

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 July 2007 (12.07.2007)

PCT

(10) International Publication Number
WO 2007/077116 A1

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

Yuan, Xue Qing Road, Hai Dian District, Beijing 100083 (CN).

(21) International Application Number:
PCT/EP2006/069852

(74) Agents: **LE DANTEC, Claude** et al.; THOMSON, 46, Quai Alphonse Le Gallo, F-92100 Boulogne Billancourt (FR).

(22) International Filing Date:
18 December 2006 (18.12.2006)

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

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
06290049.3 5 January 2006 (05.01.2006) EP

(71) Applicant (for all designated States except US): **THOMSON LICENSING** [FR/FR]; 46 Quai A. Le Gallo, F-92100 Boulogne-Billancourt (FR).

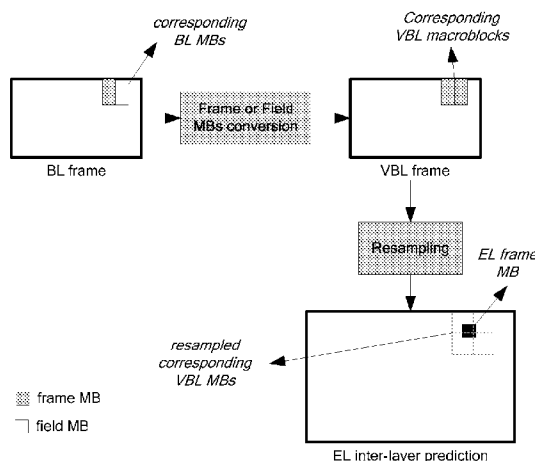
(72) Inventors; and

(75) Inventors/Applicants (for US only): **FRANCOIS, Edouard** [FR/FR]; 18 Allee Du Locar, F-35890 Bourges-Comptes (FR). **BOTTREAU, Vincent** [FR/FR]; La Riaudais Broons Sur Vilaine, F-35220 Chateaubourg (FR). **LOPEZ, Patrick** [FR/FR]; 4 Allée Du Clos De La Merceraie, F-35450 Livré Sur Changeon (FR). **VIERON, Jérôme** [FR/FR]; 5 Allee Jean Perrin, F-35137 Bedee (FR). **CHEN, Ying** [CN/CN]; Bld. 3, 103, Chenghua

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: INTER-LAYER MOTION PREDICTION METHOD



(57) Abstract: The invention relates to a method for generating for at least one macroblock of a high resolution image at least one motion predictor from motion data associated to macroblock(s) of a low resolution image. The method comprises the following steps: - defining a virtual structure made up of non overlapping macroblocks, called virtual base layer macroblocks; - identifying for the high resolution macroblock corresponding virtual base layer macroblocks in the virtual image on the basis of the location of the high resolution macroblock within the high resolution image and on the basis of the ratios of low resolution and high resolution images dimensions; - deriving, for each identified virtual base layer macroblocks, motion data from motion data associated to at least one macroblock of a corresponding low resolution macroblocks pair on the basis of the field/frame mode of the corresponding low resolution macroblocks pair and on the basis of the field/frame mode of the high resolution macroblock; - resampling motion data associated to the corresponding virtual base layer macroblocks so as to generate the at least one motion predictor.

WO 2007/077116 A1



Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTER-LAYER MOTION PREDICTION METHOD

1. Field of the invention

The invention relates to a method for generating for blocks of pixels of
5 images of a high resolution sequence at least one motion predictor from
motion data associated to blocks of pixels of images of a low resolution
sequence.

2. Background of the invention

10 State-of-art scalable hierarchical coding methods allow to encode the
information hierarchically in order that it can be decoded at different resolution
and/or quality levels. A data stream generated by a scalable coding device is
thus divided into several layers, a base layer and one or more enhancement
layers. These devices allow to adapt a unique data stream to variable
15 transmission conditions (bandwidth, error rate ...) and also to the capacities of
reception devices (CPU, characteristics of reproduction device...). A spatially
scalable hierarchical encoding (or decoding) method encodes (or decodes) a
first part of data called base layer relating to low resolution images also called
base layer images (BL images), and from this base layer encodes (or
20 decodes) at least another data part called enhancement layer relating to high
resolution images also called enhancement layer images (EL images). The
coding information or coding data relating to enhancement layer is possibly
inherited (i.e. derived) from coding information relating to the base layer by a
method called inter-layer prediction method or inter-layer inheriting method.
25 Each macroblock of a high resolution image is thus temporally predicted
according to a classical prediction mode (e.g. bidirectional prediction mode,
direct prediction mode, forward/backward prediction ...) or is predicted
according to an inter-layer prediction mode. In this former case, motion
information/data associated to a high resolution macroblock (e.g. partitioning
30 pattern associated to macroblock/block of pixels of the high resolution image
(for splitting said macroblock/block into several sub-blocks), coding modes
associated to said blocks, possibly motion vectors and one or more image
reference indices associated to some blocks allowing to reference the image
used to predict said block) has to be derived or inherited from motion data

associated to macroblocks of low resolution images whatever the format of the low or high resolution images is, i.e. progressive or interlaced.

3. Summary of the invention

5 The invention relates to a method, as part of a process of coding or decoding video signal, for generating for at least one macroblock of a high resolution image, called high resolution macroblock, at least one motion predictor from motion data associated to macroblock(s) of a low resolution image, called low resolution macroblock(s). The method comprises the
10 following steps:

- defining a virtual structure made up of non overlapping macroblocks, called virtual base layer macroblocks, the virtual structure having same dimensions as the low resolution image;
- identifying for the high resolution macroblock corresponding virtual base
15 layer macroblocks in the virtual image on the basis of the location of the high resolution macroblock within the high resolution image and on the basis of the ratios of low resolution and high resolution images dimensions;
- deriving, for each identified virtual base layer macroblocks, motion data
20 from motion data associated to at least one macroblock of a corresponding low resolution macroblocks pair on the basis of the field/frame mode of the corresponding low resolution macroblocks pair and on the basis of the field/frame mode of the high resolution macroblock;
- resampling motion data associated to the corresponding virtual base
25 layer macroblocks so as to generate the at least one motion predictor.

If the high resolution macroblock is in frame mode and if the corresponding low resolution macroblocks pair is in field mode, the deriving step comprises the steps of:

- upsampling vertically by a ratio of two the low resolution macroblocks pair
30 made up of a top macroblock and a bottom macroblock, each upsampled macroblock being made up of an up part and a down part; and
- merging motion data of the up part of the upsampled top macroblock with motion data of the up part of the upsampled bottom macroblock and merging

motion data of the down part of the upsampled top macroblock with motion data of the down part of the upsampled bottom macroblock.

If the high resolution macroblock is in field mode and if the corresponding low resolution macroblocks pair is in frame mode, the deriving
5 step comprises a step for downsampling vertically by a ratio of two motion data associated to the corresponding low resolution macroblocks pair.

The invention also relates to a device for coding a sequence of high
10 resolution images made up of macroblocks, called high resolution macroblocks, and a sequence of low resolution images, called low resolution macroblocks. The device comprises:

- first coding means for coding the low resolution images, the first coding means generating motion data for macroblocks of the low resolution images and a base layer data stream;
- 15 – inheritance means for deriving motion data for at least one macroblock of a high resolution image, called high resolution macroblock from motion data of macroblocks a low resolution image, called low resolution macroblocks; and
- second coding means for coding the high resolution images using the
20 derived motion data, the second coding means generating an enhancement layer data stream.

The inheritance means comprise:

- means for defining a virtual structure made up of non overlapping
25 macroblocks, called virtual base layer macroblocks, the virtual structure having same dimensions as the low resolution image;
- means for identifying for the high resolution macroblock corresponding virtual base layer macroblocks in the virtual image on the basis of the location of the high resolution macroblock within the high resolution image and on the basis of the ratios of low resolution and high resolution images
30 dimensions;
- means for deriving, for each identified virtual base layer macroblocks, motion data from motion data associated to at least one macroblock of a corresponding low resolution macroblocks pair on the basis of the

field/frame mode of the corresponding low resolution macroblocks pair and on the basis of the field/frame mode of the high resolution macroblock; and - means for resampling motion data associated to the corresponding virtual base layer macroblocks so as to generate the at least one motion predictor.

According to a specific embodiment, the first coding means are an MPEG-4 AVC video encoder.

The invention also relates to a device for decoding at least a coded sequence of high resolution images, the coded images arising in the form of a data stream, each image being divided in non overlapping macroblocks. The device comprising:

- first decoding means for decoding at least a first part of the data stream in order to generate low resolution images and motion data for macroblocks of the low resolution image;
- inheritance means for deriving motion data for at least one macroblock of a high resolution image from motion data of macroblocks of a low resolution image; and
- second decoding means for decoding at least a second part of the data stream using the derived motion data in order to generate high resolution images.

The inheritance means comprise:

- means for defining a virtual structure made up of non overlapping macroblocks, called virtual base layer macroblocks, the virtual structure having same dimensions as the low resolution image;
- means for identifying for the high resolution macroblock corresponding virtual base layer macroblocks in the virtual image on the basis of the location of the high resolution macroblock within the high resolution image and on the basis of the ratios of low resolution and high resolution images dimensions;
- means for deriving, for each identified virtual base layer macroblocks, motion data from motion data associated to at least one macroblock of a corresponding low resolution macroblocks pair on the basis of the

field/frame mode of the corresponding low resolution macroblocks pair and on the basis of the field/frame mode of the high resolution macroblock;

- 5 - means for resampling motion data associated to the corresponding virtual base layer macroblocks so as to generate the at least one motion predictor.

According to a specific embodiment, the first decoding means are an MPEG-4 AVC video decoder.

10 4. Brief description of the drawings

Other features and advantages of the invention will appear with the following description of some of its embodiments, this description being made in connection with the drawings in which:

- 15 - Figure 1 depicts a pair of macroblocks located vertically and coded either in frame mode (left part of the figure) or in field mode (right part of the figure);
- Figure 2 depicts a flowchart of the method according to the invention;
- Figure 3 illustrates the generation process of virtual base layer TOP and BOTTOM field macroblocks from base layer TOP and BOTTOM interlace pair of macroblocks;
- 20 - Figure 4 illustrates UP and DOWN progressive macroblocks vertical downsampling process;
- Figure 5 illustrates TOP and BOTTOM macroblocks merging process;
- 25 - Figure 6 illustrates interlace macroblock pair merging process;
- Figure 7 illustrates progressive BL macroblocks (vertical) downsampling process;
- Figure 8 depicts a block-diagram of BL progressive macroblocks cleaning process;
- 30 - Figure 9 illustrates 4x4 and 8x4 sub-partitions merging process;
- Figure 10 depicts a block-diagram of downsampled macroblock derivation process;

- Figure 11 illustrates different vertical downsampling cases;
- Figure 12 illustrates interlace BL TOP and BOTTOM macroblocks merging process;
- Figure 13 illustrates TOP and BOTTOM macroblocks in field image;
- 5 – Figure 14 illustrates TOP and BOTTOM macroblocks in frame image (MBAFF);
- Figure 15 illustrates the two steps of interlace TOP and BOTTOM macroblocks merging process;
- Figure 16 depicts a block-diagram of upsampled macroblocks content merging process;
- 10 – Figure 17 depicts virtual base layer structures;
- Figure 18 depicts the inter-layer prediction process of a given EL MB according to the invention;
- Figure 19 depicts a encoding device according to the invention; and
- 15 – Figure 20 depicts a decoding device according to the invention.

5. Detailed description of the preferred embodiments

In the SVC standard currently defined by JVT (MPEG & ITU) in the document JVT-Q202 from ISO/IEC MPEG & ITU-T VCEG entitled « Scalable Video Coding – Joint Draft 4 » and referenced as JSVM4 in the sequel, the spatial scalability is only considered for progressive material. Motion inter-layer prediction between two (or more) successive spatial layers (the base one and enhancement one(s)) are only addressed in case of progressive video sequences. The invention proposes to extend these inter-layer prediction processes in order to support any combinations of interlace/progressive scalability.

According to many video coding standards an interlaced image, which comprises a top field interlaced with a bottom field captured at different time instant, may be coded either as two field images (field image mode), i.e. the two fields of the image are coded separately, or as a frame image (frame image mode), i.e. the image is coded as a single frame. MPEG-4 AVC described in the document ISO/IEC JTC 1/SC 29/WG 11 N5546 entitled « Text of 14496-2 Third Edition » allows that decision to be made either

independently for the entire image or independently for every two vertical macroblock-pair. When the decision is made at image level it is called PAFF coding (PAFF stands for Image Adaptive Frame/Field) and when the decision is made at macroblock-pair level it is called MBAFF (stands for Macroblock Adaptive Frame/Field). More precisely, according to MPEG-4 AVC, when an interlaced image is encoded as a frame image and if MBAFF is allowed, then each pair of vertical macroblocks (MBs) may be encoded as interlaced, i.e. field mode (right part of the figure 1), or as progressive, i.e. frame mode (left part of the figure 1). On Figure 1, the grey lines represent the even lines of the images interlaced with the odd lines, i.e. the white lines, of the images. When a pair of vertical macroblocks is in field mode, each MB of the pair is said to be a field macroblock. When a pair of vertical macroblocks is in frame mode, each MB of the pair is said to be a frame macroblock.

The method according to the invention allows to generate at least one inter-layer motion predictor for a macroblock of the high resolution image (also called Enhancement Layer MacroBlock and noted EL MB) from motion data associated to some macroblocks of the low resolution images (also called Base Layer MacroBlock and noted BL MB) whatever the format (interlace or progressive) of the high and low resolution sequences. An inter-layer motion predictor, more simply called motion predictor, associated to a macroblock of a high resolution image is a macroblock to which is associated motion data (e.g. partition/sub-partition type, possibly reference indices allowing to identify reference images, motion vectors ...). According to MPEG-AVC, for each image of a sequence of images, the reference indices are selected among two reference lists (list0 et list1) defined as follows:

- each reference index is associated to a given reference image number;
- for each list, some reference indices are marked as invalid; and
- each 8x8 block, i.e. block of size 8 by 8 pixels, can use for each list a single reference index.

An invalid reference index corresponds to reference fields or frames not yet available (decoded) in the decoding process.

The motion predictor is generated using a motion downsampling method as the ESS (stands for Extended Spatial Scalability) method described in the JSVM4 document. If the high resolution image under consideration is interlaced, then two inter-layer prediction modes are potentially available to
5 encode or decode each EL macroblock:

1. inter-layer interlace (IL_I) mode: macroblock corresponding to TOP or BOTTOM field data, i.e. a field macroblock.
2. inter-layer progressive (IL_P) mode : macroblock corresponding to FRAME data, i.e. a frame macroblock.

10

The invention proposes a method for generating these inter-layer macroblock predictors. To this aim the invention consists in defining at step 190 "Virtual Base Layer" (VBL) macroblocks from the BL MBs instead of modifying in depth the method described in JSVM4. A "Virtual Base Layer" (VBL) macroblock is a structure to which motion data is associated. The use
15 of VBL structures avoids or at least significantly limits the JSVM4 upsampling process modifications, since most of the issues related to interlace are reported in the VBL data derivation. Indeed, "Virtual Base Layer" (VBL) macroblocks ensure the compatibility between BL and EL macroblocks and facilitates the direct re-use of inter-layer prediction defined in JSVM4 for
20 progressive images. Instead of applying the resampling process directly to the BL macroblocks, BL macroblocks are first converted into VBL macroblocks that all have a field/frame mode compatible with the EL macroblock field/frame mode. Practically, this method requires the use of the following structures: a
25 VBL Frame (with same dimension as BL Images), a VBL TOP field and a VBL BOTTOM field (with same width as BL Images and half height as BL Images) forming a VBL Field with same dimensions as BL image, as illustrated in Figure 17. Each VBL structure is divided into macroblocks called VBL Frame MBs, VBL TOP field MBs, and VBL BOTTOM field MBs to which motion data
30 is associated. The classical ESS downsampling method described in JSVM4 for deriving motion information for MBs of high resolution images from BL MBs is then applied using the VBL MBs instead of the original BL MBs. The inter-

layer motion prediction method may be part of a process for coding sequence of images and part of a process for decoding sequence of images. In an encoding process, all potential inter-layer motion predictors are computed. The encoding process then selects to encode current EL MB using either an

5 inter-layer prediction mode, or a classical prediction mode. If an inter-layer prediction mode is selected by the encoding process, then only the corresponding inter-layer predictor is computed by the decoding process.

In reference to Figure 2, the inter-layer motion prediction process for a current EL MB whose inter-layer mode IL_mode is either IL_P , i.e. frame mode, or

10 IL_I , i.e. field mode, achieves the following steps:

- Identify 200 the corresponding VBL macroblocks $\{c_0, c_1, \dots, c_N\}$;
- for each corresponding VBL macroblock c_i , $i=0..N$, positioned at (x_i, y_i) , derive 210 the associated motion data from motion data associated to BL macroblocks depending on the inter-layer prediction mode; and
- 15 - derive 220 of the inter-layer motion predictor from respectively VBL MBs, using the ESS process as described in JSVM4 with a ratio equal to h_{enh}/h_{base} in the vertical direction and a ratio equal to w_{enh}/w_{base} in the horizontal direction, where h_{enh} is the height and w_{enh} the width of the HL pictures and where h_{base} is the height and w_{base} the width of the BL picture.

20

The inter-layer mode IL_mode at the decoding side is derived from the bitstream. At the encoding side this mode is derived by a coding mode decision method. Such a method is for example based on a posteriori criteria, i.e. it selects the mode that minimizes a linear combination of the distortion

25 and of the coding cost.

Figure 18 illustrates as an example the inter-layer motion prediction process in a case where the low resolution image is interlaced and encoded as a frame image in MBAFF mode and where the high resolution image is interlaced and encoded as a frame image in MBAFF mode. Given an EL macroblock (filled in

30 black, and considered as a frame macroblock in this example), its corresponding VBL macroblocks are first identified. The corresponding BL

macroblocks may be of different frame/field mode than the EL macroblock. For example, on figure 18, the right pair of BL macroblocks is made up of two field macroblocks while the EL macroblock is in frame mode. In this case motion data associated to BL macroblocks is converted. This converted
 5 motion data information is therefore associated to VLB macroblocks which have the same frame/field mode as the EL macroblock. Finally, inter-layer prediction is performed from VBL macroblocks to EL macroblock using the 'standard' progressive resampling process described in JSVM4.

Frame-Field macroblocks conversion is required in the following cases.

10 *Case 1* : BL and EL frames are encoded with MBAFF mode, EL MB is field coded and corresponding BL MBs are frame coded. This requires *frame-to-field* conversion.

Case 2 : The EL MB is frame coded, the BL is a frame image and corresponding BL MBs are field coded. In that case, *field-to-frame* conversion
 15 is required.

Case 3 : BL image is progressive frame; EL MB is field coded in a frame image or belongs to a field image; and the EL field height $H_e/2$ is lower than the BL frame height H_b . In that case, *frame-to-field* conversion is required.

20 The step 200 consist in identifying the corresponding VBL macroblocks {c0,c1,...cN}. If EL MB mode is IL_P, i.e. VBL Frame MBs are used, then the corresponding VBL MBs will be determined using the following process:

$$x_s = [x * w_{base} / w_{enh}] / 16$$

$$y_s = [y * h_{base} / h_{enh}] / 16$$

25

$$x_e = [(x+15) * w_{base} / w_{enh}] / 16$$

$$y_e = [(y+15) * h_{base} / h_{enh}] / 16$$

where (x,y) are the top left sample position of the EL MB under consideration, (w_{base} , h_{base}) the base images dimensions and (w_{enh} , h_{enh}) the enhancement
 30 images dimensions. The corresponding VBL macroblocks {c0,c1,...cN} are those located inside window delimited by top left position (x_s,y_s) and bottom right position (x_e,y_e).

If EL MB mode is IL_I, i.e. VBL TOP or BOTTOM Field MBs are used, then corresponding VBL MBs will be determined using the following process:

$$x_s = [(x-x_0) * W_{base} / W_{enh}] / 16$$

$$y_s = [(y-y_0) * (h_{base}/2) / h_{enh}] / 16$$

$$5 \quad x_e = [(x+15-x_0) * W_{base} / W_{enh}] / 16$$

$$y_e = [(y+15-y_0) * (h_{base}/2) / h_{enh}] / 16$$

The corresponding VBL macroblocks {c0,c1,...cN} are those located inside window delimited by top left position (xs,ys) and bottom right position (xe,ye).

10 The step 210 consists in generating the VBL MBs and more particularly in deriving motion data associated to the VBL MBs. VBL TOP MB is generated as follows depending on BL image coding mode:

- IF BL image is coded as a Frame Image, then if BL MB positioned at (xi,2*yi) is progressive, the VBL TOP MB ci is a vertically downsampled version of the
15 MB pair comprising the BL MB positioned at (xi,2*yi+1) and the BL MB positioned at (xi,2*yi). A new downsampling process is defined in the sequel to derive a vertically downsampled macroblock version of the pair of progressive macroblocks as depicted on Figure 3 else if BL MB positioned at (xi,2*yi) is interlace, the motion data associated to VBL TOP MB ci is the copy
20 of the motion data associated to TOP BL MB at position (xi,2*yi).

- IF BL image is coded as a Field Image, the motion data associated to VBL TOP MB ci is the copy of the motion data associated to TOP BL MB at position (xi,yi).

VBL BOTTOM MB is generated as follows depending on BL image coding
25 mode:

- IF BL image is coded as a Frame Image, then if BL MB positioned at (xi,2*yi+1) is progressive, the VBL BOTTOM MB ci is a vertically downsampled version of the MB pair comprising the BL MB positioned at (xi,2*yi+1) and the BL MB positioned at (xi,2*yi). It is generated using the new
30 downsampling process as depicted on Figure 4 else if BL MB positioned at (xi,2*yi+1) is interlace, the motion data associated to VBL BOTTOM MB ci is the copy of the motion data associated to BOTTOM BL MB at position (xi,2*yi+1).

- IF BL image is coded as a Field Image, the motion data associated to VBL BOTTOM MB ci is the copy of the motion data associated to BOTTOM BL MB at position (xi,yi) .

VBL FRAME MB is generated as follows depending on BL image coding mode:

5

- IF BL image is coded as a Frame Image, as depicted on Figure 5, if BL MB positioned at (xi,yi) is progressive, then the motion data associated to the VBL Frame MB ci is a direct copy of the motion data associated to the BL MB positioned at (xi,yi) else if the BL MB positioned at (xi,yi) is interlace, the VBL

10

Frame MB ci is a merged version of the MB pair comprising the BL MB positioned at (xi,yi) and the BL MB positioned at $(xi,yi+1-2*(yi\%2))$. A new merging process is defined in the sequel that allows to merge 2 macroblocks into a progressive macroblock

In this case, the VBL Frame MBs UP and DOWN of a MB pair are identical because they are generated from the same TOP and BOTTOM BL MBs.

15

- IF BL image is coded as a Field Image, as depicted on Figure 6,

VBL Frame MB ci is a merged version of BL TOP field MB at position $(xi,yi/2)$ and BL BOTTOM field MB at position $(xi,yi/2)$. This VBL Frame MB ci is generated using the new merging process.

20

The new downsampling process for deriving a vertically downsampled macroblock version of pairs of progressive macroblocks applies on a pair of BL macroblocks (UP and DOWN), at position $(mbX, 2*mbY)$ and $(mbX, 2*mbY+1)$, encoded in progressive format. It consists in a vertical downsampling by a ratio of 2 (one example is depicted on Figure 7). The process outputs a downsampled macroblock at position (mbX, mbY) .

25

As depicted in Figure 7, each of the two BL UP and DOWN macroblocks generate a 16×8 partition after the vertical downsampling. So basically the vertically downsampled macroblock is split in two 16×8 partitions. The vertical downsampling process could generate configurations forbidden by MPEG-4 AVC standard, such as:

30

- different reference indices inside 8×8 sub-partitions (for instance if one of the BL macroblocks is partitioned in four 8×8 blocks and if 2 neighboring

up and down 8x8 blocks have not the same reference index then it would generate, once downsampled, an 8x8 block split into two 8x4 blocks having different reference indices); or

- sub-partitions with height of 2 pixels (for instance if an input 8x8 block is sub-partitioned in 8x4, it would generate once downsampled sub-partitions of size 8x2).

Therefore a cleaning process is first applied on the BL UP and DOWN MBs to avoid these configurations. Each macroblock of the pair of progressive BL macroblocks is first cleaned-up for avoiding partitions with invalid motion vectors. The process, depicted in Figure 8 for one progressive BL macroblock noted MB_{BL} applies successively on the UP and DOWN MB of the pair. It basically achieves the three following steps:

- removal 800 of invalid partitions;
- reference indices homogenization 810; and
- merging 820 of 4x4 and 4x8 sub-partitions.

The step 800 consists in removing invalid partitions, i.e. those using invalid reference index, corresponding to reference field or frame not yet available in the decoding process.

If MB_{BL} mode is not INTRA, the following process is applied:

- For each list (0 and 1), the following applies:
 - First, partitions having an invalid reference index are classified as INVALID and other partitions are classified as VALID.
 - Then if at least one partition is VALID, then for each partition classified as INVALID, removal of invalid partitions is achieved. For instance, the sub-partitioning, motion vectors and reference index of a neighboring partition of the macroblock not classified as INVALID are copied to the INVALID partition. Neighboring partitions may be scanned in the following order: horizontal, vertical, diagonal. The partition is then classified as VALID.

If for both lists, all partitions are INVALID, MB_{BL} mode is set to INVALID.

- If MB_{BL} mode is INTRA, then nothing is done.

The step 810, consists in homogenizing reference indices within MB_{BL} partition. It applies only if MB_{BL} mode is not set to INVALID.

- If BL macroblock partition is equal to 16x8 or 8x8, the following process is applied for each up partition of the macroblock.
 - If at least for one list (0 or 1), the down partition of the macroblock has a different reference index as its up partition, the following applies for each list l (0 and 1).
 - the up or down partition using the minimum reference index rl is identified (noted refPart).
 - The other partition is enforced to have the same reference index rl , sub-partitioning (if applicable) and motion vectors as refPart.
- 10 Possibly, at step 815, a partition merging process is applied. Partitions are merged into a single partition if they are not sub-partitioned and if for each list their reference index and motion vector are identical. Macroblock mode is modified as follows:
- 16x8 and 8x16 are transformed into 16x16.
- 15 - 8x8 is transformed as follows:
- 8x16 if left 8x8 blocks are merged together and right 8x8 blocks are merged together;
 - 16x8 if up 8x8 blocks are merged together and down 8x8 blocks are merged together; and
- 20 - 16x16 if the four 8x8 blocks are merged together.

The step 815 is optional.

- The step 820 consists in merging 4x4 and 8x4 sub-partitions. It applies only if MB_{BL} mode is not set to INVALID. If MB_{BL} mode is neither equal to INVALID nor to INTRA and macroblock partition is equal to 8x8, sub-partitions merging is applied to each 8x8 block of the macroblock with sub-partitioning equal to 8x4 or 4x4. 8x4 sub-partitions are merged into 8x8 sub-partitions, and 4x4 sub-partitions are merged into 4x8 sub-partitions as depicted on figure 9. For each merged area, for each list (0 and 1), the merged area motion vector is set to the average motion vector of the two parts to be merged.

- 30 After the cleaning process a downsampled macroblock named MB_D is generated. It has two corresponding BL cleaned macroblocks, BMB_{Up} and

BMbDown. The figure 10 depicts the flowchart of the process applied to generate the downsampled macroblock MB_D .

If BMbUp and BMbDown are INTRA, MB_D mode is set to INTRA.

5 Otherwise if BMbUp and BMbDown are INTRA or INVALID (three cases are considered here, BMbUp is INTRA and BMbDown is INVALID or BMbUp is INVALID and BMbDown is INTRA or both macroblocks BMbUp and BMbDown are INVALID), MB_D mode is set to INVALID.

Otherwise, the following process applies:

10 At step 1000, MB_D is split in two 16x8 parts (noted 16x8PartUp and 16x8PartDown in the sequel). Each 16x8 part has one corresponding BL frame macroblock (16x8PartUp corresponds to BMbUp, 16x8PartDown corresponds to BMbDown).

At step 1200, each 16x8 part 16x8PartZ (with Z being replaced by Up or Down) is built using the following process (illustrated in Figure 11).

- 15
- if BMbZ is INVALID or INTRA, 16x8PartZ is classified as NEUTRAL.
 - otherwise if BMbZ partition is 16x16, 16x8PartZ partition is 16x8 (Figure 11 a). For each list, 16x8PartZ reference index is those of BMbZ.
- 20
- otherwise 16x8PartZ partition is 8x8. The two 8x8 blocks of 16x8PartZ are then sub-partitioned as follows.
 - if BMbZ partition is 16x8, sub-partitioning is 8x4 for both blocks (Figure 11 b).
 - otherwise if BMbZ partition is 8x16, sub-partitioning is 8x8 for
- 25
- otherwise if BMbZ partition is 8x8, for W being replaced by left and right, the following applies (example in Figure 11 d).
 - if at least one of the W BMbZ 8x8 blocks is split in 4x8, the W 8x8 block sub-partitioning is 4x4.
- 30
- otherwise the W 8x8 block sub-partitioning is 8x4.
 - Each derived partition / sub-partition of 16x8PartZ inherits its reference indices and motion vectors as follows.

- The top left 4x4 block coordinates (b4X,b4Y) of the considered partition / sub-partition are identified.
- The reference indices and motion vectors of the 4x4 block of BMbZ with coordinates (b4X , b4Y / 2) are copied to the considered partition / sub-partition.

5

The step 1200, consists in removing NEUTRAL partition if any. If one of the two 16x8 parts is classified as NEUTRAL, its content (partitioning / sub-partitioning, reference indices and motion vectors) is copied from the other 16x8 part content.

10 The step 1300, consists in deriving MB_D partitioning as follows:

- If 16x8PartUp and 16x8PartDown partitions are 16x8, MB_D partitioning is 16x8
- Otherwise, MB_D partitioning is set to 8x8.

Possibly, at step 1400, the partition merging process of step 815 is applied.

15 This step 1400 is optional.

The new merging process for merging a TOP macroblock (i.e. coming from BL TOP field) and a BOTTOM macroblock (i.e. coming from BL BOTTOM field) into a pair of progressive macroblocks (UP and DOWN) applies to two BL macroblocks encoded in interlaced format. The first considered macroblock is a TOP macroblock BMbTop. The second one is a BOTTOM macroblock BMbBot. The output of the process is a pair of progressive UP and DOWN macroblocks as depicted on figure 12. BMbTop and BMbBot are defined as follows:

20

25

- if BL image is encoded as a Field image, BMbTop is the macroblock of BL TOP field and BMbBot is the macroblock of BL BOTTOM field at position (mbX, mbY) as depicted on Figure 13;
- otherwise (BL image is encoded as a FRAME image), BMbTop and BMbBot are macroblocks of BL Frame at position (mbX, 2*mbY) and (mbX, 2*mbY+1) – this only applies if the BL Frame macroblocks pair is encoded in interlace mode as depicted on Figure 14.

30

In reference to Figure 15, the process comprises the two following steps:

- a step 1500 for vertical upsampling of BMbTop and BMbBot with a ratio of 2, this step generating two upsampled TOP macroblocks, namely MbTopUp and MbTopDown, and two upsampled BOTTOM macroblocks, namely MbBottomUp and MbBottomDown.
- 5 - A step 1510 for merging the upsampled macroblocks MbTopUp and MbBottomUp into one macroblock MbUp, and the upsampled macroblocks MbTopDown and MbBottomDown into one macroblock MbDown.

The step 1500 consists in macroblock upsampling with a ratio 2 in the vertical direction. This step consists in applying directly the ESS upsampling process defined in JSVM4 (Joint Video Team (JVT) of ISO/IEC MPEG & ITU-
 10 T VCEG JVT-Q202, "Scalable Video Coding – Joint Draft 4", J.Reichel, H.Schwarz, M.Wien) with a ratio of 1 in the horizontal direction and a ratio of two in the vertical direction.

The step 1510 applies on two macroblocks Mbln1 and Mbln2 and
 15 generates one merged macroblock MbOut. In reference to figure 15, if the macroblocks Mbln1 and Mbln2 refer to macroblocks MbTopUp and MbBotUp respectively then MbOut refers to MbUp and if the macroblocks Mbln1 and Mbln2 refer to macroblocks MbTopDown and MbBotDown respectively then MbOut refers to MbDown.

20 According to a first embodiment, Mbln1 associated motion data is associated to MbOut.

- If MbOut is not INTRA, the invalid partitions are removed by the invalid partitions removal process described previously;
- If MbOut is classified INVALID, Mbln2 associated motion data is
 25 associated to MbOut. Then the following process applies
 - If MbOut is not INTRA, the invalid partitions are removed by the invalid partitions removal process described previously.

According to a preferred embodiment, in reference to Figure 16, the step 1510 consists in applying the following sub-steps:

30 At sub-step 1600, Mbln1 associated motion data is associated to MbOut. Then, if MbOut is not INTRA, the following process applies:

- For each list (0 and 1), the following applies:

- At sub-step 1610, partitions having an invalid reference index are classified as INVALID and other partitions are classified as VALID. INVALID partitions are processed as follows:
 - At sub-step 1620, MbOut partitioning is enforced to 8x8. The INVALID partition is divided in 8x8 partitions. MbOut mode is enforced to 8x8. For each 8x8 partition, the following applies:
 - At sub-step 1630, if the corresponding 8x8 block in Mbln2 uses a valid reference index, sub-partitioning (if any), reference index and motion vectors of this Mbln2 8x8 block are copied into the considered MbOut 8x8 partition. This partition is classified as VALID. Otherwise, the MbOut 8x8 partition is classified as INVALID.
 - If at least one MbOut 8x8 partition is VALID, then, at sub-step 1640, for each partition classified as INVALID, removal of invalid partitions is achieved. For instance the sub-partitioning, motion vectors and reference index of a neighboring partition of the macroblock not classified as INVALID are copied to the INVALID partition. Neighboring partitions may be scanned in the following order: horizontal, vertical, diagonal. The partition is then classified as VALID.

If for both lists, all partitions are INVALID, mode is set to INVALID.

At sub-step 1650, MbOut partitions may be merged by applying the partition merging process of step 815. This sub-step 1650 is optional.

In Figures 2, 8, 10 and 16, the represented boxes are purely functional entities, which do not necessarily correspond to physical separated entities. Namely, they could be developed in the form of software, or be implemented in one or several integrated circuits.

The method for generating inter-layers motion predictors may be used of a process for coding video signals and of a process for coding video signals as the one described in JSVM4.

The invention also concerns a coding device 8 depicted on figure 19. The coding device 8 comprises a first coding module 80 for coding the low resolution images. The module 80 generates a base layer data stream and motion data for said low resolution images. Preferentially the module 80 is adapted to generate a base layer data stream compatible with MPEG4 AVC standard. The coding device 8 comprises inheritance means 82 used to derive motion data for high resolution images from the motion data of the low resolution images generated by the first coding module 80. The inheritance means 82 are adapted to implement the steps 200 to 220 of the method according to the invention. The coding device 8 comprises a second coding module 81 for coding the high resolution images. The second coding module 81 uses the motion data derived by the inheritance means 82 in order to encode the high resolution images. The second coding module 81 thus generates an enhancement layer data stream. Preferentially, the coding device 8 also comprises a module 83 (for example a multiplexer) that combines the base layer data stream and the enhancement layer data stream provided by the first coding module 80 and the second coding module 81 respectively to generate a single data stream. Therefore, if a EL MB is encoded by the second coding module 81 using the inter-layer prediction mode, the motion data related to said EL MB are not coded in the data stream (or only partly coded since quarter-pel motion refinement may possibly be coded) since they are derived from the motion data related to the BL MBs that are provided by the module 80. This allows to save some bits. On the other hand, if a EL MB is encoded using a classical mode (e.g. bidirectional mode) then motion data related to said EL MB are coded in the data stream.

The invention also concerns a decoding device 9 depicted on figure 20 for decoding high resolution images from a data stream generated for example with the coding device 8. The decoding device 9 comprises a first decoding module 91 for decoding a first part of the data stream, called base layer data stream, in order to derive low resolution images and motion data for said low resolution images. Preferentially the module 91 is adapted to decode a data stream compatible with MPEG4 AVC standard. The decoding device 9 comprises inheritance means 82 used to derive motion data for high

resolution images from the motion data of the low resolution images generated by the first decoding module 91. The inheritance means 82 are adapted to implement the steps 200 to 220 of the method according to the invention. The decoding device 9 comprises a second decoding module 92 for
5 decoding a second part of the data stream, called enhancement layer data stream. The second decoding module 92 uses the motion data derived by the inheritance means 82 in order to decode a second part of the data stream. The second decoding module 92 thus generates the high resolution images. Advantageously, the device 9 comprises also an extracting module 90 (e.g. a
10 demultiplexer) for extracting from the received data stream the base layer data stream and the enhancement layer data stream.

According to another embodiment the decoding device receives two data streams: a base layer data stream and an enhancement layer data stream. In this case the device 9 does not comprise an extracting module 90.

Claims

1. Method as part of a process of coding or decoding video signal for
5 generating for at least one macroblock of a high resolution image, called high
resolution macroblock, at least one motion predictor from motion data
associated to macroblock(s) of a low resolution image, called low resolution
macroblock(s), characterized in that it comprises the following steps:
- 10 - defining (190) a virtual structure made up of non overlapping
macroblocks, called virtual base layer macroblocks, said virtual structure
having same dimensions as said low resolution image;
 - identifying (200) for said high resolution macroblock corresponding virtual
base layer macroblocks in said virtual image on the basis of the location of
15 said high resolution macroblock within said high resolution image and on
the basis of the ratios of low resolution and high resolution images
dimensions;
 - deriving (210), for each identified virtual base layer macroblocks, motion
data from motion data associated to at least one macroblock of a
corresponding low resolution macroblocks pair on the basis of the
20 field/frame mode of said corresponding low resolution macroblocks pair
and on the basis of the field/frame mode of said high resolution
macroblock;
 - resampling (220) motion data associated to said corresponding virtual
base layer macroblocks so as to generate said at least one motion
25 predictor.
2. Method according to claim 1, wherein, if said high resolution macroblock is
in frame mode and if said corresponding low resolution macroblocks pair is in
field mode, the deriving step (210) comprises the steps of:
- 30 - upsampling vertically by a ratio of two (1500) said low resolution
macroblocks pair made up of a top macroblock and a bottom macroblock,
each upsampled macroblock being made up of an up part and a down part;
and

- merging (1510) motion data of the up part of the upsampled top macroblock with motion data of the up part of the upsampled bottom macroblock and merging (1510) motion data of the down part of the upsampled top macroblock with motion data of the down part of the upsampled bottom
5 macroblock.

3. Method according to claim 1 or claim 2, wherein, if said high resolution macroblock is in field mode and if said corresponding low resolution macroblocks pair is in frame mode, the deriving step (210) comprises a step
10 for downsampling vertically by a ratio of two motion data associated to said corresponding low resolution macroblocks pair.

4. Device (8) for coding a sequence of high resolution images made up of macroblocks, called high resolution macroblocks, and a sequence of low
15 resolution images, called low resolution macroblocks, said device comprising:

- first coding means (80) for coding said low resolution images, said first coding means generating motion data for macroblocks of said low resolution images and a base layer data stream;
- inheritance means (82) for deriving motion data for at least one
20 macroblock of a high resolution image, called high resolution macroblock from motion data of macroblocks a low resolution image, called low resolution macroblocks; and
- second coding means (81) for coding said high resolution images using said derived motion data, said second coding means generating an
25 enhancement layer data stream;

characterized in that the inheritance means (82) comprise:

- means for defining a virtual structure made up of non overlapping macroblocks, called virtual base layer macroblocks, said virtual structure having same dimensions as said low resolution image;
- 30 - means for identifying for said high resolution macroblock corresponding virtual base layer macroblocks in said virtual image on the basis of the location of said high resolution macroblock within said high resolution

image and on the basis of the ratios of low resolution and high resolution images dimensions;

- means for deriving, for each identified virtual base layer macroblocks, motion data from motion data associated to at least one macroblock of a corresponding low resolution macroblocks pair on the basis of the field/frame mode of said corresponding low resolution macroblocks pair and on the basis of the field/frame mode of said high resolution macroblock; and
- means for resampling motion data associated to said corresponding virtual base layer macroblocks so as to generate said at least one motion predictor.

5. Device according to claim 4, wherein said first coding means are an MPEG-4 AVC video encoder.

15

6. Device (9) for decoding at least a coded sequence of high resolution images, the coded images arising in the form of a data stream, each image being divided in non overlapping macroblocks, said device comprising:

- first decoding means (91) for decoding at least a first part of said data stream in order to generate low resolution images and motion data for macroblocks of said low resolution image;
- inheritance means (82) for deriving motion data for at least one macroblock of a high resolution image from motion data of macroblocks of a low resolution image; and
- second decoding means (92) for decoding at least a second part of said data stream using said derived motion data in order to generate high resolution images;

25

characterized in that the inheritance means (82) comprise:

- means for defining a virtual structure made up of non overlapping macroblocks, called virtual base layer macroblocks, said virtual structure having same dimensions as said low resolution image;
- means for identifying for said high resolution macroblock corresponding virtual base layer macroblocks in said virtual image on the basis of the

30

location of said high resolution macroblock within said high resolution image and on the basis of the ratios of low resolution and high resolution images dimensions;

- 5 - means for deriving, for each identified virtual base layer macroblocks, motion data from motion data associated to at least one macroblock of a corresponding low resolution macroblocks pair on the basis of the field/frame mode of said corresponding low resolution macroblocks pair and on the basis of the field/frame mode of said high resolution macroblock; and
- 10 - means for resampling motion data associated to said corresponding virtual base layer macroblocks so as to generate said at least one motion predictor.

- 7. Device according to claim 6, wherein said first decoding means are an
15 MPEG-4 AVC video decoder.

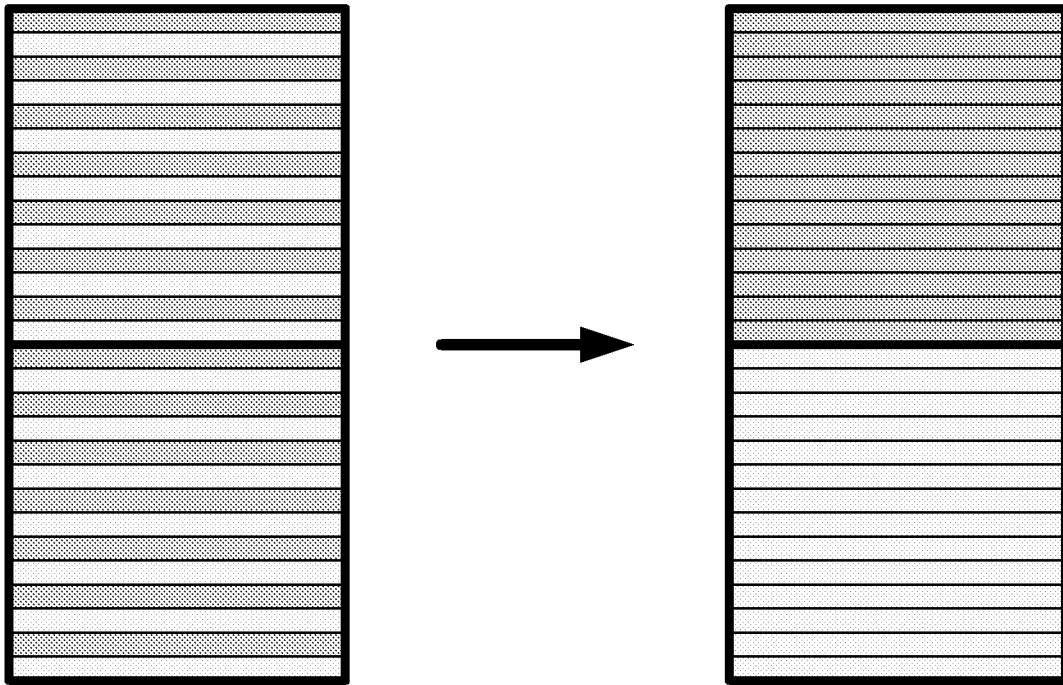


FIG.1

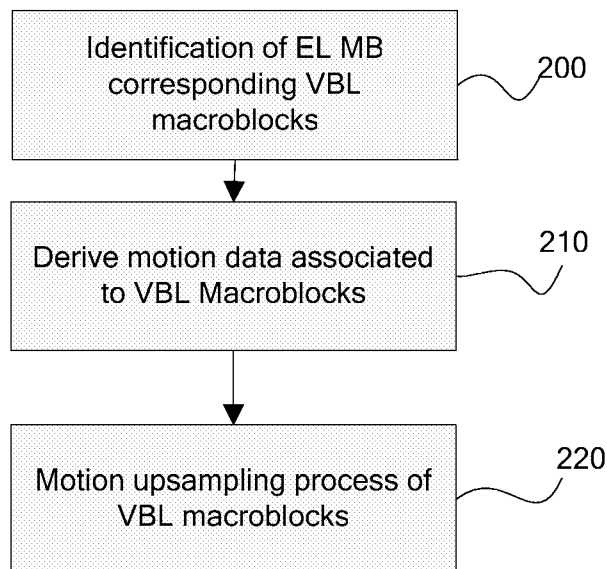


FIG.2

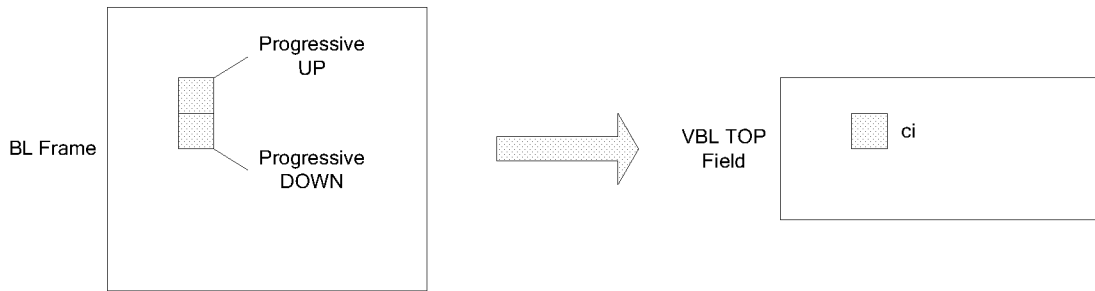


FIG.3

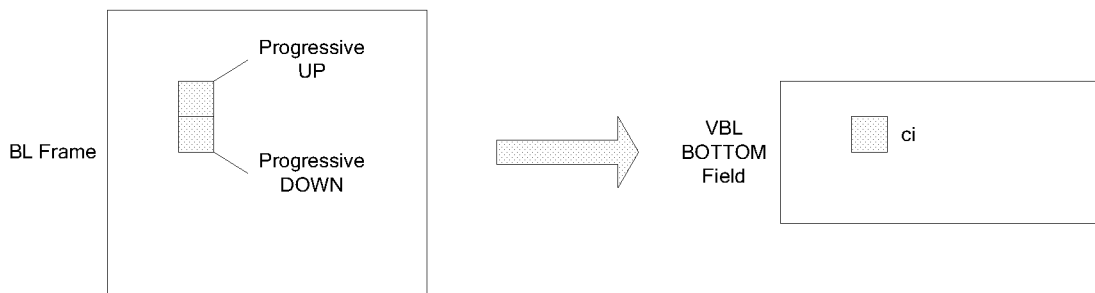


FIG.4

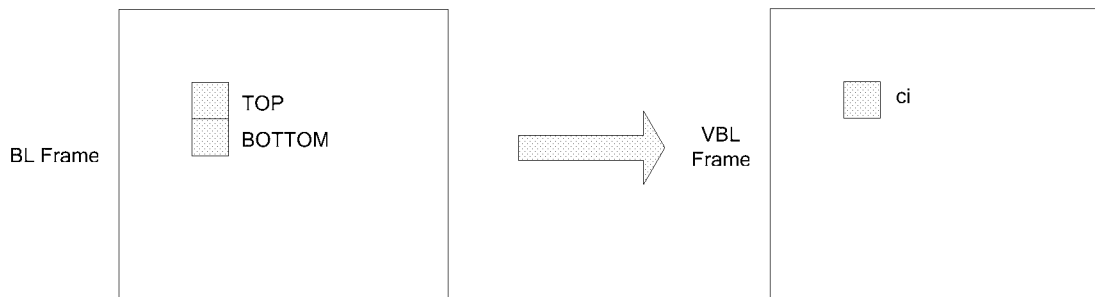


FIG.5

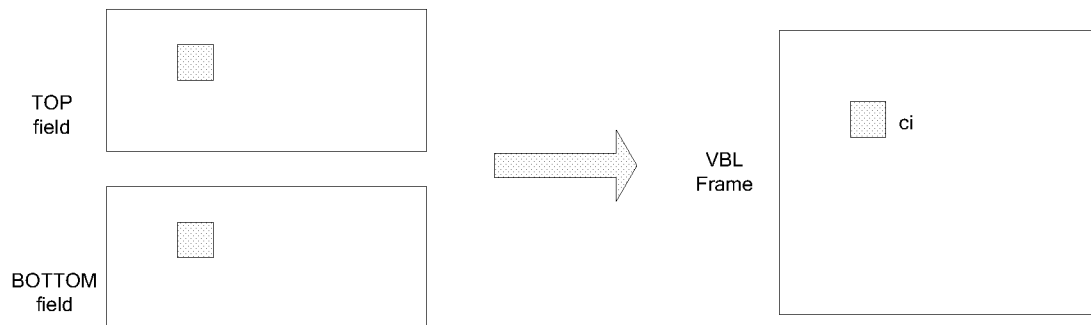


FIG.6

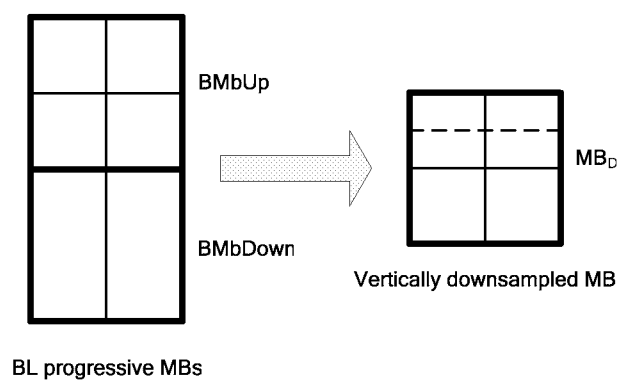


FIG.7

4/10

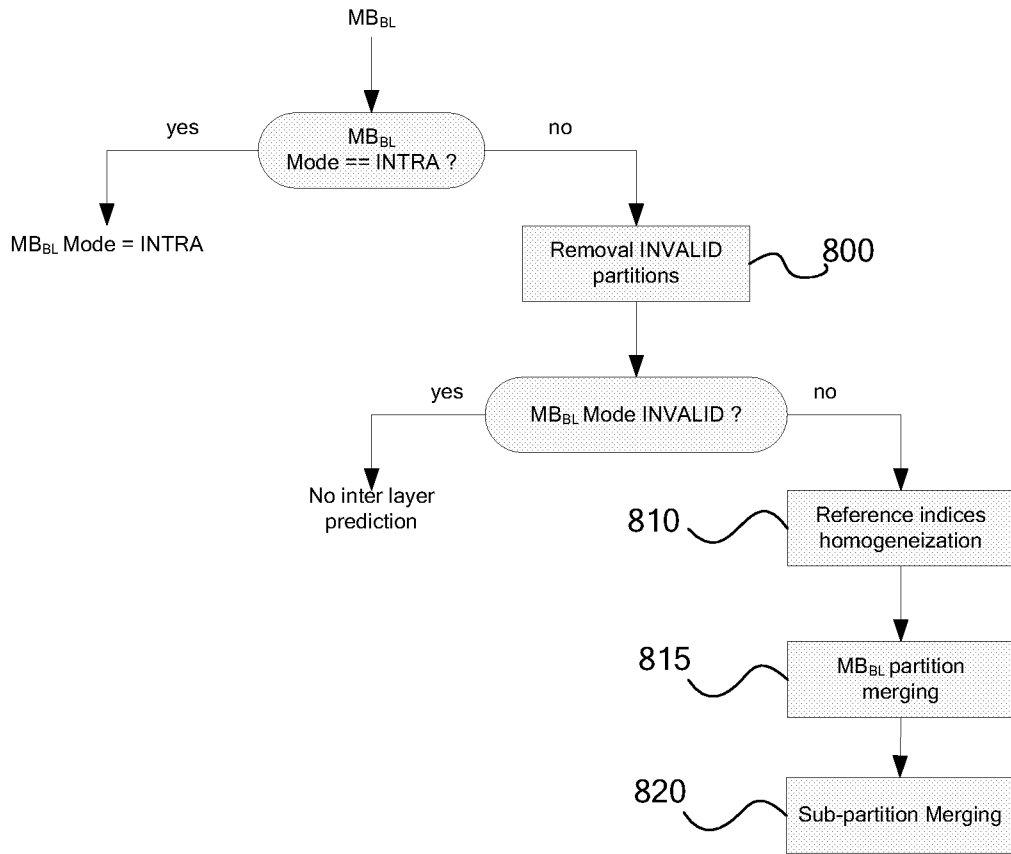


FIG.8

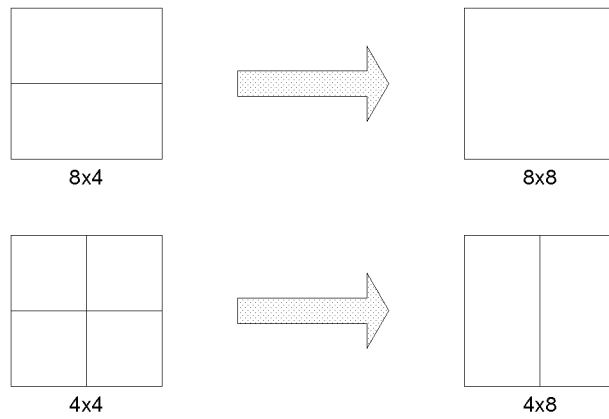


FIG.9

5/10

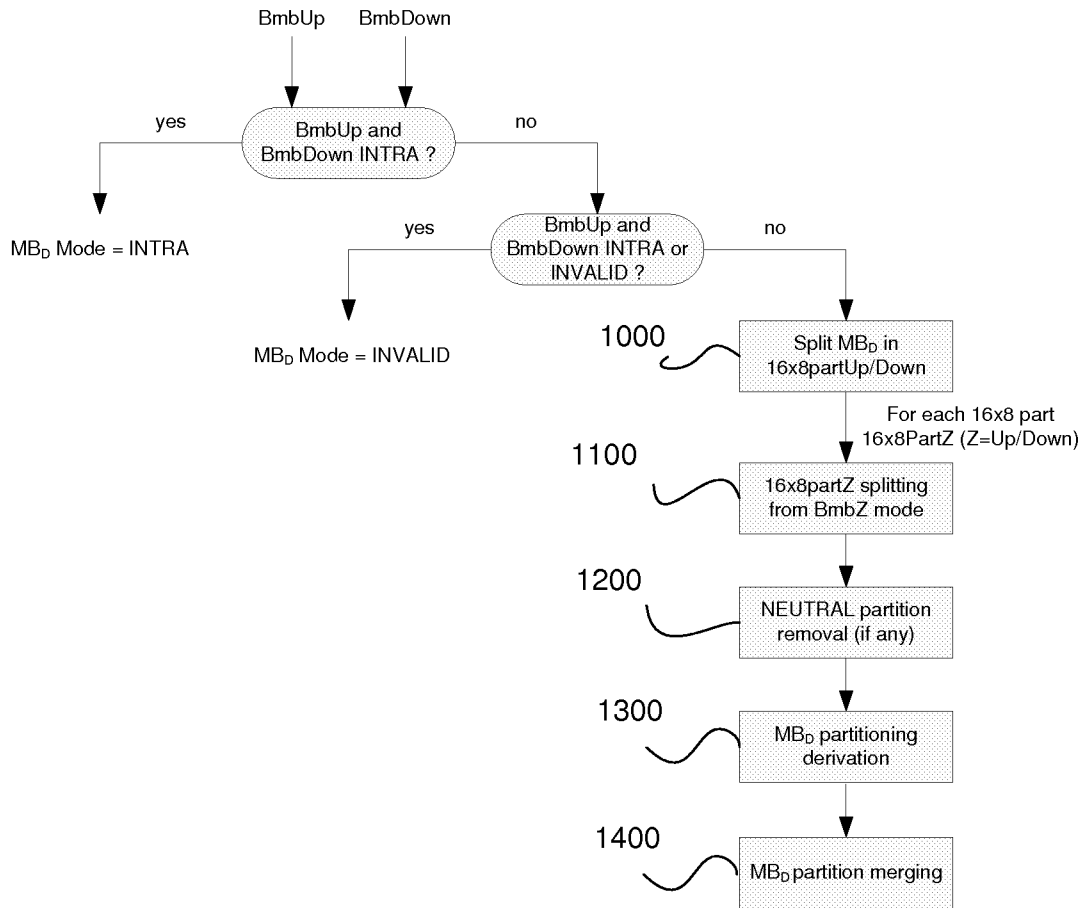


FIG.10

6/10

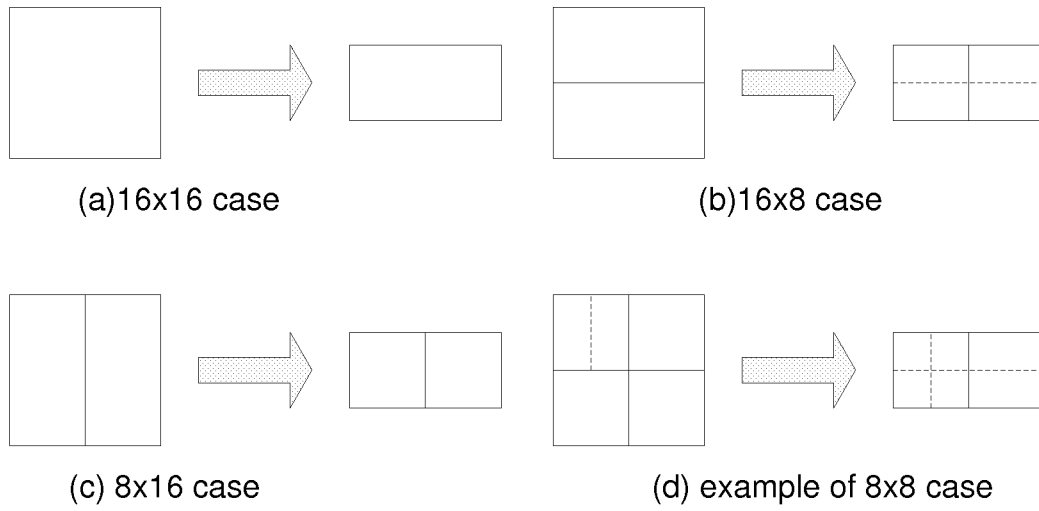


FIG.11

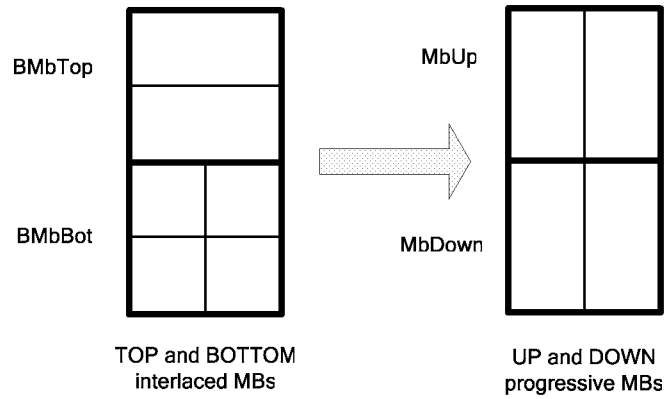


FIG.12

7/10

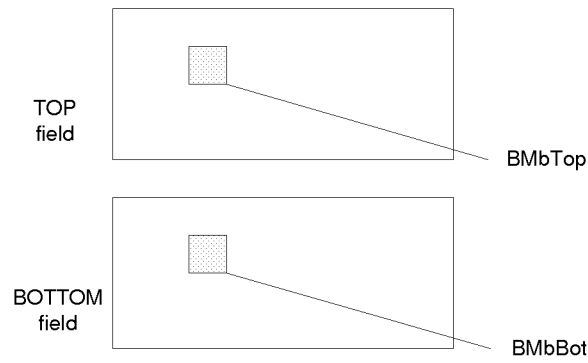


FIG.13

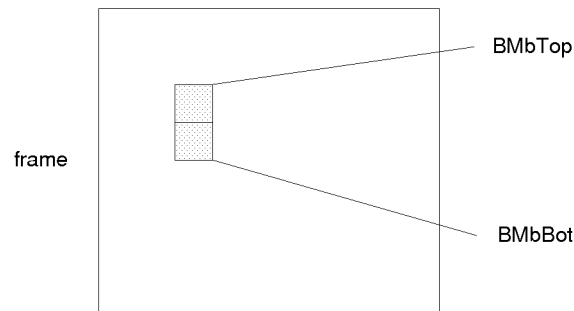


FIG.14

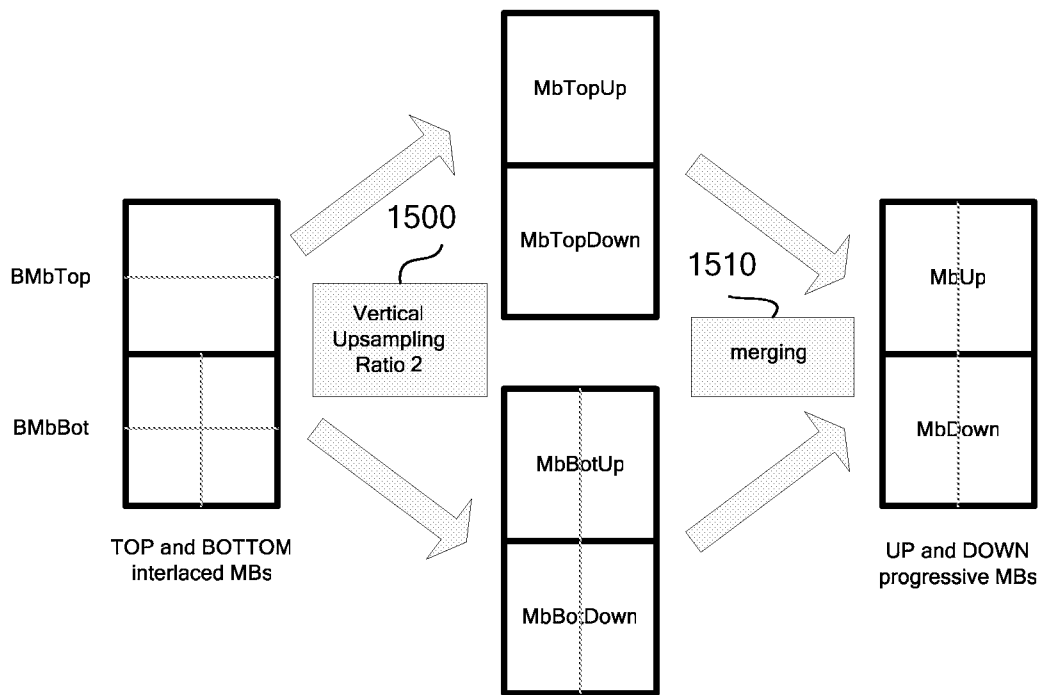


FIG.15

8/10

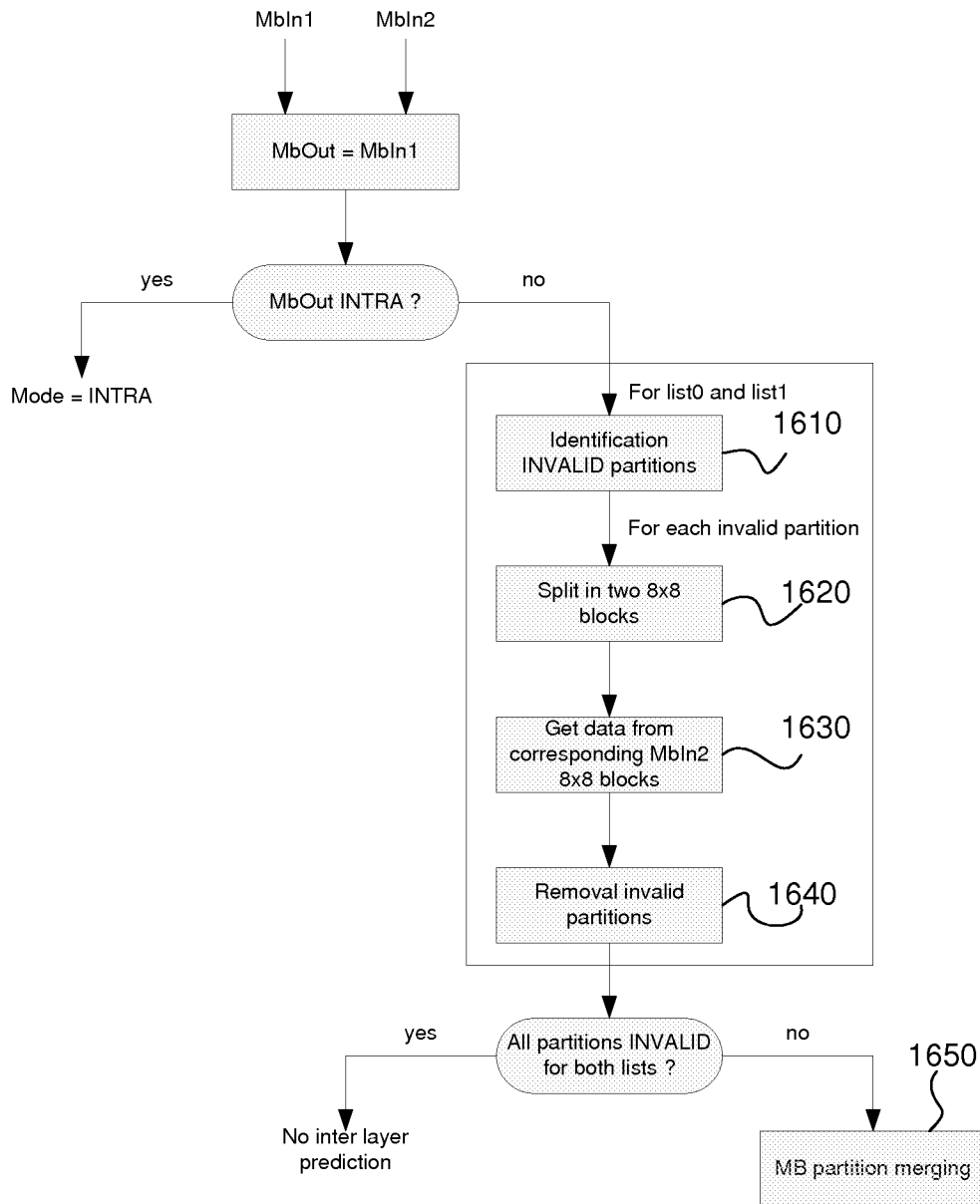


FIG.16

9/10

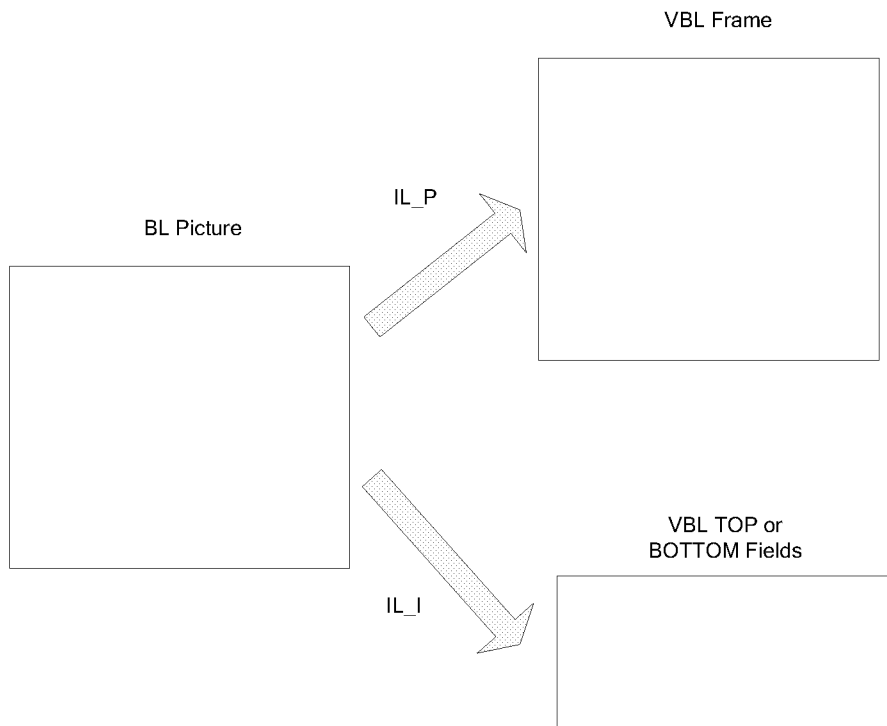


FIG.17

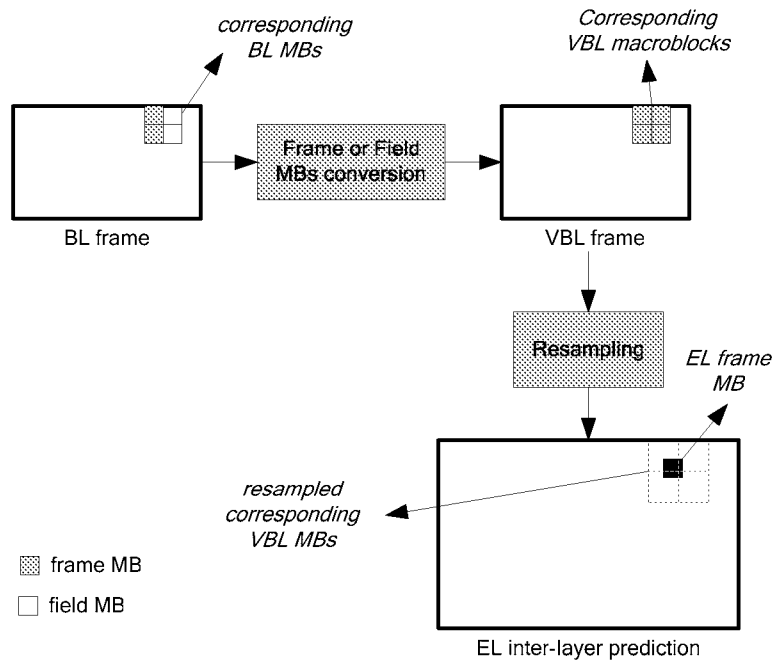


FIG.18

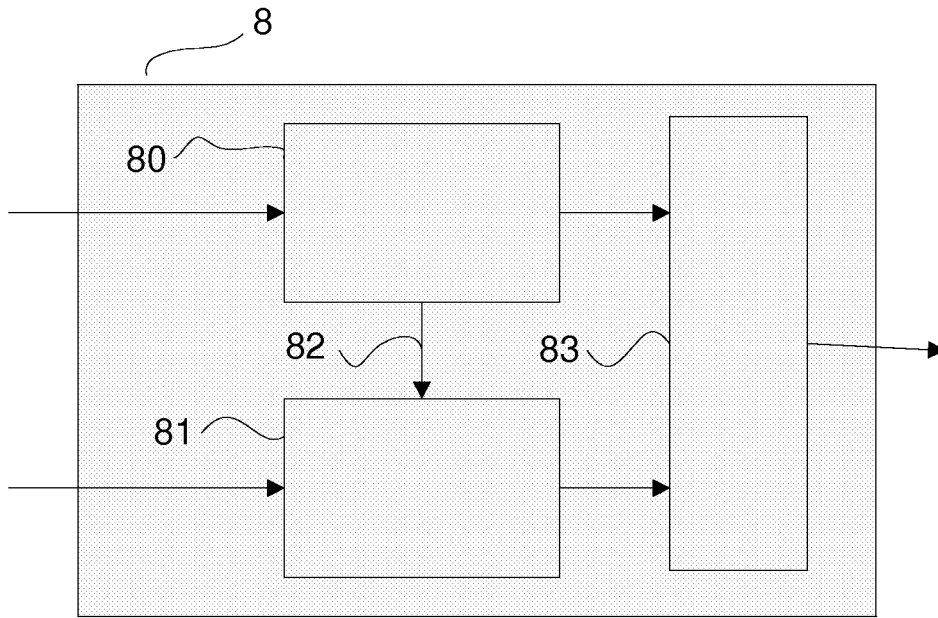


FIG.19

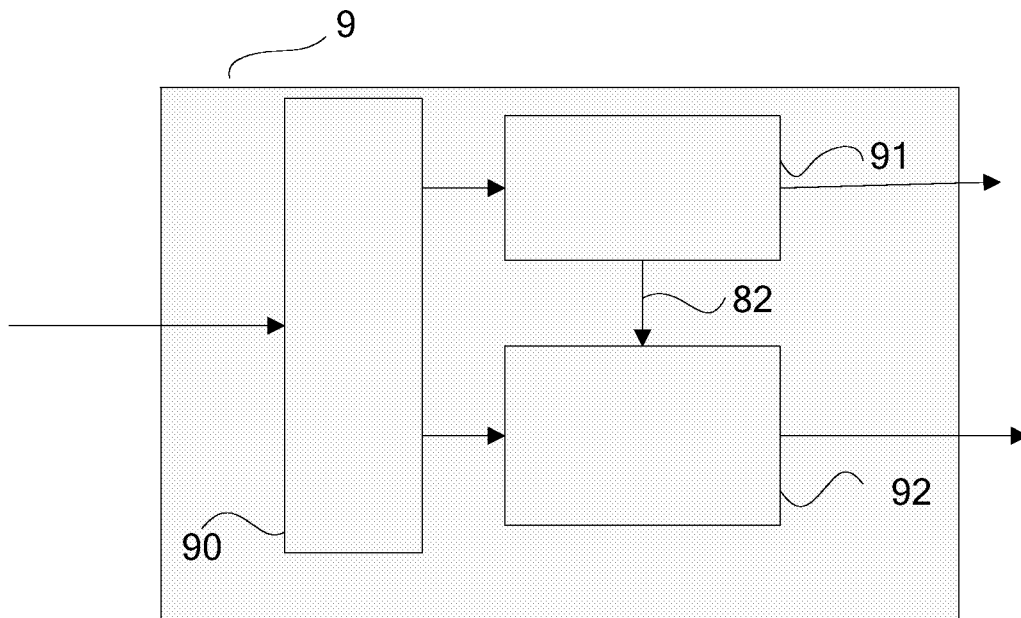


FIG.20

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2006/069852

A. CLASSIFICATION OF SUBJECT MATTER INV. H04N7/26		
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 practical, search terms used) EPO-Internal, INSPEC, 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	REICHEL J ET AL: "JOINT SCALABLE VIDEO MODEL JSVM-4" JOINT VIDEO TEAM (JVT) OF ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 AND ITU-T SG16 Q6), XX, XX, no. JVT Q202, 31 October 2005 (2005-10-31), pages 1-41, XP002372058	1,4-7
A	cited in the application paragraph [1.2.3.2] - paragraph [1.2.5.1.4]	2,3,5,7
----- -/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.		
<input type="checkbox"/> 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 document 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	
8 March 2007	19/03/2007	
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, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	KONTOPODIS, D	

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2006/069852

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WEE S J ET AL: "Field-to-frame transcoding with spatial and temporal downsampling" IMAGE PROCESSING, 1999. ICIP 99. PROCEEDINGS. 1999 INTERNATIONAL CONFERENCE ON KOBE, JAPAN 24-28 OCT. 1999, PISCATAWAY, NJ, USA, IEEE, US, vol. 4, 24 October 1999 (1999-10-24), pages 271-275, XP010368707 ISBN: 0-7803-5467-2 abstract paragraph [5.2.3] -----	1-7