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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 June 2007 (14.06.2007)

PCT

(10) International Publication Number
WO 2007/065796 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/EP2006/068786

(22) International Filing Date:
23 November 2006 (23.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0553713 5 December 2005 (05.12.2005) FR

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

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **VIERON, Jérome** [FR/FR]; 5 Allie Jean Perrin, F-35137 Bedee (FR). **BOTTREAU, Vincent** [FR/FR]; La Riaudais Broons Sur Vilaine, F-35220 Chateaubourg (FR). **LOPEZ, Patrick** [FR/FR]; 4 Allée Du Clos De La Merceraye, F-35450 Livré Sur Changeon (FR).

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

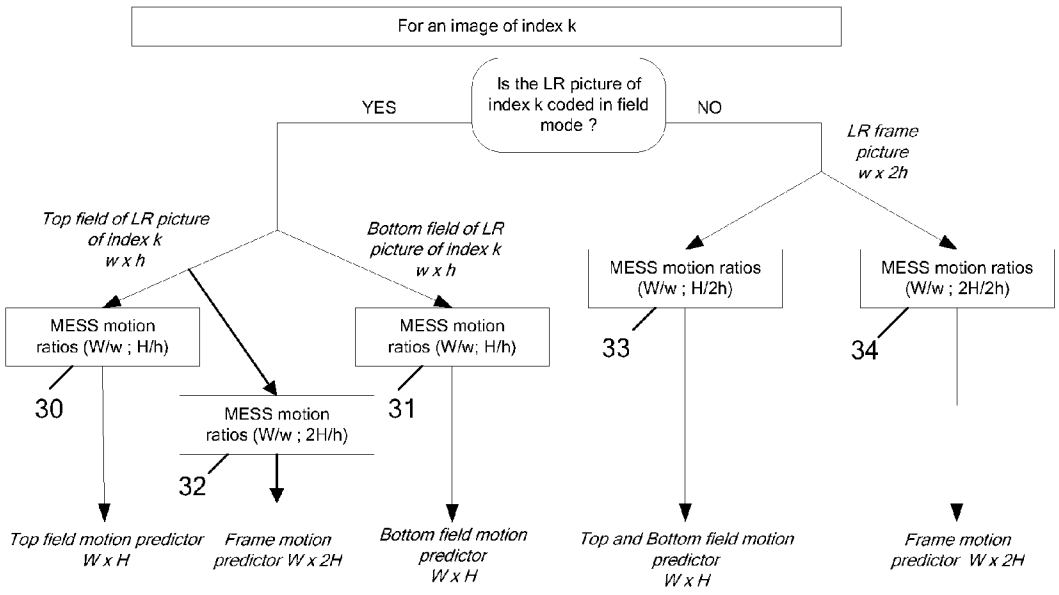
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, 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.

(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).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) **Title:** METHOD OF PREDICTING MOTION AND TEXTURE DATA



(57) **Abstract:** The invention relates to a method for generating for at least one block of pixels of a picture of a sequence of interlaced pictures at least one motion predictor and at least one texture predictor from motion data, respectively 10 texture data, associated with the pictures of a sequence of low resolution interlaced pictures.

WO 2007/065796 A2

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.

METHOD OF PREDICTING MOTION AND TEXTURE DATA

1. Background of the invention

The invention relates to a method for generating, for pictures of a high resolution interlaced sequence, at least one motion predictor and, where appropriate, at least one texture predictor from motion data and, where appropriate, texture data associated with pictures of a low resolution interlaced sequence.

2. State of the art

Hierarchical encoding methods with spatial scalability are known. Scalability represents the ability to stagger information to make it decodable at multiple resolution and/or quality levels. More specifically, a data stream generated by this type of encoding method is divided into several layers, in particular a basic layer and one or more enhancement layers. These methods are used in particular to adapt a single data stream to variable transport conditions (bandwidth, error ratios, etc.), and to the expectations of the customers and the varying capabilities of their receivers (CPU, specifications of the display device, etc.). In the particular case of spatial scalability, the part of the data stream corresponding to low resolution pictures of the sequence can be decoded independently of the part of the data stream corresponding to high resolution pictures. On the other hand, the part of the data stream corresponding to the high resolution pictures of the sequence can be decoded only from the part of the data stream corresponding to the low resolution pictures.

Hierarchical encoding with spatial scalability makes it possible to encode a first data part called basic layer, relative to the low resolution pictures and, from this basic layer, a second data part called enhancement layer, relative to the high resolution pictures. Normally, each macroblock of the high resolution picture is temporally predicted according to a conventional prediction mode (for example, bidirectional prediction mode, direct prediction method, early prediction mode, etc.) or indeed is predicted according to an inter-layer prediction method. In this latter case, motion data (for example, a partitioning of the macroblock into blocks, possibly motion vectors and reference picture indices) and, where appropriate, texture data associated with a block of pixels of the high resolution picture is deduced or inherited from the motion data, respectively texture data, associated with blocks of pixels of a low resolution picture. However, the known methods do not allow

such predictors to be generated in the case where the low resolution sequence and the high resolution sequence are interlaced.

3. Summary of the invention

The object of the invention is to overcome at least one of the drawbacks of the prior art.

The invention relates to a method for generating for at least one block of pixels of a picture of a sequence of high resolution interlaced pictures, called high resolution sequence, at least one motion predictor from motion data associated with the pictures of a sequence of low resolution interlaced pictures, called low resolution sequence, of the same temporal frequency as the high resolution sequence. Each interlaced picture comprises a top field interlaced with a bottom field and is able to be coded in frame mode or in field mode. Each field of an interlaced picture has associated with it a temporal reference. The method makes it possible in particular to generate the at least one motion predictor for the at least one block of pixels of the high resolution picture:

- from the motion data associated with at least one block of pixels of the top field of a low resolution picture of the same temporal reference as the top field of the high resolution picture if the low resolution picture is coded in field mode; and/or
- from the motion data associated with at least one block of pixels of the bottom field of the low resolution picture of the same temporal reference as the bottom field of the high resolution picture if the low resolution picture is coded in field mode; and/or
- from the motion data associated with at least one block of pixels of a low resolution picture comprising a field with the same temporal reference as the top field or bottom field of the high resolution picture if the low resolution picture is coded in frame mode.

According to a first embodiment, the at least one motion predictor is generated for the at least one block of pixels of the high resolution picture by sub-sampling the motion data associated with the at least one block of pixels of the top field of a low resolution picture of the same temporal reference as

the top field of the high resolution picture with a horizontal inter-layer ratio in the horizontal direction of the picture and a vertical inter-layer ratio in the vertical direction of the picture.

According to another embodiment, the at least one motion predictor is generated for the at least one block of pixels of the high resolution picture by sub-sampling the motion data associated with the at least one block of pixels of the bottom field of the low resolution picture of the same temporal reference as the bottom field of the high resolution picture with a horizontal inter-layer ratio in the horizontal direction of the picture and a vertical inter-layer ratio in the vertical direction of the picture.

According to another embodiment, the at least one motion predictor is generated for the at least one block of pixels of the high resolution picture by sub-sampling the motion data associated with the at least one block of pixels of the low resolution picture comprising a field with the same temporal reference as the top field or bottom field of the high resolution picture with a horizontal inter-layer ratio in the horizontal direction of the picture and a vertical inter-layer ratio in the vertical direction of the picture.

The method also makes it possible to generate, for the block of pixels of the high resolution picture, at least one texture predictor:

- from the texture data associated with at least one block of pixels of the top field of a low resolution picture of the same temporal reference as the top field of the high resolution picture; and/or
- from the texture data associated with at least one block of pixels of the bottom field of a low resolution picture of the same temporal reference as the bottom field of the high resolution picture; and/or
- from the texture data associated with at least one block of pixels of each of the top and bottom fields of the low resolution picture.

Advantageously, the motion data associated with the low resolution pictures comprises motion vectors.

Preferably, the motion vectors associated with a low resolution picture coded in frame mode or with each of the top and bottom fields of a low resolution picture coded in field mode have the same parity.

The method is advantageously used by a method of encoding high resolution pictures from low resolution pictures and by a method of decoding high resolution pictures from low resolution pictures.

Preferably, the low resolution pictures are encoded according to the MPEG-4 AVC standard.

4. List of figures

The invention will be better understood and illustrated by means of exemplary embodiments and advantageous implementations, by no means limiting, given with reference to the appended figures in which:

- figure 1 represents a sequence of low resolution pictures and a sequence of high resolution pictures that are interlaced and have the same temporal frequency;
- figure 2 illustrates the method of generating texture predictors according to the invention in the case where the sequence of low resolution pictures and the sequence of high resolution pictures are interlaced;
- figure 3 illustrates the method of generating motion predictors according to the invention in the case where the sequence of low resolution pictures and the sequence of high resolution pictures are interlaced;
- figure 4 illustrates the sub-sampling by a factor of 2 in the horizontal direction of the picture of two macroblocks MB1 and MB2 of a low resolution picture and the resulting partitioning for the corresponding predictor macroblock MB_pred.

5. Detailed description of the invention

The invention relates to an inter-layer prediction method which consists in generating motion predictors and, where appropriate, texture predictors for pictures of an ordered sequence of high resolution interlaced pictures, called high resolution sequence, from pictures of an ordered sequence of low resolution interlaced pictures, called low resolution sequence. The sequences are divided into groups of pictures (GOP). Each low and high resolution

picture comprises a top field interlaced with a bottom field. In figure 1, an interlaced picture of index k is made up of a top field referenced kT and a bottom field referenced kB. A temporal reference is associated with each field of the high and low resolution pictures. A field of a high resolution picture and a field of a low resolution picture having the same temporal reference coincide vertically. The low resolution pictures, also referenced LR pictures, have a width w (w representing a number of pixels or columns) and a height of 2h (2h representing a number of pixels or lines and means 2 multiplied by h). Each field of a low resolution picture has a width w and height h. The high resolution pictures, also referenced HR pictures, have a width W (W representing a number of pixels or columns) and a height of 2H (2H representing a number of pixels or lines and meaning 2 multiplied by H). Each field of a high resolution picture has a width W and a height H. In the embodiment described, the interlaced pictures can be encoded either in field picture mode, i.e. each field is encoded as a separate picture, or even in frame picture mode, i.e. the two fields are encoded together. The lines of a picture are numbered from 0 and therefore the first line is an even line and the second line (numbered 1) is an odd line.

The invention therefore consists in generating, for pictures of the high resolution sequence or for at least one block of pixels of the latter, at least one motion predictor and, where appropriate, at least one texture predictor. A texture predictor associated with a high resolution picture or with at least one block of pixels of a high resolution picture is a picture or a prediction block which associates with each of its pixels texture data (for example, a luminance value and, where appropriate, chrominance values), which is generated from texture data associated with at least one picture (or field) or at least one block of pixels of a low resolution picture (or at least one block of pixels of a field) according to a method of sub-sampling the texture such as the ESS method applied to the texture (ESS standing for Extended Spatial Scalability) which is described in sections S.8.3.6.4 and S.8.5.14.2 of document ISO/IEC MPEG & ITU-T VCEG, entitled "Joint Scalable Video Model JSVM3 Annex-S", referenced JVT-P202, J.Reichel, H.Schwarz, M.Wien. This document is referenced JSVM3 below. A motion predictor associated with a high resolution picture or with at least one block of pixels of a high resolution picture is

defined as a prediction picture or a prediction block with which is associated motion data (for example, a type of partitioning, possibly reference picture indices making it possible to identify the reference pictures to which the motion vectors point). The motion predictor is generated from motion data associated with at least one picture (or field) or at least one block of pixels of a low resolution picture (or at least one block of pixels of a field) according to a motion sub-sampling method such as the ESS method applied to the motion which is described in section S.8.4.1.6.3 of JSVM3, or such as the modified ESS method, described below, derived from the ESS method applied to the motion. The modified ESS method, referenced MESS in figure 3, makes it possible in particular to process high and/or low resolution interlaced sequences. More specifically, it makes it possible to deal with the case where the height or the width of the high resolution picture is less than that of the low resolution picture. Furthermore, it makes it possible advantageously to avoid having the motion predictors include invalid motion vectors, i.e. vectors that point to unavailable reference pictures, when the prediction method according to the invention is used by a hierarchical encoding or decoding method.

According to the modified ESS method, an intermediate motion predictor is generated by sub-sampling by 2 the motion data associated with the low resolution picture, more particularly, the motion data associated with each of the macroblocks of the low resolution picture, in the vertical direction of the picture, in the horizontal direction of the picture or in both directions. The method of sub-sampling by 2 is repeated in the vertical direction of the picture as long as the height of said intermediate predictor is greater than the height of the high resolution picture and it is repeated in the horizontal direction of the picture as long as the width of said intermediate predictor is greater than the width of the high resolution picture. The sub-sampling consists in particular in dividing by two the coordinates of the motion vectors associated with the blocks of pixels. For example, with reference to figure 4, based on two macroblocks MB1 or MB2 of the low resolution picture possibly divided into blocks of pixels, a macroblock MB of the intermediate motion predictor is generated. The size of the blocks of pixels in a macroblock is indicated above said macroblock. For example, in the second line of figure 4, the macroblock MB1 is not divided, the macroblock MB2 is divided into two blocks measuring

8 by 16 pixels (denoted 8x16) and the macroblock MB generated from these two macroblocks is divided into four 8x8 blocks, two of which are divided into 4x8 blocks. The indices of reference pictures are made uniform between the blocks of 8 by 8 pixel size within a macroblock MB, and isolated intra-type blocks within a macroblock MB are deleted in the same way as in the ESS inter-layer prediction method applied to the motion and described in JSVM3.

The motion predictor associated with the high resolution picture is generated from the last intermediate motion predictor generated in this way, by applying the ESS method with an inter-layer ratio equal to $\frac{W}{w_i}$ in the horizontal direction of the picture and $\frac{2H}{2h_i}$ in the vertical direction of the picture, where w_i and $2h_i$ are respectively the width and the height of the last intermediate motion predictor generated. Furthermore, for each prediction macroblock, the motion vector inheritance method is modified so as not to generate invalid motion vectors, i.e. vectors that point to fields or frame pictures that are not available in the temporal breakdown process. In this case, if all the motion vectors associated with a prediction macroblock MB_pred are invalid then the inter-layer motion prediction is not authorized for this macroblock. Otherwise (i.e. if at least one of the vectors is valid), the ESS prediction method applied to the motion is used.

The method according to the invention, illustrated by figures 1 to 3, is described for a picture but can be applied to a part of a picture and in particular to a macroblock. It makes it possible for example to handle the case of a low resolution sequence interlaced in the SD format, i.e. of dimension 720 by 288 pixels, 60 Hz and of a high resolution sequence interlaced in the 1080i format, i.e. of dimension 1920 by 540 pixels, 60 Hz.

Texture predictors associated with a high resolution picture of index k in figure 1 are generated in the following manner as illustrated by figure 2:

- A texture predictor of dimension W by H is generated 20 for the top field of the high resolution picture on the basis of the texture data of the top field of the low resolution picture of index k by applying the ESS

method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{H}{h}$ in the vertical direction of the picture;

- A texture predictor of dimension W by H is generated 21 for the bottom field of the high resolution picture on the basis of the texture data of the bottom field of the low resolution picture of index k by applying the ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{H}{h}$ in the vertical direction of the picture.
- A frame texture predictor of dimension W by 2H is generated 22 by interlacing the texture predictors associated with the top and bottom fields.

According to an embodiment illustrated in figure 2 by the dotted lines, the frame texture predictor is generated 23 on the basis of texture data of the low resolution frame picture, ie. the top and bottom fields of the low resolution picture interlaced, by applying the ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{2H}{2h}$ in the vertical direction of the picture.

If the low resolution picture of index k is coded in field mode, motion predictors associated with a high resolution picture of index k in figure 1 are generated in the following manner as illustrated by figure 3:

- a motion predictor of dimension W by H is generated 30 for the top field of the high resolution picture on the basis of the motion data of the top field of the low resolution picture of index k for the top field of the high resolution picture on the basis of the low resolution picture of index k by applying the modified ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{H}{h}$ in the vertical direction of the picture;

- A motion predictor of dimension W by H is generated 31 for the bottom field of the high resolution picture on the basis of the motion data of the bottom field of the low resolution picture of index k by applying the modified ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{H}{h}$ in the vertical direction of the picture; and

- A frame motion predictor of dimension W by 2H is generated 32 on the basis of the motion data of the bottom field of the low resolution picture of index k by applying the modified ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{2H}{h}$ in the vertical direction of the picture.

In the other cases, i.e. if the low resolution picture of index k is coded in frame mode, motion predictors associated with a high resolution picture of index k in figure 1 are generated in the following manner as illustrated by figure 3:

- A motion predictor of dimension W by H is generated 33 for the top field of the high resolution picture on the basis of the motion data of the frame low resolution picture of index k by applying the modified ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{2H}{h}$ in the vertical direction of the picture;
- A motion predictor of dimension W by H is generated 33 for the bottom field of the high resolution picture on the basis of the motion data of the frame low resolution picture of index k by applying the modified ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{2H}{h}$ in the vertical direction of the picture; and

- A frame motion predictor of dimension W by 2H is generated 34 on the basis of the motion data of the frame low resolution picture of index k by applying the modified ESS method with an inter-layer ratio of $\frac{W}{w}$ in the horizontal direction of the picture and $\frac{2H}{2h}$ in the vertical direction of the picture.

In this case, the motion predictors associated with the bottom field and top field of the high resolution picture are identical.

If the method is used by a coding method all motion predictors (respectively texture predictors) may be generated in order to select the most appropriate one according to a given criteria, e.g. a rate distorsion criteria. If said method is used by a decoding method then a single motion predictor (respectively a single texture predictor) is generated, the type of predictor being specified in the bitstream.

Of course, the invention is not limited to the abovementioned exemplary embodiments. In particular, those skilled in the art can apply any variant to the embodiments described and combine them to benefit from their different advantages. For example, the method according to the invention can be applied to a part of the high resolution picture. In practice, it is possible to generate motion and/or texture predictors for blocks of pixels (for example, macroblocks measuring 16 by 16 pixels) of the high resolution picture from motion and/or texture data associated with blocks of pixels of the low resolution pictures. Similarly, the invention has been described in the case where the top field of an interlaced picture is displayed first ("top field first" case) and can be extended directly to the case where the bottom field is displayed first ("bottom field first" case) by reversing the top and bottom fields. Moreover, the invention can also be extended to the case of several high resolution sequences (i.e. several enhancement layers). Furthermore, the invention is advantageously used by a method of encoding or decoding a sequence of pictures or video. Preferably, the sequence of low resolution pictures is encoded according to the MPEG4 AVC encoding standard defined

in document ISO/IEC 14496-10 ("Information technology -- Coding of audio-visual objects -- Part 10: Advanced Video Coding").

Claims

1. Method for generating for at least one block of pixels of a picture of a sequence of high resolution interlaced pictures, called high resolution sequence, at least one motion predictor from motion data associated with the pictures of a sequence of low resolution interlaced pictures, called low resolution sequence, of the same temporal frequency as said high resolution sequence, each interlaced picture comprising a top field interlaced with a bottom field and able to be coded in frame mode or in field mode, each field of an interlaced picture having associated with it a temporal reference, characterized in that, said at least one motion predictor is generated for said at least one block of pixels of said high resolution picture:

- from the motion data associated with at least one block of pixels of the top field of a low resolution picture of the same temporal reference as the top field of said high resolution picture if said low resolution picture is coded in field mode (30, 32); and/or
- from the motion data associated with at least one block of pixels of the bottom field of said low resolution picture of the same temporal reference as the bottom field of said high resolution picture if said low resolution picture is coded in field mode (31); and/or
- from the motion data associated with at least one block of pixels of a low resolution picture comprising a field with the same temporal reference as the top field or bottom field of the high resolution picture if said low resolution picture is coded in frame mode (33, 34).

2. Method according to Claim 1, wherein said at least one motion predictor is generated for said at least one block of pixels of said high resolution picture by subsampling (30, 32) said motion data associated with said at least one block of pixels of said top field of a low resolution picture of the same temporal reference as the top field of said high resolution picture with a horizontal inter-layer ratio in the horizontal direction of the picture and a vertical inter-layer ratio in the vertical direction of the picture.

3. Method according to Claim 1, wherein said at least one motion predictor is generated for said at least one block of pixels of said high resolution picture by subsampling (31) said motion data associated with said at least one block of pixels of said bottom field of said low resolution picture of the same temporal reference as the bottom field of said high resolution picture with a horizontal inter-layer ratio in the horizontal direction of the picture and a vertical inter-layer ratio in the vertical direction of the picture.

4. Method according to Claim 1, wherein said at least one motion predictor is generated for said at least one block of pixels of said high resolution picture by subsampling (33, 34) said motion data associated with said at least one block of pixels of said low resolution picture comprising a field with the same temporal reference as the top field or bottom field of the high resolution picture with a horizontal inter-layer ratio in the horizontal direction of the picture and a vertical inter-layer ratio in the vertical direction of the picture.

5. Method according to one of Claim 1, wherein, for said block of pixels of said high resolution picture, at least one texture predictor is generated:
- from the texture data associated with at least one block of pixels of the top field of a low resolution picture of the same temporal reference as the top field of said high resolution picture (20); and/or
- from the texture data associated with at least one block of pixels of the bottom field of a low resolution picture of the same temporal reference as the bottom field of said high resolution picture (21); and/or
- from the texture data associated with at least one block of pixels of each of said top and bottom fields of said low resolution picture (22).

6. Method according to one of Claims 1 to 5, wherein the motion data associated with the low resolution pictures comprises motion vectors.

7. Method according to Claim 6, wherein the motion vectors associated with a low resolution picture coded in frame mode or with each of the top and bottom fields of a low resolution picture coded in field mode have the same parity.

8. Method according to one of Claims 1 to 7 wherein said method is used by a method of encoding high resolution pictures from low resolution pictures.

9. Method according to Claim 8 wherein the low resolution pictures are encoded according to the MPEG-4 AVC standard.

10. Method according to Claims 1 to 7 wherein said method is used by a method of decoding high resolution pictures from low resolution pictures.

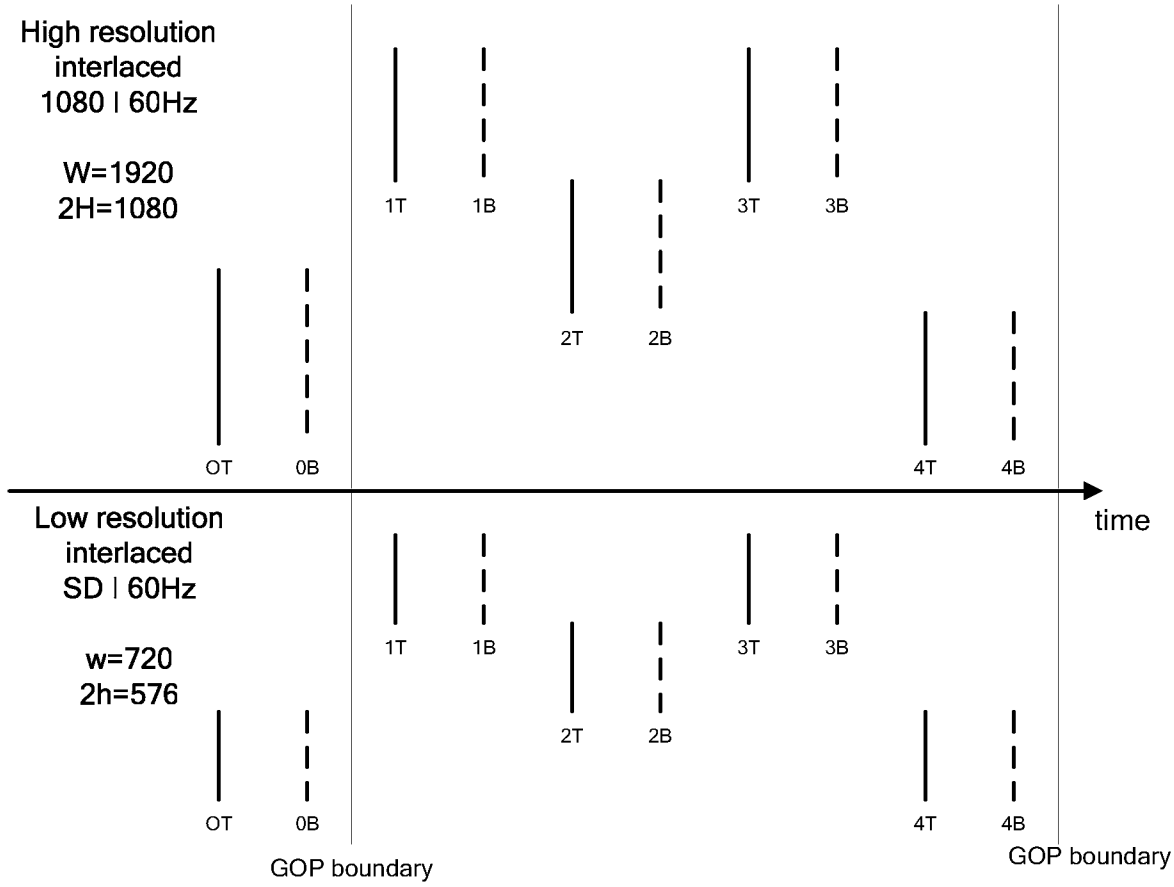

High resolution
interlaced
1080 I 60Hz
W=1920
2H=1080
1T
1B
3T
3B
2T
2B
OT
0B
4T
4B
time
Low resolution
interlaced
SD I 60Hz
w=720
2h=576
1T
1B
3T
3B
2T
2B
OT
0B
4T
4B
GOP boundary
GOP boundary

FIG.1

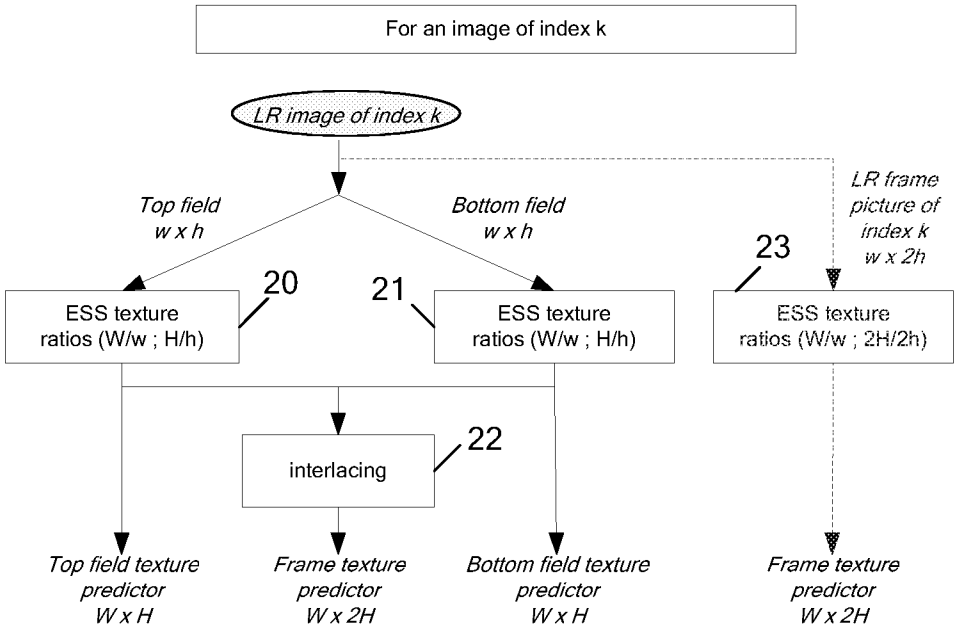

For an image of index k
LR image of index k
Top field
w x h
Bottom field
w x h
LR frame
picture of
index k
w x 2h
20
21
23
ESS texture
ratios (W/w ; H/h)
ESS texture
ratios (W/w ; H/h)
ESS texture
ratios (W/w ; 2H/2h)
22
interlacing
Top field texture
predictor
W x H
Frame texture
predictor
W x 2H
Bottom field texture
predictor
W x H
Frame texture
predictor
W x 2H

FIG.2

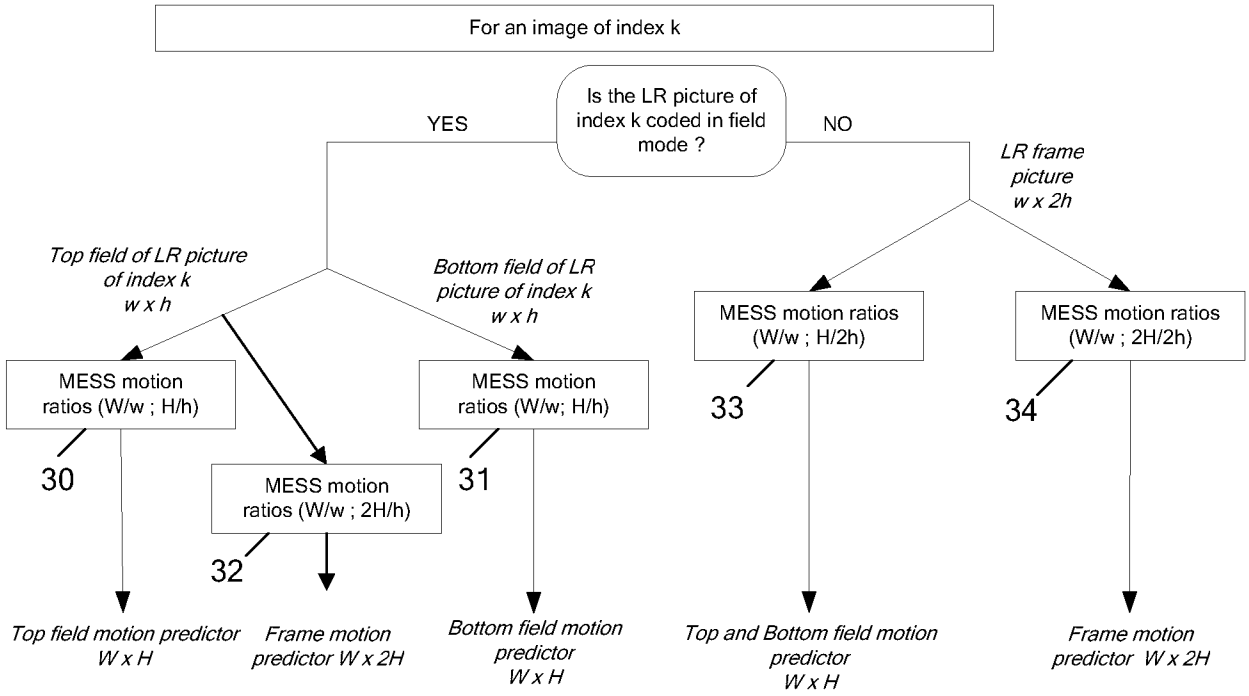

For an image of index k
Is the LR picture of index k coded in field mode ?
YES
NO
LR frame picture w x 2h
Top field of LR picture of index k w x h
Bottom field of LR picture of index k w x h
MESS motion ratios (W/w ; H/h)
MESS motion ratios (W/w; H/h)
MESS motion ratios (W/w ; H/2h)
MESS motion ratios (W/w ; 2H/2h)
30
31
32
33
34
MESS motion ratios (W/w ; 2H/h)
Top field motion predictor W x H
Frame motion predictor W x 2H
Bottom field motion predictor W x H
Top and Bottom field motion predictor W x H
Frame motion predictor W x 2H

FIG.3

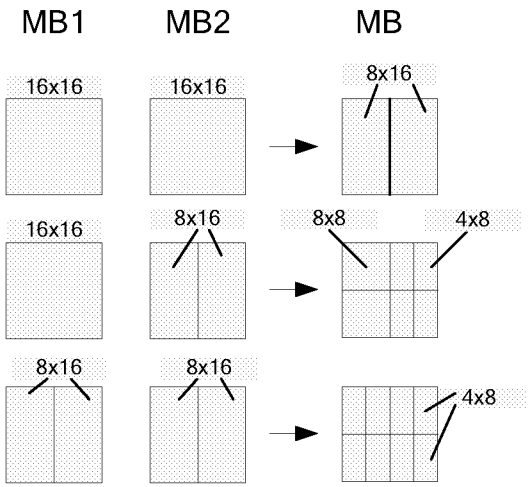

MB1
MB2
MB
16x16
16x16
8x16
16x16
8x16
8x8
4x8
8x16
8x16
4x8

FIG.4