural environment, the dynamic range of the actual objects is much higher than a dynamic range that imaging devices such as cameras can image or displays can display. In order to display the actual objects on such displays in a natural way, the TMO is used for converting a High Dynamic Range (which may be hereinafter called "HDR") image to a Low Dynamic Range (which may be hereinafter called "LDR") image while maintaining good visible conditions.

Generally speaking, the TMO is directly applied to the HDR signal so as to obtain an LDR image, and this image can be displayed on a classical LDR display. There is a wide variety of TMOs, and many of them are non-linear operators.

Regarding the art in relation to the LDR/HDR video compression, using a global TMO/iTMO (inverse Tone Mapping Operations) is proposed as one possibility as explained in Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward and W. Heidrich, "On-the-fly tone mapping for backward-compatible high dynamic range image/video compression," ISCAS, 2010.

In this article, the distribution of the floating point data is taken into consideration for the minimization of the total quantization error. The algorithm is described by the following steps (the variables used here are illustrated in FIG. 1.)

Step 1: The logarithm of the luminance values is computed. Thus, for each pixel of luminance L, the following steps are based on the value $l=log_{10}(L)$. (l is still in the floating point format.)

Step 2: A histogram of the l values is computed by taking a bin size fixed to $\delta=0,1$. For example, all the pixels in the image sequence can be used to build the histogram. Thus, for each bin k $(k=1..N)$ the probability p_k that a pixel belongs to this bin is known. The value $l_k = \delta.k$ is assigned to the bin.

Step 3: A slope value is computed for each bin K from a model described by the following formula (1):

$$s_k = \frac{v_{max} \cdot p_k^{1/3}}{\delta.\sum_{k=1}^{N} p_k^{1/3}} \qquad (1)$$

where v_{max} is the maximum value of the considered integer representation ($v_{max} = 2^n-1$ if the data is quantized to n bit integers).

To avoid the risk of division by zero in the inversion equation (inverse tone mapping in 5.), if $s_k=0$, the s_k can be set at a non-null minimum value ε instead.

Step 4: Knowing the N slope values, a global tone mapping curve can be defined. For each k in $[1,N]$, a floating point number l that meets $l_k<l<=l_{k+1}$, is mapped to an integer value v defined by the following formula (2):

$$v = (l - l_k).s_k + v_k \qquad (2)$$

where the values v_k are defined from the values s_k by $v_{k+1} = \delta.s_k + v_k$ (and $v_1=0$).

The value v is then rounded to obtain an integer in the interval $[0, 2^n-1]$.

Step 5: In order to perform the inverse tone mapping, the parameters s_k $(k=1..N)$ must be transmitted to the decoder. For a given pixel of value v in the tone mapped image, firstly, the value k that meets $v_k<=v<v_{k+1}$ must be found.

The inverse equation is then expressed as the following formula (3):

$$l_{dec} = l_k + \frac{(v-v_k)}{s_k} \quad (3)$$

Here, the decoded pixel value is made $Ldec = 10^{l_{dec}}$.

Moreover, in order to apply the inverse tone mapping (iTMO), the decoder must know the curve in FIG. 1.

The term "decoded" here corresponds to a de-quantization operation that is different from the term "decoded" of the video coder/decoder.

Another possibility is to use local tone mapping operators as disclosed in M. Grundland et al, "Non linear multiresolution blending", Machine Graphis & vision International Journal Volume 15 Issue 3 Feb 2006, and Zhe Wendy Wang ; Jiefu Zhai ; Tao Zhang ; Llach, Joan "Interactive tone mapping for High Dynamic Range video". ICASSP 2010. For example the TMO laplacian pyramid may be used based on the disclosure of Peter J. Burt Edward H. Adelson. "The Laplacian Pyramid as a compact image code," IEEE Transactions on Communications, vol. COM-3l, no. 4, April 1983, Burt P.J., "The Pyramid as Structure for Efficient Computation. Multiresolution Image Processing and Analysis", Springer-Verlag, 6-35, and Zhai jiedu, Joan Llach, "Zone-based tone mapping" WO 2011/002505 A1. The efficiency of the TMO consists in the extraction of different intermediate LDR images from an HDR image where the intermediate LDR images correspond to different exposures. Thus, the over-exposed LDR image contains the fine details in the dark regions while the lighting regions (of the original HDR image) are saturated. In contrast, the under-exposed LDR image contains the fine details in the lighting zone while the dark regions are clipped.

Afterwards, each LDR image is decomposed in laplacian pyramid of n levels, while the highest level is dedicated to the lowest resolution, and the other levels provide the different spectral bands (of gradient). So, at this stage, each LDR image corresponds to a laplacian pyramid, and further we can notice that each LDR image can be rebuilt from its laplacian pyramid by using an inverse decomposition or "collapse", only if there is not a rounding miscalculation.

Finally, the tone mapping is implemented with the fusion of the different pyramid levels of the set of intermediate LDR images, and the resulting blended pyramid is collapsed so as to give the final LDR image.

In fact, the fusion of the gradients of the different spectral bands (or pyramid levels) is a non-linear process. The advantages of the type of algorithms reside on an efficient result of the tone mapping, but sometimes a lot of well-known rendering faults like halo artifacts are caused. The above references give more details on this technique.

Indeed, because this tone mapping is

non-linear, it is difficult to implement the inverse tone mapping of the LDR so as to give an acceptable prediction to a current block of HDR layer in the case of SNR (Signal-to-Noise Ratio) or spatial video scalability.

Moreover, WO2010/018137 discloses a method for modifying a reference block of a reference image, a method for encoding or decoding a block of an image with help from a reference block and device therefore and a storage medium or signal carrying a block encoded with help from a modified reference B. In the prior art, a transfer function is estimated from neighboring mean values, and this function is used to correct an inter-image prediction. However, in WO2010/018137, the approach was limited to the mean value so as to give a first approximation of the current block and the collocated one.

SUMMARY OF THE INVENTION

According to an embodiment of the present disclosure, there is provided a method comprising, building a first intermediate patch of a low dynamic range with the neighboring pixels of the collocated block of the base layer and a first prediction block predicted from neighboring pixels of a collocated block of a base layer with a coding mode of the base layer; building a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and a second prediction block predicted from neighboring pixels of a current block of an enhancement layer with the coding mode; building a patch by applying a transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain, wherein the transfer function is determined to transform the first intermediate patch to the second intermediate patch in a transform domain; predicting a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block in the patch being collocated to the current block of the enhancement layer in the second intermediate patch; and encoding a residual error between the current block of the enhancement layer and the prediction of the current block of the enhancement layer.

According to an embodiment of the present disclosure, there is provided an apparatus comprising, a first intermediate patch creation unit configured to predict a first prediction block from neighboring pixels of the collocated block of a base layer with a coding mode of the base layer and to build a first intermediate patch of a low dynamic range with the neighboring pixels of the collocated block of the base layer and the first prediction block; a second intermediate patch creation unit configured to predict a second prediction block from neighboring pixels of a current block of an enhancement layer with the coding mode and to build a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and the second prediction block; a unit to determine a transfer function to transform the first intermediate patch to the second intermediate patch in a transform domain, to build a patch by applying the transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain and to predict a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block being in the patch collocated to the current block of the enhancement layer in the second intermediate patch; and an encoder to encode a residual error between the current block of the enhancement layer and the prediction of the current block of the enhancement layer.

According to another embodiment of the present disclosure, there is provided a method comprising, decoding a residual prediction error; building a first intermediate patch of a low dynamic range with the neighboring pixels of the collocated block of the base layer and a first prediction block predicted from neighboring pixels of a collocated block of a base layer with a coding mode of the base layer; building a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and a second prediction block predicted from neighboring pixels of a current block of an enhancement layer with the coding mode; building a patch by applying a transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain, wherein the transfer function is to transform the first intermediate patch to the second intermediate patch in a transform domain; predicting a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block in the patch being collocated to the current block of the enhancement layer in the second intermediate patch; and reconstructing a block of the enhancement layer by adding the prediction error to the prediction of the current block of the enhancement layer.

According to yet another embodiment of the present disclosure, there is provided an apparatus comprising, a decoder for decoding a residual prediction error; a first intermediate patch creation unit configured to build a first intermediate patch of a low dynamic range with the neighboring pixels of a collocated block of a base layer and a first prediction block predicted from neighboring pixels of a collocated block of a base layer with a coding mode of the base layer; a second intermediate patch creation unit configured to build a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and a second prediction block predicted from neighboring pixels of a current block of an enhancement layer with the coding mode and; a unit to build a patch by applying the transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain, wherein the transfer function is to transform the first intermediate patch to the second intermediate patch in a transform domain and to predict a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block being in the patch collocated to the current block of the enhancement layer in the second intermediate patch; and a unit to add the prediction error to the prediction of the current block of the enhancement layer to reconstruct a block of the enhancement layer.

Other objects, features, and advantages of the present disclosure will become more apparent from the following detailed description when read in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a histogram of the floating point values $l=log_{10}(L)$ and its associated tone mapping curve based on the slopes s_k;

FIGs. 2A and 2B are an image of a reconstructed base layer and an image of a current block of an enhancement layer to be encoded;

FIGs. 3A through 3J are drawings illustrating an example of Intra 4x4 prediction specified in H.264 standards;

FIGs. 4A and 4B are block diagrams illustrating an apparatus for determining a prediction of a current block of an enhancement layer of the first embodiment and FIG. 4A is an encoder side and FIG. 4B is a decoder side;

FIGs. 5A and 5B are block diagrams illustrating a configuration of an apparatus for determining a prediction of a current block of an enhancement layer of a second embodiment of the present disclosure embodiment and FIG. 5A is an encoder side and FIG. 5B is a decoder side;

FIG. 6 is a block diagram illustrating a configuration of an apparatus for determining a prediction of a current block of an enhancement layer of a fourth embodiment of the present disclosure; and

Fig. 7 is a flow diagram illustrating an exemplary method for determining a prediction of a current block of an enhancement layer according to an embodiment of the present disclosure.

DETAILED DESCRIPTION OF THE EMBODIMENTS

A description is given below of embodiments of the present disclosure, with reference to the drawings.

The embodiments of the present disclosure aim to improve the processing of an inverse Tone Mapping Operations (which may be hereinafter called an “iTMO”), and the previous TMO used in a global or local (the non-linear) manner, obviously if the base layer signal is still usable.

The idea relates to, for example, an HDR SNR scalable video coding with a first tone mapped base layer l_b using a given TMO dedicated to the LDR video encoding, and a second enhancement layer l_e dedicated to the HDR video encoding. In this case (SNR scalability), for a current block b_e (to be encoded) of the enhancement layer, a block of prediction extracted from the base layer b_b (the collocated block) should be found, and the block has to be processed by inverse tone mapping.

In order to implement the inverse tone mapping of the block b_b, a function of transformation T_{be} should be estimated to allow the pixels of the patch $p_b^{'}$ (composed of a virtual block $b_b^{'}$ (homologous of b_b) and its neighbor) to be transformed to the current patch $p_e^{'}$ (composed of a virtual block $b_e^{'}$ (homologous of b_e) and its neighbor). Once T_{be} is determined, the function of transformation T_{be} can be applied to the patch p_b (composed of the block b_b and its neighbor) giving the patch p_b^T, finally the last step resides on the extraction of the block $\tilde{b}_e$ collocated to the current block in the patch p_b^T . Here, the block $\tilde{b}_e$ corresponds to the prediction of the block b_e.

Here, it should be noted that before the estimation of the transformation T_{be}, the coding mode of the collocated block b_b of the base layer is needed, or a mode of prediction is needed to be extracted from the reconstructed image (of the l_b) among the set of available coding modes (of the encoder of the enhancement layer) based on the base layer.

It is also important to notice that the entire processing steps explained above are also implemented at the decoder side as well as encoder side.

[Principle]

In order to illustrate an approach proposed in the embodiments of the present disclosure, an example based on SNR scalability is given below. In this case (SNR scalability), a block of prediction extracted from the base layer $\mathrm{b_b}$ (the collocated block) should be found for a current block $\mathrm{b_e}$ (to be encoded) of the enhancement layer, and the block of prediction has to be processed by inverse tone mapping.

FIGs. 2A and 2B illustrate an image of a reconstructed base layer and an image of a current block to be encoded separately.

The notations illustrated in FIG. 2B, relative to the current image of the enhancement layer l_e are as follows:

- The current block (unknown) to predict of the enhancement layer is: X_u^B
- The known reconstructed (or decoded) neighbor (or template) of the current block : X_k^T
- The current patch is : $X = \begin{bmatrix} X_k^T \\ X_u^B \end{bmatrix}$ (4)
- The index k and u indicate respectively «known» and «unknown».

The notations illustrated in FIG. 2A, relative to the image of the base layer l_b are as follows:

- The collocated block (known) of the base layer, (that is effectively collocated to the current block to predict of the enhancement layer) is: Y_k^B
- The known reconstructed (or decoded) neighbor (or

template) of the current block is : Y_k^T

• The collocated patch (collocated of X) is:

$$Y = \begin{bmatrix} Y_k^T \\ Y_k^B \end{bmatrix} \quad (5)$$

The goal is to determine a block of prediction for the current block X_u^B from the block Y_k^B. In fact, the transformation will be estimated between the patches Y and X, this transformation corresponding to a kind of inverse tone mapping.

Obviously, in the context of video compression, the block X_u^B is not available (remember that the decoder will implement the same processing), but there are a lot of possible modes of prediction that could provide a first approximation (more precisely prediction) of the current block X_u^B. Here, the first approximation of the current block X_u^B and its neighbor X_k^T compose the intermediate patch X′ of the patch X.

After that, the first approximation of the block X_u^B is used so as to find a transformation function $Trf\ (l_b \rightarrow l_e)$ which allows the intermediate patch of X to be transformed into the intermediate patch of Y (respectively noticed X′ and Y′), and this transformation is finally applied to the initial patch Y allowing the definitive block of prediction to be provided.

[First Embodiment]

A description is given of a first embodiment of a method and an apparatus for determining a prediction of a current block of an enhancement layer, with reference to FIGs. 3A through 3J and 4.

More specifically, the first embodiment of the present disclosure is about the SNR scalability, that is to say, the same spatial resolution between the LDR base layer and the HDR enhancement layers. In addition, in the first embodiment, the collocated block Y_k^B of the current block X_u^B had been encoded with one of the intra coding modes of the coder of the enhancement layer, for example, the intra modes of H.264 standard defined in MPEG-4 AVC/H.264 and described in the document ISO/IEC 14496-10.

With the coding mode of index m of the block Y_k^B and with the neighboring pixels of Y_k^T, it is possible to reconstruct the block of prediction $Y_{prd,m}^B$.

FIGs. 3A through 3J are drawings illustrating Intra 4x4 predictions specified in H.264 standards. As illustrated in FIGs. 3A through 3J, the N (here in case of H264 N=9) different intra mode predictions are offered in the H.264 standards.

In H.264, Intra 4x4 and Intra 8x8 predictions correspond to a spatial estimation of the pixels of the current block to be coded based on the neighboring reconstructed pixels. The H.264 standard specifies different directional prediction modes in order to elaborate the pixel prediction. Nine (9) intra prediction modes are defined on 4x4 and 8x8 block sizes of the macroblock (MB). As depicted in FIG.3, eight (8) of these modes consist of a 1D directional extrapolation of the pixels (from the left column and the top line) surrounding the current block to predict. The intra prediction mode 2 (DC mode) defines the predicted block pixels as the average of available surrounding pixels.

In the example of intra 4x4, the predictions are built as illustrated in FIG. 3A through 3J.

For example, as illustrated in FIG. 3C, in mode 1 (horizontal), the pixels e, f, g, and h are predicted with (left column) the reconstructed pixel J.

Moreover, as illustrated in FIG. 3G, in mode 5, as a first example, "a" is predicted by (Q+A+1)/2. Similarly, as a second example, "g" and "p" are predicted by (A+2B+C+2)/4.

Here, returning to the problem discussed above, it is preferable to build a prediction of the current block X_u^B, for the purpose of utilizing the same m index mode of prediction than one used in the base layer and the current neighbor X_k^T that provide the block of prediction: $X_{prd,m}^B$.

Here, two intermediate patches X′ and Y′ can be composed as the following formulas (6) and (7).

The current intermediate patch X′:

$$X' = \begin{bmatrix} X_k^T \\ X_{prd,m}^B \end{bmatrix} \quad (6)$$

The intermediate patch Y′ of the base layer: $Y' = \begin{bmatrix} Y_k^T \\ Y_{prd,m}^B \end{bmatrix}$ (7)

The desired transform Trf is computed between Y′ and X′, in a Transform Domain (TF), and the transformation could be Hadamard, Discrete Cosine Transform (DCT), Discrete Sine Transform (DST) or Fourier transform and the like. The following formulas (8) and (9) are provided.

$$T_{X'} = \text{TF}\ (X') \quad (8)$$

$$T_{Y'} = \text{TF}\ (Y') \quad (9)$$

The formula TF (Y′) corresponds to the 2D transform "TF" (for example, DCT) of the patch Y′.

The next step is to compute the transfer function Trf that allows $T_{Y'}$ to be transformed to $T_{X'}$ in which the following formulas (10) and (11) are applied to each couple of coefficients.

If

(abs ($T_{X'}$ (u,v)) > th and abs ($T_{Y'}$ (u,v) > th))

then

$$Trf\ (u,v) = T_{X'}\ (u,v) / T_{Y'}\ (u,v) \quad (10)$$

else

$$Trf\ (u,v) = 0 \quad (11)$$

end if

Here, u and v are the transfer transform coordinates of the coefficients of $T_{X'}$ $T_{Y'}$ $and\,Trf$, and th is a threshold of a given value, which avoids singularities in the Trf transfer function. For example, th could be equal to 1 in the context of H.264 or HEVC standards compression. HEVC (High Efficiency Video Coding) is described in the document, B. Bross, W.J. Han, G. J. Sullivan, J.R. Ohm, T. Wiegand JCTVC-K1003, "High Efficiency Video Coding (HEVC) text specification draft 9," Oct 2012.

The function Trf is applied to the transformation (TF) of the initial patch of the base layer Y which gives the patch Y'' after inverse transform (TF^{-1}). The patch Y'' is composed of the template Y''^{T} and the block Y''^{B}_{m} as shown by formulas (12) through (14).

$$Y'' = \begin{bmatrix} Y''^{T} \\ Y''^{B}_{m} \end{bmatrix} \quad (12)$$

with $$Y'' = TF^{-1}(T_{Y'}) \quad (13)$$

and $$T_{Y'} = TF(Y).Trf \quad (14)$$

The formula $TF(Y).Trf$ corresponds to the application of the transfer function Trf to the components of the transform patch T_Y of the initial patch Y of the base layer, and this application is performed for each transform component (of coordinates u and v) as shown by formula (15).

$$T_{Y'}(u,v) = T_Y(u,v).Trf(u,v) \quad (15)$$

Finally, the prediction of the current block X_u^B resides on the extraction of the block $Y_m''^B$ from the patch Y'', and the notation m indicating that the block of prediction is built with help from m intra mode index of the base layer.

FIGs. 4A and 4B are block diagrams illustrating an apparatus for determining a prediction of a current block of an enhancement layer of the first embodiment. The principle of this description of intra SNR scalability is also illustrated in the FIGs. 4A and 4B.

With reference to FIGs. 4A and 4B, Local inter-layer LDR HDR prediction is described.

So as to clarify the description and particularly the decoder, we describe the SNR Scalable Video Coding (SVC) scheme:

(1) Firstly the base layer
(2) And secondly the enhancement layer

At the encoder (or coder) side shown in FIG.

4A, and the decoder side shown in FIG. 4B, knowing that the proposal focuses on the inter layer (bl->el) prediction.

At the coder and the decoder sides, only the intra image prediction mode, using the intra mode (m) is described, because our inter layer prediction mode uses intra mode (m). So it is well known that the function of the prediction unit (using a given RDO (Rate Distortion Optimizations) criterion) resides on the determination of the best prediction mode from:

(1) The intra and inter image predictions at the base layer level

(2) The intra, inter image and inter layer predictions (our new prediction mode) at the enhancement layer level

<u>Signification of the index</u>:

k: known

u: unknown

B: block

T: neighbor of the block (usually called "Template" in the video compression domain)

Pred: prediction

m: index of the intra coding mode from N available modes

Y, X, Y', X', and Y'' are patches which are composed of a block and a template with reference to FIGs. 2A and 2B

<u>Coder side (unit 400) in FIG. 4A:</u>

An original block 401 b_e is tone mapped using the TMO 406 that gives the original tone mapped block b_{bc}.

Base layer (bl)

We consider the original base layer block b_{bc} to encode

a) With the original block b_{bc} and the (previous decoded) images stored in the reference frames buffer 426, the motion estimator (motion estimation unit) 429 finds the best inter image prediction block with a given motion vector (temporal prediction unit) and the temporal prediction (Temp Pred Pred) unit 430 gives the temporal prediction block. From the available intra prediction modes (illustrated with the FIG. 3, in case of H264) and neighboring reconstructed (or decoded) pixels the spatial prediction (Sp Pred) unit 428 gives the intra prediction block.

b) If the mode decision process (unit 425) chooses the intra image prediction mode (of m index, from N intra available modes), the residual error prediction rb is computed(by the combiner 421) with the difference between the original block b_{bc} and the prediction block $\tilde{b}_b$ ($Y^{B}_{prd,m}$)

c) After, the residual error prediction rb is transformed and quantized to r_{bq} by T Q unit 422 and finally entropy coded by entropy coder unit 423 and sent in the bitstream base layer.

d) The decoded block is locally rebuilt, by adding (with the combiner 427) the inverse transformed and dequantized by T^{-1} Q^{-1} unit 424 prediction error block r_{bdq} to the prediction block $\tilde{b}_b$ giving the reconstructed (base layer) block

e) The reconstructed (or decoded) frame is stored in the (bl) reference frames buffer 426.

Enhancement layer (el)

We can notice that the structure of the coder of the enhancement layer is similar to the coder of the base layer, for example the units 407, 408, 409 and 413 have the same function than the respective units 425, 426, 429 and 430 of the coder of the base layer in terms of coding mode decision, temporal prediction and reference frames buffer. We consider now the original enhancement layer block b_e to encode.

f) For the block of the enhancement layer, if the collocated block of the base layer is coded in intra image mode, then we consider the intra mode (of m index) of this collocated block (S705 of the method 700 shown in Fig. 7).

g) With this intra mode (of m index) of the base layer we determine:
   - o determine or re-use the intra block of prediction ($\tilde{b}_b$) $Y^B_{prd,m}$ at the base layer level with bl Spatial Pred (Sp pred) unit 428 (S710, Fig. 7),
   - o a first intermediate patch Y′ with the neighbor (Y^T_k) of collocated block (Y^B_k) and the block of prediction $Y^B_{prd,m}$ (S715, Fig. 7) then: formula (7)

h) similarly with this intra mode (of m index) of the base layer we determine:
   - o An intermediate intra block of prediction $X^B_{prd,m}$ at the enhancement layer level (with el Spatial Pred (Sp pred) unit 412; S720, Fig. 7),
   - o And a second intermediate patch X′ with the neighbor (X^T_k) of current block (b_e) and the intermediate block of prediction $X^B_{prd,m}$ (S725, Fig. 7) then: formula (6)

i) In the transform domain (for example, DCT) we determine the transfer function Trf from the patch Y′

to the patch X′ using the formulas (8) to (11) (S730, Fig. 7).

j) Now we consider the initial (decoded) patch of the base layer Y composed of the collocated block (Y_k^B) and its neighborY_k^T, then formula (5) (S735-S740 in Fig. 7)

1. We apply a transformation (for example, DCT) to the patch Y: $\mathbf{TF(Y)}$
2. the Trf function is now applied in the transform domain such as: $T_{Y''} = \mathbf{TF(Y)}.\boldsymbol{Trf}$
3. an inverse transform (for example, DCT^{-1}) is computed on $T_{Y''}$ giving $\mathbf{Y''} = \mathbf{TF^{-1}}(T_{Y''})$ where the resulting patch is composed as the formula (12)
4. finally the prediction which corresponds to the block $\mathbf{Y_m''^B}$ is extracted from the patch Y′′.

All the steps from f to j are realized in the "Pred el/bl (Trf)" unit 411 in FIG. 4A.

k) the error residual r_e, between the enhancement layer block $\mathbf{b_e}$ and the inter-layer prediction ($\mathbf{Y_m''^B}$) (using the combiner 402) computed at the steps f to j, is transformed and quantized re_q (T Q unit 403) and entropy coded by entropy coder unit 404 and sent in the enhancement layer bitstream

l) Finally the decoded block is locally rebuilt, by adding (with the combiner 410) the inverse transformed and dequantized prediction error block by T^{-1} Q^{-1} unit 405, re_{dq} to the prediction $\mathbf{Y_m''^B}$, and the reconstructed (or decoded) image is stored in the (el) reference frames buffer 408.

Decoder side (unit 450) in FIG. 4B:

Base layer (bl)

a) from the bl bitstream, for a given block, the entropy decoder (entropy decoder unit) 471 decodes the quantized error prediction r_{bq} and the associated coding intra mode of m index

b) the residual error prediction r_{bq} is dequantized and inverse transformed by T^{-1} Q^{-1} unit 472 to r_{bdq},

c) With help from the m intra mode, the "spatial prediction (Sp Pred)" unit 475 and "prediction" unit 474 with the decoded neighboring pixel, give the block of Intra-image prediction $\tilde{b}_b$ or $Y^B_{prd,m}$.

d) The decoded block is locally rebuilt, by adding (with the combiner 473) the decoded and dequantized prediction error block r_{bdq} to the prediction block $\tilde{b}_b$ (or $Y^B_{prd,m}$) giving the reconstructed block of the base layer.

e) The reconstructed (or decoded) frame is stored in the reference frames buffer 476, the decoded frames being used for the next (bl) intra image prediction and inter prediction (using the motion compensation unit 477).

Enhancement layer (el)

f) From the el bitstream, for a given block, the entropy decoder 451 decodes the quantized error prediction r_{eq}.

g) The residual error prediction r_{eq} is dequantized and inverse transformed by T^{-1} Q^{-1} unit 452 and output r_{edq}.

h) If the coding mode of the block to decode corresponds to our inter-layer mode, then we consider the intra mode (of m index) of the collocated block of the base layer.

i) With this intra mode (of m index) of the base layer we determine:

- Determine or re-use the intra block of prediction ($\tilde{b}_b$) $Y^B_{prd,m}$ at the base layer level (with bl Spatial Pred (Sp pred)unit 475),

- o A first intermediate patch Y′ with the neighbor (Y_k^T) of collocated block (Y_k^B) and the block of prediction $Y_{prd,m}^B$ then formula (7).

j) Similarly with this intra mode (of m index) of the base layer we determine:
- o An intermediate intra block of prediction $X_{prd,m}^B$ at the enhancement layer level with el Spatial Pred (Sp pred) unit 455,
- o And a second intermediate patch X′ with the neighbor (X_k^T) of current block (b_e) and the intermediate block of prediction $X_{prd,m}^B$ then formula (6).

k) In the transform domain (for example, DCT) we determine the transfer function Trf from the patch Y′ to the patch X′ using the formulas (8) to (11).

l) Now we consider the initial (decoded) patch of the base layer Y composed of the collocated block (Y_k^B) and its neighbor Y_k^T, then formula (5).

1. We apply a transformation (for example, DCT) to the patch Y : $\mathbf{TF(Y)}$
2. The Trf function is now applied in the transform domain such as: $T_{Y''} = \mathbf{TF(Y)}.Trf$
3. An inverse transform (for example, DCT^{-1}) is computed on $T_{Y''}$ giving $\mathbf{Y''} = \mathbf{TF^{-1}}(T_{Y''})$ where the resulting patch is composed as following:

$$\mathbf{Y''} = \begin{bmatrix} \mathbf{Y''^T} \\ \mathbf{Y_m''^B} \end{bmatrix} \quad (12)$$

4. Finally the prediction corresponds to the block

Y''^{B}_{m} is extracted from the patch Y''.

All the steps from h to l are realized in the "Pred el/bl (Trf)" unit 457, we can notice that the steps h to l are strictly the same to the steps f to j of the coder (of the first embodiment); obviously if the el coder chooses this inter-layer prediction mode by the mode decision of the el coder 407.

m) The el decoded block is built, by adding (with the combiner 453) the decoded and dequantized prediction error block r_{edq} to the prediction block Y''^{B}_{m} (via the prediction unit 454) giving the reconstructed (el) block.

n) The reconstructed (or decoded) image is stored in the (el) reference frames buffer 456, the decoded frames being used for the next (el) intra image prediction and inter prediction (using the motion compensation unit 458)

As described above, the apparatus of the first embodiment can be configured as illustrated by FIGs. 4A and 4B, by which the method of the first embodiment can be performed.

According to the method and apparatus for determining a prediction of a current block of an enhancement layer, by utilizing the coding mode of the collocated block of the base layer, the prediction of the current block of the enhancement layer can be readily and accurately obtained.

[Second Embodiment]

In the first embodiment, the intra mode of prediction of the base layer can be used in the objective to have first approximation of the current block and the collocated blocks, and the next steps correspond to the algorithm detailed with the formulas (8) through (14).

In a second embodiment, a description is given below of a more complex situation in which the encoder algorithms used to encode the base layer and the enhancement layer are different from each other, so that the modes of prediction are not compatible. A simple example can correspond to a base layer encoded with JPEG2000 (e.g., which is described in The JPEG-2000 Still Image Compression Standard, ISO/IEC JTC Standard, 1/SC29/WG1, 2005, and Jasper Software Reference Manual (Version 1.900.0), ISO/IEC JTC, Standard 1/SC29/WG1, 2005) and an enhancement layer encoded with H.264. In this situation, the first embodiment is not applicable, because the m intra mode is not available in the (for example, JPEG2000) base layer.

To solve this problem, testing the modes of prediction (available in the encoder of the enhancement layer) is performed on the pixels of the base layer to check those decoded pixels are obviously available, and finally the best intra mode is selected, according to a given criterion.

The current and the collocated patches of the enhancement and base layer are shown by the following formulas (16) and (17).

The current patch is: $X = \begin{bmatrix} X_k^T \\ X_u^B \end{bmatrix}$ (16)

The collocated patch (collocated of X) is: $Y = \begin{bmatrix} Y_k^T \\ Y_k^B \end{bmatrix}$ (17)

The selection of the best intra mode (of m index) is realized from a set S={m_0, …, m_{n-1}} of n possible intra modes (for example those corresponding to the modes shown in FIG.3). For this purpose, a virtual prediction error is computed with the virtual prediction $Y^B_{prd,J}$ (of the collocated block Y^B_k) according to a given mode of j index, and an error of virtual prediction ER_{j} between the block Y^B_k and the virtual prediction $Y^B_{prd,j}$ as shown by the following formula (18).

$$ER_{\mathrm{j}} = \sum_{\mathrm{p}\in Y^B_k}\left(Y^B_k(p) - Y^B_{prd,j}(p)\right)^2 \qquad (18)$$

Here, p corresponds to the coordinates of the pixel in the block to predict $Y^B_k$and the block of virtual prediction $Y^B_{prd,j}$; $Y^B_k(p)$ is a pixel value of the block to predict Y^B_k; and $Y^B_{prd,j}(p)$ is a pixel value of the block of virtual prediction according to the intra mode of index j.

The best virtual prediction mode is given by the minimum of the virtual prediction error from the n available intra modes prediction as the following formula (19).

$$J_{\mathbf{mode}} = \underset{\mathrm{j}}{\mathbf{Argmin}}\{ER_{\mathrm{j}}\} \qquad (19)$$

Here, it is remarked that the metric used to calculate the virtual prediction error by formula (18) is not limited to the sum of square error (SSE), other metrics are possible: sum of absolute difference (SAD), sum of absolute Hadamard transform difference (SATD).

The virtual prediction $Y^B_{prd,J_{\mathrm{mode}}}$ appropriated to the collocated block Y^B_k is obtained, and then the same mode (J_{mode}) is used so as to compute a virtual prediction ($X^{B}_{prd,J_{\text{mode}}}$) dedicated to the current block (X^{B}_{u}) of the enhancement layer.

The new intermediates patches are provided as the following formulas (20) and (21).

The current intermediate patch X′:

$$X' = \begin{bmatrix} X^{T}_{k} \\ X^{B}_{prd,J_{\text{mode}}} \end{bmatrix} \quad (20)$$

The intermediate patch Y′ of the base layer:

$$Y' = \begin{bmatrix} Y^{T}_{k} \\ Y^{B}_{prd,J_{\text{mode}}} \end{bmatrix} \quad (21)$$

Now, the process to find the (definitive) prediction of the current block from the base layer using a transfer function Trf is similar to the processing given by the previous formulas (8) and (9), once the intermediate virtual prediction blocks $Y^{B}_{prd,J_{\text{mode}}}$ and $X^{B}_{prd,J_{\text{mode}}}$ are obtained.

Having the transfer function Trf, this function is applied to the patch Y that gives, after inverse transform, the patch Y″ from which the desired prediction is extracted, as shown by formula (22).

$$Y'' = \begin{bmatrix} \mathrm{Y}''^{\mathrm{T}} \\ \mathrm{Y}''^{\mathrm{B}}_{J_{\text{mode}}} \end{bmatrix} \quad (22)$$

In formula (22), the prediction of the current block is $\mathrm{Y}''^{\mathrm{B}}_{J_{\text{mode}}}$. Here the process is similar to those used to the formula (12) by using the formulas (13), (14) and (15) with here the virtual mode J_{mode}.

The principle of this description of intra SNR scalability is illustrated in FIGs. 5A and 5B. FIG. 5 is a block diagram illustrating a configuration of an apparatus for determining a prediction of a current block of an enhancement layer of a second embodiment of the present disclosure.

Coder side (unit 500) in Fig 5A:

An original HDR image im_{el}, composed of block $\mathbf{b_e}$ 501, is tone mapped using the TMO 506 that gives the original tone mapped image im_{bl}.

Base layer (bl)

We consider the original base layer image im_{bl} to encode. With a given video encoder 531 the image is encoded with the coder 531 and locally decoded by the local in-loop decoder 532. The local decoded images are stored in the "reconstructed images buffer" 533. The resulting encoded images are sent in the base layer bitstream.

Enhancement layer (el)

We consider now the original enhancement layer block $\mathbf{b}_e$ to encode.

a) For the current block of the enhancement layer, we consider all intra coding modes available of the enhancement layer encoder intra mode (of m index),
   - We find (formula (19), with " Jmode= Argminj {ER_j}" unit 542) the best (of Jmode index) prediction mode dedicated to the collocated block (of the base layer) from the neighboring pixels of this collocated block, (according to a given criterion (formula (19)), and the encoding modes of the enhancement layer encoder).

b) With this intra mode (of Jmode index) of the enhancement layer we determine:
   - The intra block of prediction $Y^B_{prd,J_{mode}}$ at the base layer level (with bl Spatial Pred (Sp Pred) unit 541),
   - A first intermediate patch Y′ with the neighbor (Y^T_k) of collocated block (Y^B_k) and the block of prediction $Y^B_{prd,J_{mode}}$then formula (21).

c) Similarly with this intra mode (of Jmode index) of the base layer we determine:
   - An intermediate intra block of prediction $X^B_{prd,J_{mode}}$ at the enhancement layer level (with el Spatial Pred (Sp Pred) unit 512),
   - And a second intermediate patch X′ with the neighbor (X^T_k) of current block (b_e) and the intermediate block of prediction $X^B_{prd,J_{mode}}$ then formula (20).

d) In the transform domain (for example, DCT) we determine the transfer function Trf from the patch Y′ to the patch X′ using the formulas (8) to (11).

e) Now we consider the initial (decoded) patch of the base layer Y composed of the collocated block (Y^B_k) and its neighborY^T_k, then formula (5).
   1. We apply a transformation (for example, DCT) to the patch Y : $\mathbf{TF(Y)}$
   2. The Trf function is now applied in the transform domain such as: $T_{Y''} = \mathbf{TF(Y)}.Trf$
   3. An inverse transform (for example, DCT^{-1}) is computed on $T_{Y''}$ giving $\mathbf{Y''} = \mathbf{TF^{-1}}(T_{Y''})$ where the resulting patch is composed as formula (22).

4. Finally the prediction corresponds to the block $\mathbf{Y}''^{\mathbf{B}}_{J_{\mathbf{mode}}}$ is extracted from the patch Y''.

All the steps from b to e are realized in the "Pred el/bl (Trf)" unit 511.

f) The error residual (computed using the combiner 502) r_e, between the enhancement layer block $\mathbf{b}_e$ and the inter-layer prediction ($\mathbf{Y}''^{\mathbf{B}}_{J_{\mathbf{mode}}}$) computed at the steps a to e, is transformed and quantized re_q by T , Q unit 503 and entropy coded by entropy coder 504 and sent in the enhancement layer bitstream.

g) Finally the decoded block is locally rebuilt, by adding (using the combiner 514) the inverse transformed and dequantized prediction error block by T^{-1} Q^{-1} unit 505 from re_{dq} to the prediction $\mathbf{Y}''^{\mathbf{B}}_{J_{\mathbf{mode}}}$, and the reconstructed (or decoded) image is stored in the (el) reference frames buffer 508.

About the others units 507 and 509 the function is respectively dedicated to the classical coding mode decision and the motion estimation for the inter-image prediction.

Decoder side (unit 550) Fig 5B (unit 550):

Base layer (bl)

From the bl bitstream, the base layer sequence is decoded with the decoder 584. The reconstructed image buffer 582 stores the decoded frames used to the inter-layer prediction.

Enhancement layer (el)

a) From the el bitstream, for a given block, the entropy decoder 551 decodes the quantized error prediction r_{eq}

b) The residual error prediction r_{eq} is dequantized and inverse transformed by T^{-1} Q^{-1} unit 552 to generate r_{edq}.

c) If the coding mode of the block to decode corresponds to our inter-layer mode, then we need of an intra mode (of Jmode index) of the collocated block of the base layer.

  - For the current block of the HDR layer, we consider all intra coding modes available of the enhancement layer encoder intra mode (of Jmode index),
  - Find (formula (19), and " Jmode= Argminj $\{ER_j\}$" unit 581) the best (of Jmode index) prediction mode dedicated to the collocated block (of the base layer) from the neighboring pixels of this collocated block (according to a given criterion (formula (19)), and the encoding modes of the enhancement layer encoder)

d) With this intra mode (of Jmode index) of the enhancement layer we determine:

  - The intra block of prediction $Y^B_{prd,J_{mode}}$ at the base layer level with bl Spatial Pred (bl Sp Pred) unit 583,
  - A first intermediate patch Y' with the neighbor (Y_k^T) of collocated block (Y_k^B) and the block of prediction $Y^B_{prd,J_{mode}}$ then formula (21).

e) Similarly with this intra mode (of Jmode index) of the base layer we determine:
   - o An intermediate intra block of prediction $X^{B}_{prd,J_{mode}}$ at the enhancement layer level with el Spatial Pred (Sp Pred) unit 555,
   - o And a second intermediate patch X′ with the neighbor (X_k^T) of current block (b_e) and the intermediate block of prediction $X^{B}_{prd,J_{mode}}$ then formula (20).

f) In the transform domain (for example, DCT) we determine the transfer function Trf from the patch Y′ to the patch X′ using the formulas (8) to (11).

g) Now we consider the initial (decoded) patch of the base layer Y composed of the collocated block (Y_k^B) and its neighborY_k^T, then formula (5).
   1. We apply a transformation (for example, DCT) to the patch Y : $\mathbf{TF(Y)}$
   2. The Trf function is now applied in the transform domain such as: $T_{Y''} = \mathbf{TF(Y)}.Trf$
   3. An inverse transform (for example, DCT^{-1}) is computed on $T_{Y''}$ giving $\mathbf{Y''} = \mathbf{TF^{-1}}(T_{Y''})$ where the resulting patch is composed as formula (22).
   4. Finally the prediction corresponds to the block $\mathbf{Y}''^{\mathbf{B}}_{J_{\mathbf{mode}}}$ is extracted from the patch Y′′.

All the steps from c to g are realized in the “Pred el/bl (Trf)” unit 557, we can notice that the steps d to h are strictly the same to the steps b to e of the coder (of the second embodiment); obviously if the el coder chooses this inter-layer prediction mode by mode decision of the el coder (unit 507).

h) The el decoded block is built, by adding (using the combiner 553)) the decoded and dequantized prediction error block (unit 552) r_{edq} to the prediction block $Y_{J_{mode}}^{\prime\prime B}$ (via the prediction unit 554 and unit 557) giving the reconstructed (el) block.

i) The reconstructed (or decoded) image is stored in the (el) reference frames buffer 556, the decoded frames being used for the next (el) intra image prediction and inter image prediction using the motion compensation unit 558

According to the method and apparatus for determining a prediction of a current block of an enhancement layer, even when the coding mode of the base layer is different from that of the enhancement layer, the appropriate inter layer coding mode is selected, and then the prediction of the current block can be obtained.

[Third Embodiment]

A description of a method and an apparatus for determining a prediction of a current block of an enhancement layer is given below of a third embodiment of the present disclosure.

In spatial scalability, the spatial resolution of the base layer (l_e) and the enhancement layer (l_b) are different from each other, but regarding the availability of the mode of prediction of the base layer, there are different possibilities.

More specifically, a description is given below of a case in which the spatial scalability is in the same video coding standard, similarly to the first embodiment.

If the size of the current block (X_u^B) is the same as the collocated up-sampled of the block (Y_k^B) of the base layer, the prediction mode m of the base layer can be utilized, and the processing explained in the first embodiment can be applied to this case. For example (in case of spatial scalability NxN → 2Nx2N), a given 8x8 current block has a 4x4 collocated block in the base layer. Then, the intra mode m corresponds to the intra coding mode used to encode this 4x4 block (of l_b layer) and the 8x8 block of prediction $Y_{prd,m}^B$ could be the up-sampled prediction of the base layer (4x4→8x8), or the prediction $Y_{prd,m}^B$ could be computed on the up-sampled image of the base layer with the same m coding mode. As the first embodiment, once obtained the base layer and enhancement layer intermediate prediction blocks, the base layer and enhancement layer intermediate patchs are built. After from the two intermediate patchs, the transfer function is estimated using the formula 8 to 11. Finally, the transfer function is applied to the up-sampled and transformed (ex DCT) patch of the base layer, the inter layer prediction being extracted as in the first embodiment.

In contrast, if the size of the current block (X_u^B) is different from the up-sampled of the block (Y_k^B) of the base layer, the coding mode m is not really available. In this case, the principle explained in the second embodiment can be-used. In other words, the best coding mode m has to be estimated in the up-sampled base layer, the remaining processing (dedicated to the inter-layer prediction) being the same than the second embodiment;

knowing that the estimated transfer function (Trf) is applied to the up-sampled and transformed (ex DCT) base-layer patch.

[Fourth Embodiment]

A description of a method and an apparatus for determining a prediction of a current block of an enhancement layer is given below of a fourth embodiment of the present disclosure.

Based on LDR/HDR scalable video coding, a fourth embodiment of the present disclosure provides a coding mode choice algorithm for the block of the base layer, in order to re-use the selected mode to build the prediction ($l_b \rightarrow l_e$) with the technique provided in the first embodiment. The choice of the coding mode ,at the base layer level, may cause the inherent distortions at the two layers level.

Here, the RDO (Rate Distortion Optimization) technique serves to address the distortions of LDR and HDR and the coding costs of the current HDR and collocated LDR blocks, and the RDO criterion gives the prediction mode that provides the best compromise in terms of reconstruction errors and coding costs of the base and enhancement layers. To this end, the classical RDO criteria for the two layers are provided as the following formulas (23) and (24).

$$\text{LDR: } \mathrm{Cst_{bl}} = \mathrm{Dist_{bl}} + \lambda_{\mathrm{bl}}.\mathrm{B_{bl}^{cst}} \qquad (23)$$

$$\text{HDR: } \mathrm{Cst_{el}} = \mathrm{Dist_{el}} + \lambda_{\mathrm{el}}.\mathrm{B_{el}^{cst}} \qquad (24)$$

The terms $\mathrm{B_{bl}^{cst}}$ and $\mathrm{B_{el}^{cst}}$ are composed of the coding cost of the DCT coefficients of the error residual of prediction of the base layer and the enhancement layer, respectively, and the syntax elements (block size, coding mode ...) contained in the header of the blocks (B_{bl}^{cst} and B_{el}^{cst}) that allow the predictions to be rebuilt at the decoder side.

Considering the example of the block Y_{or}^{B} (being the original block) of the base layer, the quantized coefficients of the error residual of prediction after inverse quantization and inverse transform (for example, DCT^{-1}), this residual error added to the prediction provides the reconstructed (or decoded) block (Y_{dec}^{B}). With the original block Y_{or}^{B} and the decoded one Y_{dec}^{B}, the base layer distortion associated to this block is provided as the following formula (25).

$$\mathrm{Dist_{bl}} = \sum_{p \in Y_{or}^{B}} \left(Y_{or}^{B}(p) - Y_{dec}^{B}(p) \right)^2 \qquad (25)$$

In the RDO criteria, a well-known parameter λ_{bl} is used so as to give the best compromise rate distortion. In this example, the best mode, among N possible modes, is provided as the following formula (26).

$$\mathbf{J_{mode}^{bl} = \underset{j}{Argmin}\{Cst_{bl}^{j}\}} \qquad (26)$$

It is possible to re-write the formulas (23) and (24) in other form as shown by formulas (27) and (28).

$$\text{LDR: } \mathrm{Cst_{bl}^{'}} = \frac{\mathrm{Dist_{bl}}}{\lambda_{bl}} + \mathrm{B_{bl}^{cst}} \qquad (27)$$

$$\text{HDR: } \mathrm{Cst_{el}^{'}} = \frac{\mathrm{Dist_{el}}}{\lambda_{el}} + \mathrm{B_{el}^{cst}} \qquad (28)$$

The formulas (27) and (28) can be mixed with a blending parameter α that allows a global compromise between base layers and enhancement layers as the following formula (29).

$$Cst' = \left(\frac{Dist_{\mathrm{bl}}}{\lambda_{\mathrm{bl}}} + B_{bl}^{cst}\right).\left(1-\alpha\right) + \left(\frac{Dist_{\mathrm{el}}}{\lambda_{\mathrm{el}}} + B_{el}^{cst}\right).\ \alpha \qquad (29)$$

with

$$0 \le \alpha \le 1$$

The best mode (according to formula (29)) gives the mode of the base layer, which produces the minimum global cost Cst' via one of the N coding modes of the base layer as shown by the following formula (30).

$$\mathrm{J}_{\mathrm{mode}}^{\mathrm{bl}} = \underset{\mathrm{j}}{\mathrm{Argmin}}\left\{Cst'_j\right\} \qquad (30)$$

From this formula (30), the following matters are noted.

If α =0, the situation corresponds to the algorithm proposed in the first embodiment, in which the coding mode (of index m) of the base layer can be used in order to build the inter-layer prediction (bl→el) via the transfer function Trf and finally provides the inter-layer prediction $\mathrm{Y}_{\mathrm{m}}^{\prime\prime\mathrm{B}}$ with m $= \mathrm{J}_{mode}^{\mathrm{bl}}$

On the contrary, if α =1, the choice of the coding mode principally focuses on the enhancement layer, and there is a risk of the base layer containing a lot of visual artifacts.

If α =0.5, a compromise between the two layers is necessary. In this case, it is important to notice that the choice of coding mode of the base layer is really based on the impact not only at the base layer level but also at the enhancement layer level, more precisely :

- The impact on the base layer according to the choice of the base layer coding mode
- And the impact on the enhancement layer using the entire process explained in the first embodiment i.e. the inter layer prediction based on the previous base layer coding mode

FIG. 6 shows a block diagrams illustrating an apparatus for determining a prediction of a current block of an enhancement layer of the fourth embodiment.

With reference to FIG. 6, local inter-layer prediction is described. For the description, only the intra image prediction mode, using the intra mode (m) is described, because our inter layer prediction mode uses intra mode (m).

Notice that, only the coder side is described because in the fourth embodiment the associated decoder is the same than the first embodiment and corresponds to the decoder illustrated by the Fig 4.b.

Coder side (unit 600) in FIG. 6:

An original block 601 b_e is tone mapped using the TMO 606 that gives the original tone mapped block b_{bc}.

Notice that in the specific case of inter-layer prediction of the fourth embodiment, the units 625 and 607 (corresponding to the coding mode decision units of the base and enhancement layers) are not used. In that case the unit

642 replace the units 625 and 607, in fact the unit 642 selects the best intra J_{mode}^{bl} mode using the formula 30 and sends that mode ($J_{\mathrm{mode}}^{\mathrm{bl}}$) to the units 625 and 607.

Base layer intra coding mode selection ($J_{\mathrm{mode}}^{\mathrm{bl}}$)in unit 642

For a given blending parameter α that allows a global compromise between base layers and enhancement layers as the following formula (29, and for each N available intra prediction modes (illustrated with the FIG. 3, in case of H264) We operate N iterations on the coding modes:

Loop on N intra modes of m index {

a) With the neighboring reconstructed (or decoded) pixels of the base layer the spatial prediction and the intra coding mode m (m being an index), the (Sp Pred) unit 658 gives an intra base layer prediction block

b) With the neighboring reconstructed (or decoded) pixels of the enhancement layer the spatial prediction and the same m intra coding mode (Sp Pred) unit 612 gives an intermediate intra enhancement layer prediction block

- The unit 611 builds the patch of the base layer composed of the intra base layer neighbor and the block of prediction of the step (a)
- The unit 611 builds the patch of the enhancement layer composed of intra enhancement layer neighbor and the block of prediction of the step (b)
- In the transform domain (for example, DCT) determine (in unit 611)the transfer function Trf

from the patch Y′ to the patch X′ using the formulas (8) to (11).

- o Still in unit 611,
  - ▪ consider the initial (decoded) patch of the base layer Y composed of the collocated block (Y_k^B) and its neighbor Y_k^T, then formula (5)
  - ▪ apply a transformation (for example, DCT) to the patch Y: $\mathbf{TF(Y)}$
  - ▪ apply the Trf function is applied in the transform domain such as: $T_{Y''} = \mathbf{TF(Y)}.Trf$
  - ▪ inverse transform (for example, DCT^{-1}) $T_{Y''}$ giving $\mathbf{Y''} = \mathbf{TF^{-1}}(T_{Y''})$ where the resulting patch is composed as the formula (12)
  - ▪ extracted the prediction corresponding to the block $\mathbf{Y_m''^B}$ from the patch Y′′

c) In units 642, the best mode (according to formula (29)) is selected, which produces the minimum global cost $\mathrm{Cst'}$ via one of the N coding modes (formula (30))

} end Loop on N intra modes of m index

Finally the best intra $J_{\mathrm{mode}}^{\mathrm{bl}}$ is sent to the base layer spatial prediction unit 658 and decision unit 607 and to the enhancement layer unit 611.

Once the $J_{\mathrm{mode}}^{\mathrm{bl}}$ found, the remaining of the process is similar to the description of coder of the first embodiment, knowing that the base layer intra mode index m = $J_{\mathrm{mode}}^{\mathrm{bl}}$.

Base layer (bl)

We consider the original base layer block b_{bc} to encode

d) With the original block b_{bc} and the (previous decoded) images stored in the reference frames buffer 626, the motion estimator (motion estimation unit) 629 finds the best inter image prediction block with a given motion vector (temporal prediction unit) and the temporal prediction (Temp Pred Pred) unit 630 gives the temporal prediction bloc

e) If the mode decision process (unit 625) chooses the intra image prediction mode (of m= J_{mode}^{bl} index, the residual error prediction rb is computed (by the combiner 621) with the difference between the original block b_{bc} and the prediction block $\tilde{b}_b$ ($Y_{prd,m}^{B}$)

f) After, the residual error prediction rb is transformed and quantized to r_{bq} by T Q unit 622 and finally entropy coded by entropy coder unit 623 and sent in the bitstream base layer.

g) The decoded block is locally rebuilt, by adding (with the combiner 657) the inverse transformed and dequantized by T^{-1} Q^{-1} unit 624 prediction error block r_{bdq} to the prediction block $\tilde{b}_b$ giving the reconstructed (base layer) block

h) The reconstructed (or decoded) frame is stored in the (bl) reference frames buffer 626.

<u>Enhancement layer (el)</u>

We can notice that the structure of the coder of the enhancement layer is similar to the coder of the base layer, for example the units 607, 608, 609 and 613 have the same function than the respective units 625, 626, 629 and

630 of the coder of the base layer in terms of coding mode decision, temporal prediction and reference frames buffer. We consider now the original enhancement layer block b_e to encode.

i) For the block of the enhancement layer, if the collocated block of the base layer is coded in intra image mode, then we consider the intra mode (of m index with m= J^{bl}_{mode}) of this collocated block.

j) With this intra mode (of m index) of the base layer we determine:
  - determine or re-use the intra block of prediction ($\tilde{b}_b$) $Y^B_{prd,m}$ at the base layer level with bl Spatial Pred (Sp pred) unit 658,
  - a first intermediate patch Y′ with the neighbor (Y^T_k) of collocated block (Y^B_k) and the block of prediction $Y^B_{prd,m}$ then: formula (7)

k) similarly with this intra mode (of m index) of the base layer we determine:
  - An intermediate intra block of prediction $X^B_{prd,m}$ at the enhancement layer level (with el Spatial Pred (Sp pred) unit 612),
  - And a second intermediate patch X′ with the neighbor (X^T_k) of current block (b_e) and the intermediate block of prediction $X^B_{prd,m}$ then: formula (6)

l) In the transform domain (for example, DCT) we determine the transfer function Trf from the patch Y′ to the patch X′ using the formulas (8) to (11).

m) Now we consider the initial (decoded) patch of the base layer Y composed of the collocated block (Y^B_k) and its neighborY^T_k, then formula (5)

5. We apply a transformation (for example, DCT) to the patch Y: $TF(Y)$

6. the Trf function is now applied in the transform domain such as: $T_{Y''} = \mathbf{TF}(\mathbf{Y}).\boldsymbol{Trf}$
7. an inverse transform (for example, DCT^{-1}) is computed on $T_{Y''}$ giving $\mathbf{Y}'' = \mathbf{TF}^{-1}(T_{Y''})$ where the resulting patch is composed as the formula (12)
8. finally the prediction corresponds to the block $\mathbf{Y}_{\mathbf{m}}^{\prime\prime\mathbf{B}}$ is extracted from the patch Y''.

All the steps from j to m are realized in the "Pred el/bl (Trf)" unit 611.

n) the error residual r_e, between the enhancement layer block $\mathbf{b_e}$ and the inter-layer prediction ($\mathbf{Y}_{\mathbf{m}}^{\prime\prime\mathbf{B}}$) (using the combiner 602) computed at the steps j to m, is transformed and quantized re_q (T Q unit 603) and entropy coded by entropy coder unit 604 and sent in the enhancement layer bitstream

o) Finally the decoded block is locally rebuilt, by adding (with the combiner 610) the inverse transformed and dequantized prediction error block by T^{-1} Q^{-1} unit 605, re_{dq} to the prediction $\mathbf{Y}_{\mathbf{m}}^{\prime\prime\mathbf{B}}$, and the reconstructed (or decoded) image is stored in the (el) reference frames buffer 608.

As described above, the embodiments of the present disclosure relates to the SNR and spatial scalable LDR/HDR video encoding with the same or different encoders for the two layers. The LDR video can be implemented from the HDR video with any tone mapping operators: global or local, linear or non-linear. In the scalable solution of the embodiments, the inter layer prediction is implemented on the fly without additional specific meta-data.

The embodiments of the present disclosure concern both the encoder and the decoder. The embodiments of the present disclosure applied to decoding processes generally disclosed, and the decoding is detectable according to the embodiments of the present disclosure.

The embodiments of the present disclosure can be applied to image and video compression. In particular, the embodiments of the present disclosure may be submitted to the ITU-T or MPEG standardization groups as part of the development of a new generation encoder dedicated to the archiving and distribution of LDR/HDR video content.

All examples and conditional language recited herein are intended for pedagogical purposes to aid the reader in understanding the disclosure and the concepts contributed by the inventor to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions, nor does the organization of such examples in the specification relate to a showing of the superiority or inferiority of the disclosure.

CLAIMS

1. A method comprising:

building (S715) a first intermediate patch of a low dynamic range with the neighboring pixels of the collocated block of the base layer and a first prediction block predicted from neighboring pixels of a collocated block of a base layer with a coding mode of the base layer;

building (S725) a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and a second prediction block predicted from neighboring pixels of a current block of an enhancement layer with the coding mode;

building (S735) a patch by applying a transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain, wherein the transfer function is determined (S730) to transform the first intermediate patch to the second intermediate patch in a transform domain;

predicting (S740) a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block in the patch being collocated to the current block of the enhancement layer in the second intermediate patch; and

encoding a residual error between the current block of the enhancement layer and the prediction of the current block of the enhancement layer.

2. The method as claimed in claim 1, wherein the base layer is tone mapped using a tone mapping operator dedicated to a low dynamic range video.

3. The method as claimed in claim 1, wherein a first coding mode of the collocated block of the base layer is used for the coding mode when the first coding mode is available for the current block of the enhancement layer.

4. The method as claimed in claim 1, wherein the coding mode is obtained by selecting a most appropriate coding mode from possible coding modes when a first coding mode of the collocated block of the base layer is not available for the current block of the enhancement layer.

5. The method as claimed in claim 4, wherein the selecting the most appropriate coding mode is performed by selecting a coding mode that minimizes a difference between the collocated block of the base layer and a virtual prediction of the collocated block of the base layer with each of the possible coding modes of the enhancement layer.

6. The method as claimed in claim 1, wherein a first coding mode of the collocated block of the base layer is used for the coding mode if the size of the current block of the enhancement layer is the same as the size of up-sampled collocated block of the base layer.

7. The method as claimed in claim 1, wherein a first coding mode of the collocated block of the base layer is selected by taking into account a compromise in terms of reconstruction errors in the base and enhancement layers and coding costs of the base and enhancement layers.

8. An apparatus (400) comprising:

a first intermediate patch creation unit (428) configured to predict a first prediction block from neighboring pixels of the collocated block of a base layer with a coding mode of the base layer and to build a first intermediate patch of a low dynamic range with the neighboring pixels of the collocated block of the base layer and the first prediction block;

a second intermediate patch creation unit (412) configured to predict a second prediction block from neighboring pixels of a current block of an enhancement layer with the coding mode and to build a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and the second prediction block;

a unit (411) to determine a transfer function to transform the first intermediate patch to the second intermediate patch in a transform domain, to build a patch by applying the transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain and to predict a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block being in the patch collocated to the current block of the enhancement layer in the second intermediate patch; and

an encoder (404) to encode a residual error between the current block of the enhancement layer and the prediction of the current block of the enhancement layer.

9. The apparatus as claimed in claim 8, wherein the base layer is tone mapped using a tone mapping operator dedicated to a low dynamic range video.

10. The apparatus (400) as claimed in claim 8, wherein a first coding mode of the collocated block of the

base layer is used as the coding mode when the first coding mode is available for the current block of the enhancement layer.

11. The apparatus (500) as claimed in claim 8, wherein a most appropriate coding mode from possible coding modes is selected when a first coding mode of the collocated block of the base layer is not available for the current block of the enhancement layer.

12. The apparatus (500) as claimed in claim 11, wherein the most appropriate coding mode is selected by selecting a coding mode that minimizes a difference between the collocated block of the base layer and a virtual prediction of the collocated block of the base layer with each of the possible coding modes of the enhancement layer.

13. The apparatus (400) as claimed in claim 8, wherein a first coding mode of the collocated block of the base layer is used for the coding mode if the size of the current block of the enhancement layer is the same as the size of up-sampled collocated block of the base layer.

14. The apparatus (600) as claimed in claim 8, wherein a first coding mode of the collocated block of the base layer is selected by taking into account a compromise in terms of reconstruction errors in the base and enhancement layers and coding costs of the base and enhancement layers.

15. A method comprising:
decoding a residual prediction error;
building (S715) a first intermediate patch of

a low dynamic range with the neighboring pixels of the collocated block of the base layer and a first prediction block predicted from neighboring pixels of a collocated block of a base layer with a coding mode of the base layer;

building (S725) a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and a second prediction block predicted from neighboring pixels of a current block of an enhancement layer with the coding mode;

building (S735) a patch by applying a transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain, wherein the transfer function is to transform the first intermediate patch to the second intermediate patch in a transform domain;

predicting (S740) a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block in the patch being collocated to the current block of the enhancement layer in the second intermediate patch; and

reconstructing a block of the enhancement layer by adding the prediction error to the prediction of the current block of the enhancement layer.

16. An apparatus (450) comprising:

a decoder (451) for decoding a residual prediction error;

a first intermediate patch creation unit (475) configured to build a first intermediate patch of a low dynamic range with the neighboring pixels of a collocated block of a base layer and a first prediction block predicted from neighboring pixels of a collocated block of a base layer with a coding mode of the base layer;

a second intermediate patch creation unit (455) configured to build a second intermediate patch of a high dynamic range with the neighboring pixels of the current block of the enhancement layer and a second prediction block predicted from neighboring pixels of a current block of an enhancement layer with the coding mode and;

a unit (457) to build a patch by applying the transfer function to a transformed initial patch of the base layer in a transform domain and then applying an inverse transform to the resulting patch so as to return in a pixel domain, wherein the transfer function is to transform the first intermediate patch to the second intermediate patch in a transform domain and to predict a prediction of the current block of the enhancement layer by extracting a block from the patch, the extracted block being in the patch collocated to the current block of the enhancement layer in the second intermediate patch; and

a unit (453) to add the prediction error to the prediction of the current block of the enhancement layer to reconstruct a block of the enhancement layer.

FIG.1

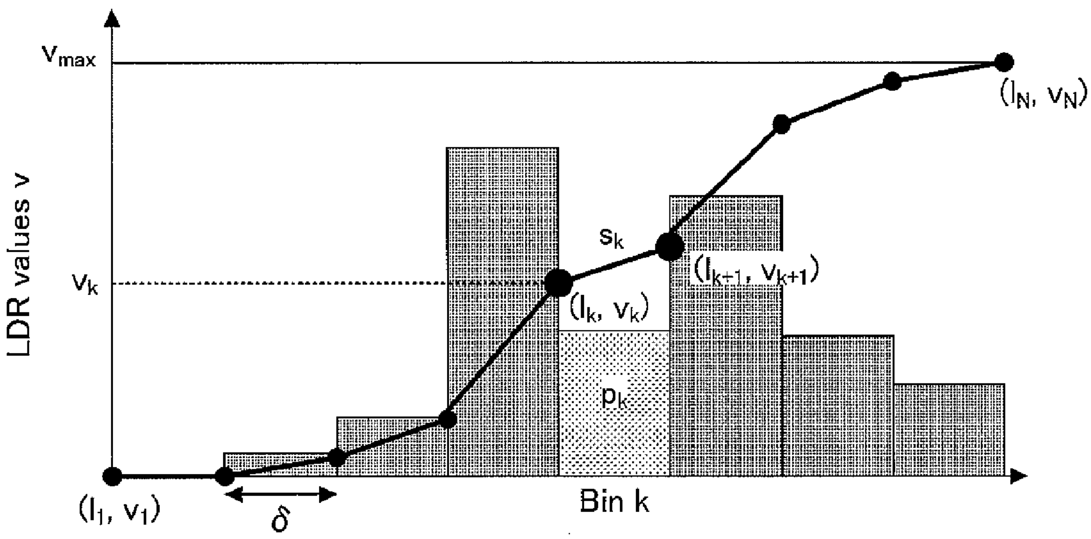


FIG.2A

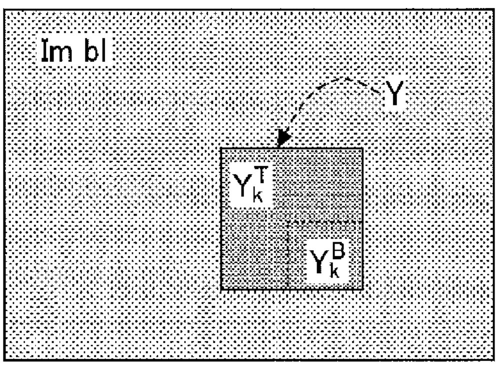


FIG.2B

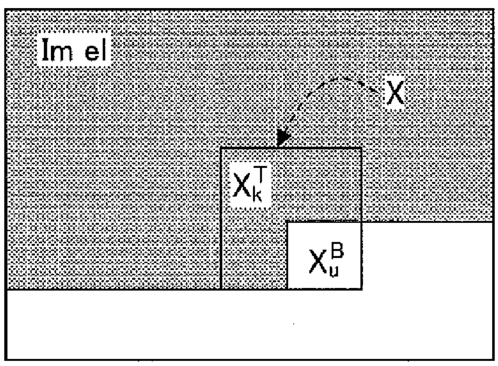


FIG.3A

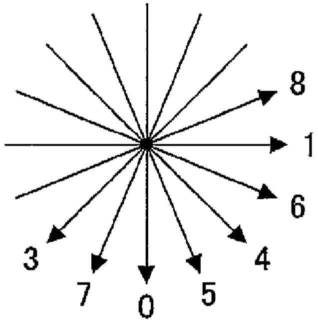
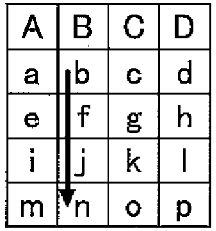
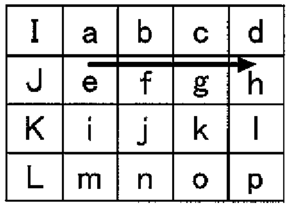


FIG.3B



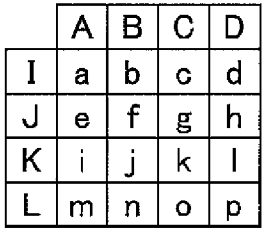
MODE 0
VERTICAL

FIG.3C



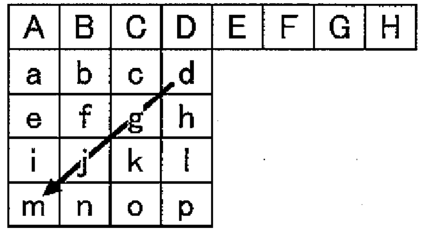
MODE 1
HORIZONTAL

FIG.3D



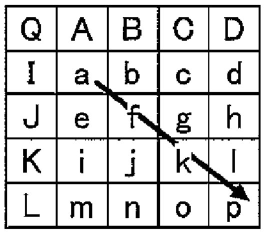
MODE 2
DC

FIG.3E



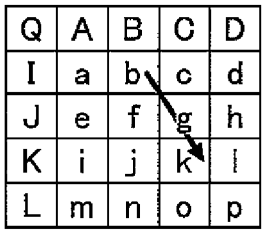
MODE 3
DIAGONAL
DOWN-LEFT

FIG.3F



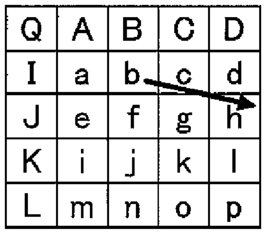
MODE 4
DIAGONAL
DOWN-RIGHT

FIG.3G



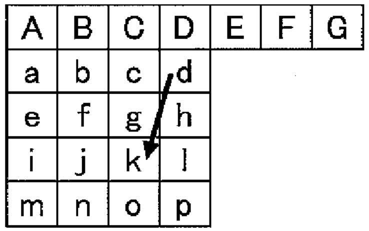
MODE 5
VERTICAL RIGHT

FIG.3H



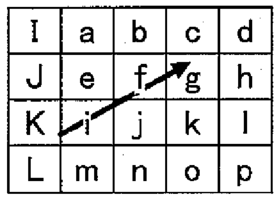
MODE 6
HORIZONTAL
DOWN

FIG.3I



MODE 7
VERTICAL LEFT

FIG.3J



MODE 8
HORIZONTAL UP

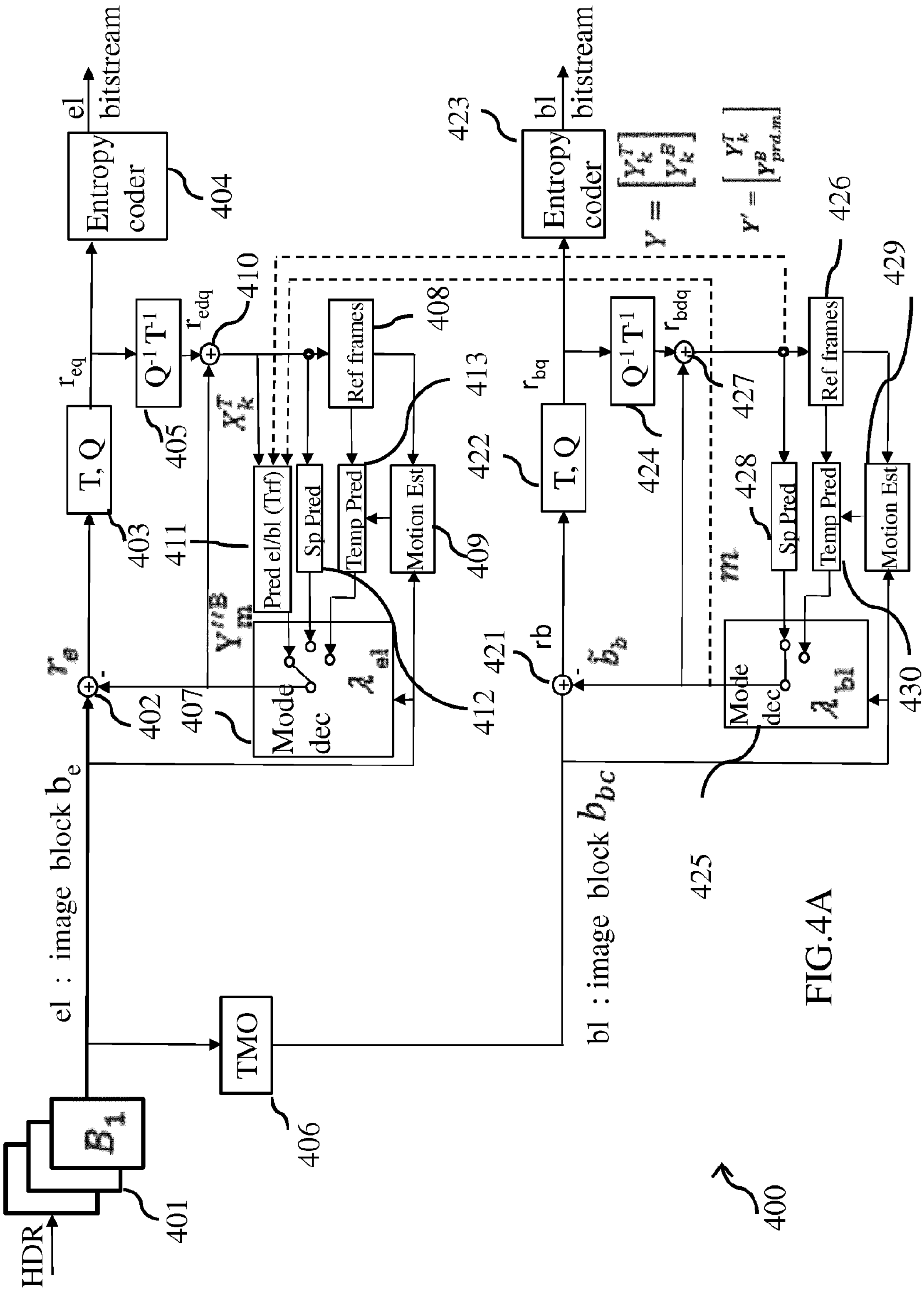

HDR
401
406
TMO
el : image block be
bl : image block bbc
402
403
T, Q
req
Q-1 T-1
405
redq
410
Entropy coder
404
el bitstream
407
Mode dec
λel
411
Pred el/bl (Trf)
Sp Pred
Temp Pred
Motion Est
Ref frames
408
409
412
413
421
rb
422
rbq
424
rbdq
427
423
bl bitstream
425
Mode dec
λbl
428
426
429
430
400

FIG.4A

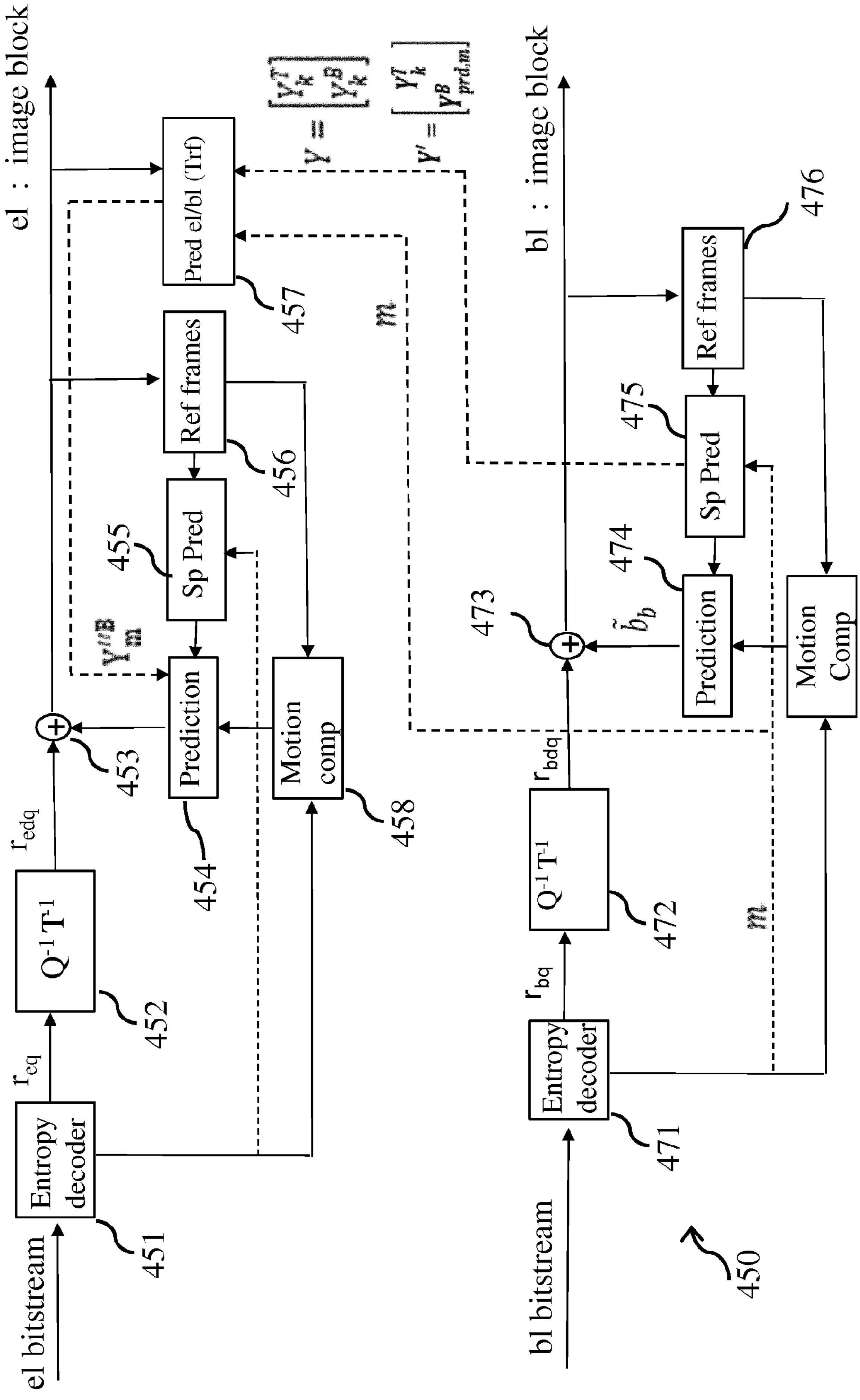

el : image block
Pred el/bl (Trf)
Ref frames
Sp Pred
Prediction
Motion comp
Entropy decoder
Q-1 T-1
el bitstream
r_eq
r_edq
Y''B m
m
Y =
Y' =
451
452
453
454
455
456
457
458
bl : image block
Ref frames
Sp Pred
Prediction
Motion Comp
Entropy decoder
Q-1 T-1
bl bitstream
r_bq
r_bdq
m
471
472
473
474
475
476
477
450

FIG.4B

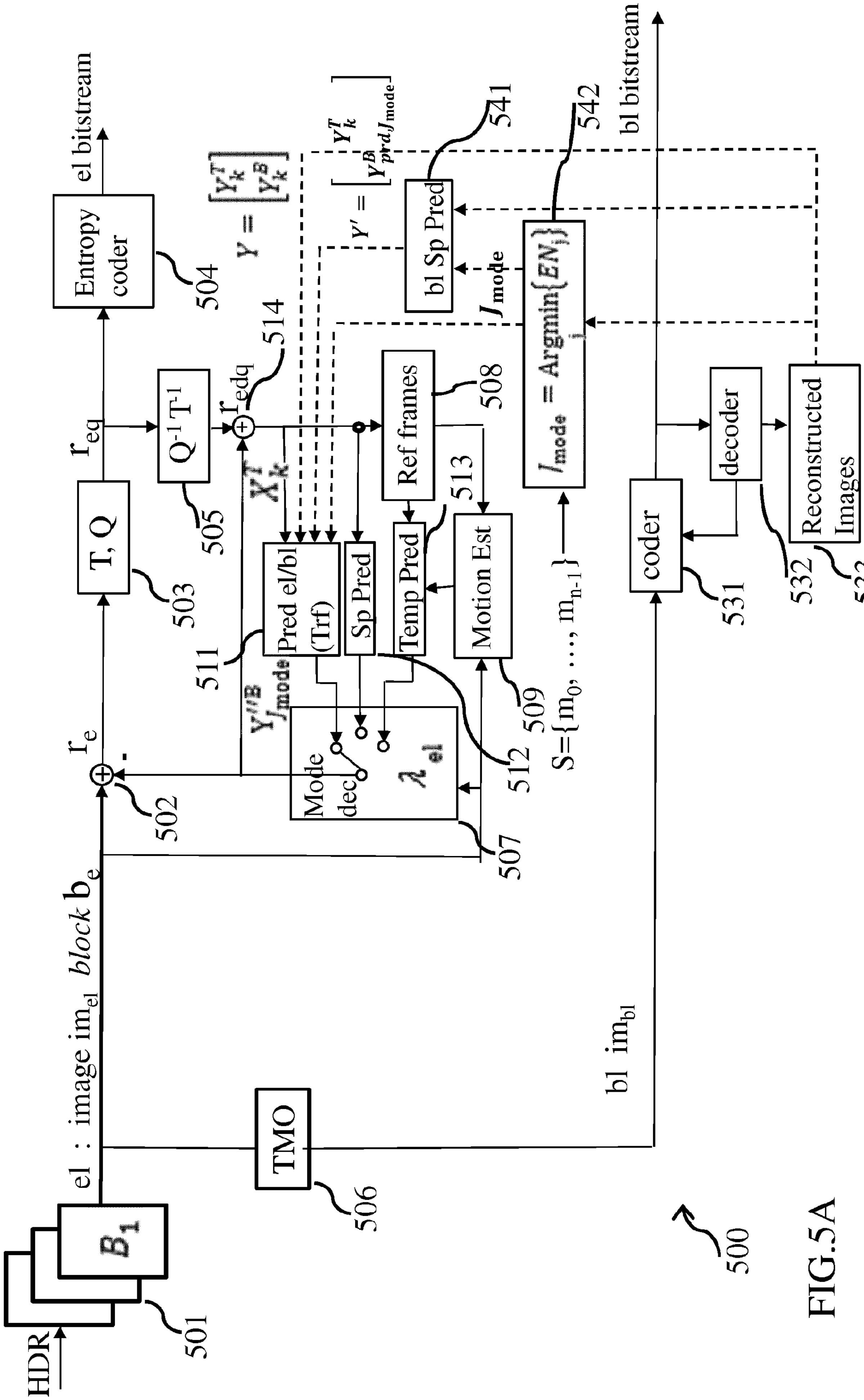

HDR
501
el : image im_el block b_e
TMO
506
502
r_e
T, Q
503
r_eq
Q^-1 T^-1
505
r_edq
514
Entropy coder
504
el bitstream
Y = [Y_k^T Y_k^B]
Y' = [Y_k^T Y^B_prd J_mode]
bl Sp Pred
541
I_mode = Argmin{EN_j}
542
J_mode
X_k^T
Y''^B_J_mode
Pred el/bl (Trf)
511
Sp Pred
512
Temp Pred
513
Ref frames
508
Motion Est
509
Mode dec
λ_el
507
S={m_0, ..., m_n-1}
bl im_bl
coder
531
decoder
532
Reconstructed Images
533
bl bitstream
500

FIG.5A

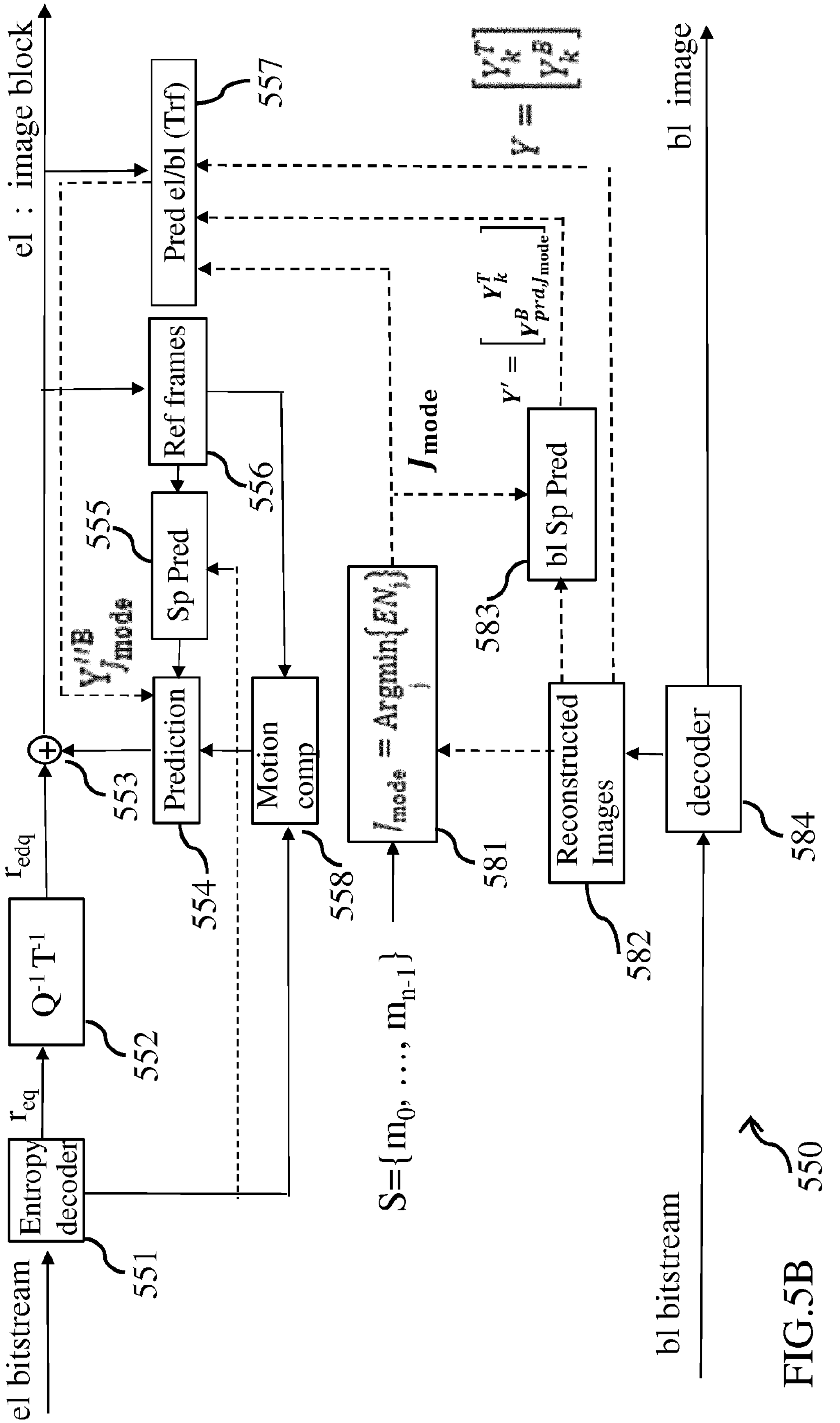

el bitstream
Entropy decoder
551
r_{eq}
Q^{-1} T^{-1}
552
r_{edq}
553
Prediction
554
Sp Pred
555
Ref frames
556
Pred el/bl (Trf)
557
Motion comp
558
el : image block
$Y''^{B}_{J_{mode}}$
$S=\{m_0, \ldots, m_{n-1}\}$
$J_{mode} = \text{Argmin}_j\{EN_j\}$
581
J_{mode}
Reconstructed Images
582
bl Sp Pred
583
$Y' = \begin{bmatrix} Y_k^T \\ Y^B_{prdJ_{mode}} \end{bmatrix}$
$Y = \begin{bmatrix} Y_k^T \\ Y_k^B \end{bmatrix}$
decoder
584
bl bitstream
bl image
550

FIG.5B

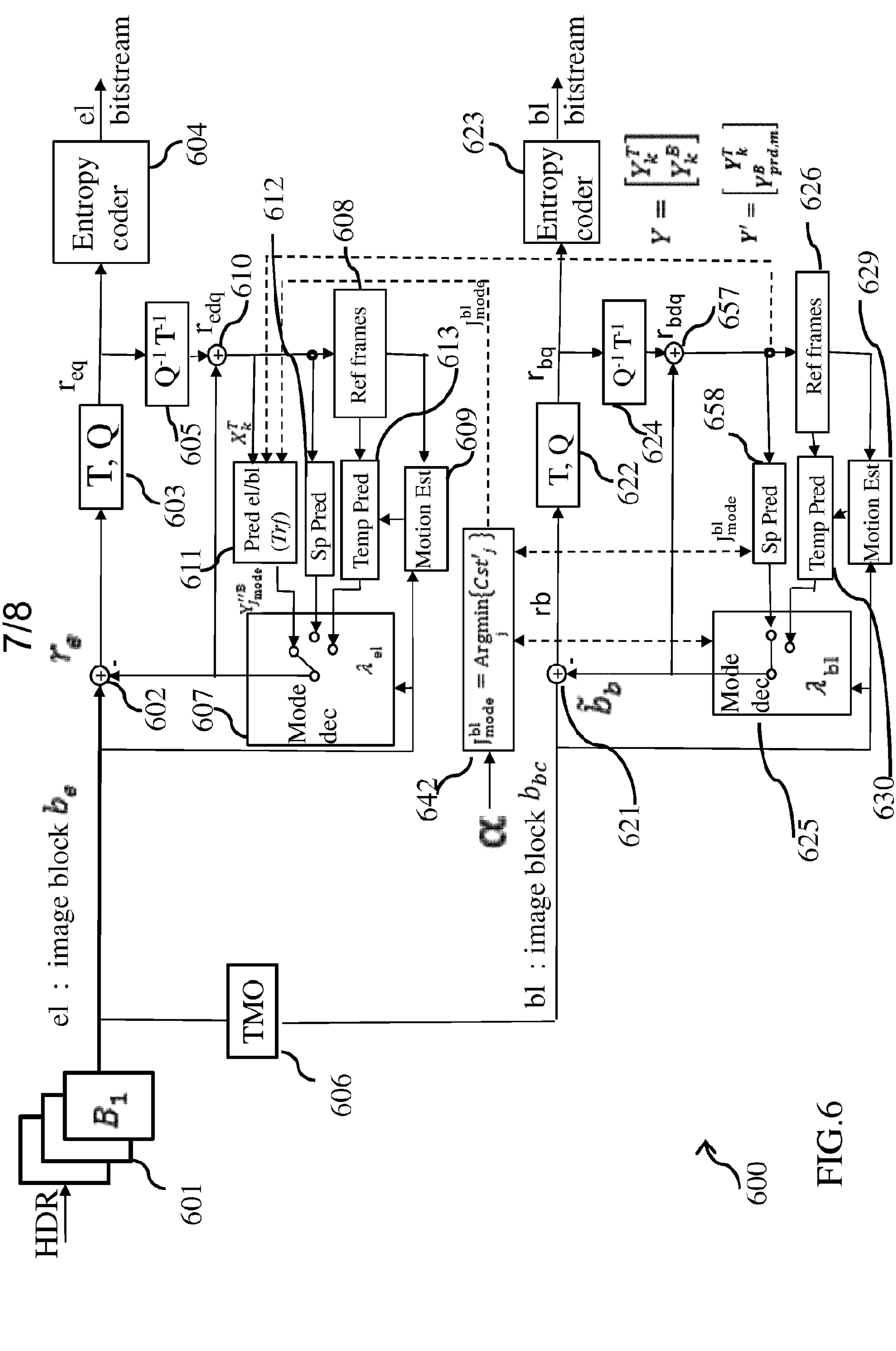

7/8
HDR
B_1
601
TMO
606
el : image block b_e
bl : image block b_{bc}
r_e
602
603
T, Q
r_{eq}
Entropy coder
604
el bitstream
$Q^{-1}T^{-1}$
605
r_{edq}
610
612
608
Ref frames
613
Motion Est
609
Temp Pred
Sp Pred
Pred el/bl (Trf)
611
X_k^T
$Y''^B_{J_{mode}}$
J^{bl}_{mode}
Mode dec
λ_{el}
607
$J^{bl}_{mode} = \mathrm{Argmin}_j\{Cst'_j\}$
642
α
rb
$\tilde{b}_b$
621
T, Q
622
r_{bq}
$Q^{-1}T^{-1}$
624
r_{bdq}
657
Entropy coder
623
bl bitstream
$Y = \begin{bmatrix} Y_k^T \\ Y_k^B \end{bmatrix}$
$Y' = \begin{bmatrix} Y_k^T \\ Y^B_{prd,m} \end{bmatrix}$
626
Ref frames
629
658
Sp Pred
Temp Pred
Motion Est
630
Mode dec
λ_{bl}
625
600

FIG.6

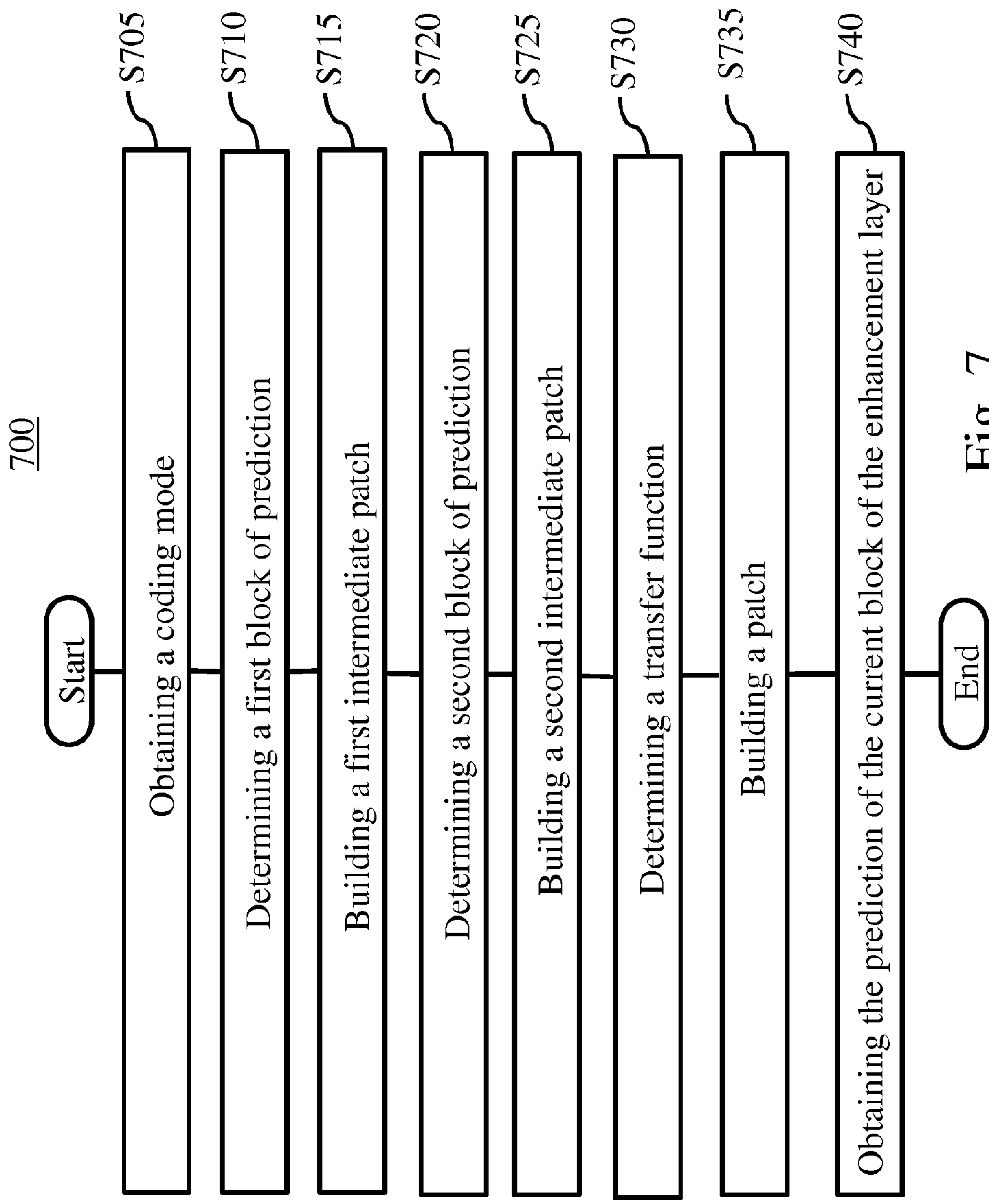

700
Start
Obtaining a coding mode
S705
Determining a first block of prediction
S710
Building a first intermediate patch
S715
Determining a second block of prediction
S720
Building a second intermediate patch
S725
Determining a transfer function
S730
Building a patch
S735
Obtaining the prediction of the current block of the enhancement layer
S740
End

Fig. 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/064868

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N19/11 H04N19/157 H04N19/176 H04N19/593
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.
Y	WU Y ET AL: "CE1: SVC study on inter-layer prediction: bit-depth scalability", 25. JVT MEETING; 82. MPEG MEETING; 21-10-2007 - 26-10-2007; SHENZHEN,CN; (JOINT VIDEO TEAM OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16),, no. JVT-Y081, 21 October 2007 (2007-10-21) , XP030007285, ISSN: 0000-0137 page 2 first paragraph -----	1-16

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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
1 August 2016	09/08/2016

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	Regidor Arenales, R

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/064868

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WINKEN M ET AL: "CE2: SVC bit-depth scalability", 24. JVT MEETING; 81. MPEG MEETING; 29-6-2007 - 5-7-2007; GENEVA, CH;(JOINT VIDEO TEAM OF ISO/IEC JTC1/SC29/WG11 AND ITU-T SG.16),, no. JVT-X057, 4 July 2007 (2007-07-04), XP030007164, ISSN: 0000-0082 figure 1	1-16
Y	WO 2014/082982 A1 (THOMSON LICENSING) 5 June 2014 (2014-06-05) page 3, line 14 - line 24 page 14, line 10 - line 16	1-16
Y	US 2014/140392 A1 (XU JUN [US] ET AL) 22 May 2014 (2014-05-22) paragraph [0140] - paragraph [0146]	3-5, 10-12
A	WO 2014/106746 A1 (CANON KK [JP]; CANON EUROP LTD [GB]) 10 July 2014 (2014-07-10) page 27, line 20 - line 32	1-16

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2016/064868

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014082982 A1	05-06-2014	EP 2926557 A1	07-10-2015
		US 2015304656 A1	22-10-2015
		WO 2014082982 A1	05-06-2014
US 2014140392 A1	22-05-2014	NONE	
WO 2014106746 A1	10-07-2014	GB 2509901 A	23-07-2014
		GB 2509998 A	23-07-2014
		US 2015341638 A1	26-11-2015
		WO 2014106746 A1	10-07-2014