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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 October 2006 (12.10.2006)

PCT

(10) International Publication Number
WO 2006/106039 A1

(51) International Patent Classification:
H04N 7/26 (2006.01) *H04N 7/24* (2006.01)

(21) International Application Number:
PCT/EP2006/060873

(22) International Filing Date: 20 March 2006 (20.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
05300253.1 6 April 2005 (06.04.2005) 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)*: **CHEN, Ying** [CN/CN]; Bld. 3, 103, Chenghua Yuan, Xue Qing Road, Hai Dian District, Beijing 100083 (CN). **ZHAI, Jiefu** [CN/CN]; Bld. 3, 103, Chenghua Yuan, Xue Qing Road, Hai Dian District, Beijing 100083 (CN). **YIN, Peng** [CN/US]; 65 Warwick Road, West Windsor, New Jersey 08550 (US). **PANDIT, Purvin, Bibhas** [IN/US]; 23 Pear Tree Lane, Franklin Park, NJ 08823 (US).

(74) Agent: **RITTNER, Karsten**; Deutsche Thomson-Brandt GmbH, European Patent Operations, Karl-wiechert-allee 74, 30625 Hannover (DE).

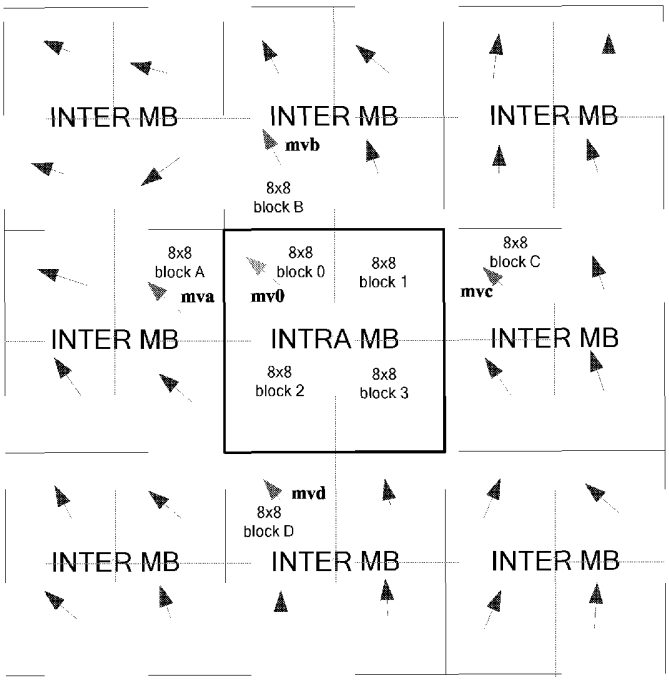
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC,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

(54) Title: METHOD AND APPARATUS FOR ENCODING ENHANCEMENT LAYER VIDEO DATA



(57) Abstract: A method for improving the performance of the BLSkip mode in SVC includes the steps of upsampling the motion field of the base layer, interpolating the motion vectors for the intra MBs, interpolating the 8x8 block motion field to a 4x4 block motion field, and generating a MV predictor for a 4x4 block in BLSkip mode using neighbor candidates.

WO 2006/106039 A1

Method and apparatus for encoding enhancement layer video data

Field of the invention

This invention relates to a method for encoding video data.

Background

The H.264/AVC standard provides excellent coding efficiency but it does not consider scalable video coding (SVC). SVC may provide different layers, usually a base layer (BL) and an enhancement layer (EL). To give more functionality for the video codec, the Motion Picture Expert Group (MPEG) considered providing a standard for SVC. Various techniques were proposed, and the Joint Video Team (JVT) finally started a standard called JSVC, with a corresponding reference software description called JSVM. SVC provides temporal, SNR and spatial scalability for applications. The base layer of JSVM is compatible with H.264, and most components of H.264 are used in JSVM as specified, so that only few components need to be adjusted according to the subband structure. Among all the scalabilities, spatial scalability is the most challenging and interesting topic, since it is hard to use the redundancy between the two spatial scalable layers.

SVC provides several techniques for spatial scalability, such as IntraBL mode, residual prediction or BLSkip (base layer skip) mode. These modes can be selected on macroblock (MB) level.

IntraBL mode uses the upsampled reconstructed BL picture to predict a MB in the EL, and only encodes the residual. Residual prediction tries to reduce the energy of the motion compensation (MC) residual of the EL by subtracting the upsampled MC residual of the BL.

BLSkip mode utilizes the upsampled MV for a MB in the EL and requires only the residual to be written into the bit stream if a MB selects this mode. Thus, the BLSkip mode makes use of the redundancy between the MVs of a BL and its EL in the spatial scalability case.

In the JSVM of SVC, BLSkip modes are used for MBs of inter coded predicted (P) frames and inter coded bi-predicted (B) frame. A BL MV, which will be usually stored for each 4x4 block, will be upsampled by multiplication with two. Then the upsampled MV will correspond to an 8x8 block of the higher resolution EL. That is, if the QCIF (176x144) BL frame has (11x9) MBs and each MB has sixteen 4x4 blocks, there are 11x9x16 MVs in the BL (if there is no intra MB). When a selected MV is (h,v) and its corresponding 4x4 block has the start coordinates (x,y), then the upsampled MV is (h*2,v*2) and the corresponding 8x8 block in the high resolution frame (CIF: 352x288) is (2*x,2*y). Thus, four 4x4 blocks with start coordinates of (2x,2y), (2x+4, 2y), (2x,2y+4) and (2x+4,2y+4) will be assigned the same MV of (2h,2v).

Then during the mode decision process, when BLSkip is the current candidate, the MB (if it has the start coordinate of 2x,2y) will set the MVs by using actually four MVs. For the four 8x8 subblocks, four MVs are set for the current

MB, which have the corresponding start coordinates of (2x,2y), (2x+8,2y), (2x,2y+8) and (2x+8,2y+8).

Summary of the Invention

The MV structure mode for the described MB is actually MODE_8x8, which means that the difference of the MV will be specified in the depth of resolution 8x8 rather than 4x4.

The above-described structure means that the MC process for each 8x8 blocks can be handled by using four MVs. However, these 4 MVs are actually the same. Using this mode, the MVs are not used to be coded, since during the decode process they can also be predicted in the same way. So it saves some bits if the predictor is somewhat similar to the true motion. The above solution used by SVC is simple and of low complexity. However, the upsampled MV is not very accurate because of two reasons. One is that the MV have only 1/2 pixel accuracy, but the mostly used MV for MC have 1/4 pixel accuracy; another is that for each 8x8 block, its four 4x4 subblocks use the same MV, which might also be not accurate enough from the resolution point of view.

The problem to be solved by the invention is therefore to improve the accuracy for predicted MVs, particularly for the BLSkip mode.

The present invention gives better motion field upsampling and interpolation methods for spatial scalability so as to improve the coding efficiency of the inter pictures of SVC.

The present invention first smoothes and interpolates the motion field constructed by 8x8 MVs to a better granularity

motion field constructed by 4x4 MVs, and then during the prediction process uses scalar median, vector median and average methods to select a better predictor for the BLSkip mode. As a result, the disclosed MV prediction technique achieves improved coding efficiency, particularly for the BLSkip mode in SVC JSVM 1.0.

According to the invention, a method for encoding higher layer video data includes the steps of
upsampling the motion field of the base layer,
interpolating the motion vectors for the intra coded macroblocks,
interpolating the 8x8 block motion field to a 4x4 block motion field, and
generating in base layer skip mode a motion vector predictor for a 4x4 block, wherein motion vectors from neighbor blocks are used.

Further according to the invention, an apparatus for encoding higher layer video data includes
means for upsampling the motion field of the base layer,
means for interpolating the motion vectors for the intra coded macroblocks,
means for interpolating the 8x8 block motion field to a 4x4 block motion field, and
means for generating in base layer skip mode a motion vector predictor for a 4x4 block, wherein motion vectors from neighbor blocks are used.

Advantageous embodiments of the invention are disclosed in the dependent claims, the following description and the figures.

Brief description of the drawings

Exemplary embodiments of the invention are described with reference to the accompanying drawings, which show in

Fig.1 8x8 MV interpolation for Intra MB;

Fig.2 bilinear interpolation for MVs of 4x4 block; and

Fig.3 relevant blocks for bilinear interpolation of 4x4 block.

Detailed description of the invention

Currently, only the MV upsampled from the co-located MV of the base layer (low resolution layer) is used as a predictor for the enhancement layer (high resolution layer), the accuracy of the upsampled MV is low (usually 1/2 pixel accuracy if we use 1/4 pixel accuracy in our motion estimation (ME) process). And the MV partition is only for 8x8 block. That is, every 4 4x4 block of an 8x8 block will share the same MV. Our invention focus on improving the accuracy of the MV, especially for the BLSkip mode by first using motion field smoothing technique and then give more MVs as predictor for a current MB to determine it will choose the BLSkip mode or not. Given a better predictor (predictors) for the BLSkip mode, which has a high percentage for the MBs in a picture, we can improve the coding efficiency.

In order to achieve more coding efficiency, this invention improves the BLSkip mode by giving a better MV predictor

for a MB that selects BLSkip mode. To get a more accurate MV that will be closer to the true motion, we first utilize several methods to smooth the motion field of the BL, and then utilize strategies to choose the MV for a MB of the EL from the motion field as well as from the neighboring MVs from the current layer.

A motion field is defined in a 2D area P as following: $MV(P) \subset N^2$, where D is a set of lattice pixels and N is the integer field.

Because the default MV accuracy is quarter (1/4) pixel, each MV mv(p) actually is a product of 4 and the replacement of the ME.

For the upsampled motion field from the BL, we can look on it as a function defined in the area P made up of by 8x8 lattice pixels. So P is $\{(P_x, P_y) \mid P_x = 8x, P_y = 8y, \text{ for any } 0 \le x < 22, and 0 \le y < 18\}$. Here, the upper bound of x and y is actually determined by the width and height of the resolution of the frame after upsampling: width/8 and height/8.

Actually if a MB is intra, it has no MV, so there is no definition of the *MV(P)* for the corresponding four values in *P*. So we first give definitions for those values in P by interpolation then using the neighboring MVs.

When we have a definition for any value in the 8x8 lattice definition area of P, we will construct another motion field: $MV(P4) \subset N^2$, where the function is defined in the area of 4x4 lattice pixels.

1. Define the motion field in the intra area

In H.264/AVC, even inter pictures (P or B) have intra MBs. In SVC, the same features are kept. So, at the first step, we give definitions for the lattices that correspond to the intra MBs.

A possible solution is to use bilinear interpolation to get the 8x8 MVs for an intra MB. The first step is to find for an 8x8 block in an Intra MB from each direction the nearest 8x8 blocks that do not belong to Intra MB.

If the start coordinate of the 16x16 MB is (x,y), then the so-called nearest MVs have the following coordinates if the nearby (neighboring) MBs are Inter MBs:

Block 0: A(x-8,y) , B(x ,y-8), C(x+16,y) , D(x ,y+16)
Block 1: A(x-8,y) , B(x+8,y-8), C(x+16,y) , D(x+8,y+16)
Block 2: A(x-8,y+8), B(x ,y-8), C(x+16,y+8), D(x ,y+16)
Block 3: A(x-8,y+8), B(x ,y-8), C(x+16,y+8), D(x+8,y+16)

In Fig.1, the four nearest MVs in different directions (mva,mvb,mvc,mvd) are used to interpolate the MV of block 0 (mv0).

Actually, in the upsampled motion field, four Intra MBs will appear the same time, since a 16x16 Intra MB (whether it is INTRA_16x16 or INTRA_4x4) will turn to a 32x32 block, which can be divided into four 16x16 Intra MBs. If a nearby MB is still Intra coded, the inventive method continues searching in the same direction until encountering an Inter MB, and then uses it. After we get the four 8x8 blocks, we can carry on the MV interpolation process.

We define the distance of two blocks to be calculated by $Dist(P_A - P_B) = |x_a - x_b| + |y_a - y_b|$, where P_A and P_B are the start coordinate of the 8x8 blocks.

E.g. the nearest 8x8 blocks for block 0 in Fig.1 are blocks A,B,C,D. And the distance from 0 to A, B, C, D respectively are D_a=8, D_b=8, D_c=16, D_d=16. The interpolated MV in block 0 is calculated by:

$$mv_0 = ((D_c \cdot mv_a + D_a.mv_c)/(D_a + D_c) + (D_d \cdot mv_b + D_b.mv_d)/(D_b + D_d))/2$$

2. Smooth and interpolate the 8x8 motion field to 4x4 motion field

Now, the motion field is defined in any pixel of the 8x8 lattice. The next step is to create a motion field defined in a 4x4 lattice. One possible solution is illustrated in Fig.2. E.g. the weight for the 4x4 block A (bl4x4A) is 9/16, 3/16, 3/16, 1/16 regarding the 8x8 blocks 0,1,2,3 (bl8x8 0…3) respectively.

In other words, the weight or impact of an 8x8 block on a 4x4 block depends on the number of its 2x2 subblocks that are included or direct neighbours to the 4x4 block. These are 9 from bl8x8_0, 3 from bl8x8_1 and bl8x8_2 each and 1 from bl8x8_3, as shown in Fig.3 corresponding to Fig.2.

For each central point of a 4x4 block, there are four nearest central points of the 8x8 blocks, which contain a MV (mv_a,mv_b,mv_c,mv_d). The MV of a 4x4 block is interpolated by the following formula:

$$mv = \sum_{i=0}^{3} mv_i \cdot weight_v \cdot weight_h$$

where subscript x is A,B,C,D as shown in Fig.2. There are two weights for each predictor MV in the formula.

$$Weight_v = (x_p - x_0)/(x_1 - x_0)$$

$$Weight_h = (y_p - y_0)/(y_4 - y_0)$$

Other solutions are also available for interpolation and smoothing of the motion field.

3. Motion vector prediction for the BLSkip mode

When during the mode decision process the current mode candidate is BLSkip, the motion compensation mode will be set as MODE_4x4, because the BL motion field is actually based on 4x4 lattices.

A simple solution is to utilize as the MV for the BLSkip mode for each 4x4 block of a MB the co-located MV of the 4x4 block of the upsampled motion field.

Another solution is to use the neighboring MVs to predict the MV for the current 4x4 block. The candidates can be MVs of the four blocks (left 4x4 block, above 4x4 block, down 4x4 block and right 4x4 block) or those of the eight blocks (by adding left above 4x4 block, left down 4x4 block, right above 4x4 block and right down 4x4 block).

For the certain candidate MVs, several methods can be used to get the predictor: Average the candidates, Scalar Median the candidates or Vector Median the candidates.

Average means that the horizontal value and the vertical value of the MV are generated by averaging the respective horizontal or vertical values of the candidates.

Scalar Median is also generated by getting the median of horizontal values and vertical values respectively. To get the median value of several values, a simple solution is to sort the values first and get the value in the sorted list using the median index.

The Vector Median method gets out the MV that has the smallest sum of distance from others. So the solution of the following equation is the vector median MV.

$$mv_{VM} = \arg\min_{mv_i}\left(\sum_{i \neq j} Dist(mv_i, mv_j)\right)$$

For low pass pictures, or if the reference frame is far from the current frame, the Average method will give better results. If the distance of the two frames is close, Median or Vector Median method will be better. One adaptive MV prediction strategy is to utilize the Average method for low pass frames, and utilize the Median or the Vector Median method for high pass frames.

In principle, the inventive method for improving the performance of the BLSkip mode in SVC includes the following steps:

- upsampling the motion field of the base layer;
- interpolating the motion vectors for the intra MBs;

- interpolating the 8x8 block motion field to a 4x4 block motion field;
- generating a MV predictor for a 4x4 block in BLSkip mode using neighbor candidates.

An apparatus for improving the performance of the BLSkip mode in SVC according to the invention has:

- means for upsampling the motion field of the base layer;
- means for interpolating the motion vectors for the intra MBs;
- means for interpolating the 8x8 block motion field to a 4x4 block motion field; and
- means for generating a MV predictor for a 4x4 block in BLSkip mode from neighbour candidates.

In one embodiment of the invention, MVs from the four nearest 8x8 blocks of different directions are used to predict the 8x8 blocks in Intra MBs.

In one embodiment of the invention, bilinear interpolation is used to interpolate the 8x8 motion field to a 4x4 motion field.

In one embodiment, a selection strategy, e.g. Average, Median and Vector Median, is used to get the MV from the neighboring MV candidates.

In one embodiment, an adaptive MV selection strategy is used that takes into account the distance between the current frame and the reference frame: if the distance is short, Median or Vector Median is preferred, otherwise the Average.

In short, JSVM used the upsampled BL MV as the MV for the higher resolution layer for the BLSkip mode. However the upsampled MV is not good enough and can be improved. All the upsampled MVs are actually half pixel accuracy and each four 4x4 blocks in an 8x8 block share the same MV. In the present invention, we provide method and apparatus to give better motion vector for the BLSkip mode.

The core of the invention is the following: In order to achieve more coding efficiency, this invention improves the BLSkip mode by giving better MV predictor for a MB that selects a BLSkip mode. To get a more accurate MV that will be closer to the true motion, we first utilize several methods to smooth the motion field of the BL, and then utilize strategies to choose the MV for a MB of the EL from the motion field as well as from the neighboring MVs from the current layer.

The main advantage of the present invention is that the coding efficiency of inter coded pictures of SVC is improved, which is achieved by better motion field upsampling and adaptive interpolation methods.

Claims

1. Method for encoding spatial enhancement layer video data including the steps of
   - upsampling the motion field of the base layer;
   - assigning the motion vectors of the upsampled motion field to 16x16 macroblocks of the enhancement layer;
   - generating motion vectors for 8x8 blocks from the motion vectors of 16x16 macroblocks, wherein the motion vectors for intra coded 8x8 blocks (block0, ...,block3) are interpolated from the motion vectors of the four nearest inter coded 8x8 blocks (blockA,...,blockD); and
   - interpolating motion vectors of 4x4 blocks from the motion vectors of the 8x8 blocks (block0,..., block3).

2. Method according to claim 1, wherein for said generating the motion vectors of intra coded 8x8 blocks (block0) a bilinear interpolation is used.

3. Method according to claim 2, wherein the bilinear interpolation considers the distance (D_a,D_b,D_c,D_d) from a current 8x8 block to the four nearest inter coded 8x8 blocks (blockA,...,blockD) according to $mv_0 = ((D_c \cdot mv_a + D_a.mv_c)/(D_a + D_c) + (D_d \cdot mv_b + D_b.mv_d)/(D_b + D_d))/2$.

4. Method according to any of the previous claims, wherein for said interpolating motion vectors of 4x4 blocks (bl4x4A,...,bl4x4D) only motion vectors of 8x8 blocks of the current 16x16 macroblock are used.

5. Method according to claim 4, wherein said interpolating motion vectors of 4x4 blocks (bl4x4A, ...,bl4x4D) comprises assigning weights to the motion vectors of 8x8 blocks of the current 16x16 macroblock, wherein the weight assigned to a motion vector of an 8x8 block depends on the number of its 2x2 subblocks that are direct neighbours to the 4x4 block or included in the 4x4 block.

6. Method according to any of the previous claims, further comprising the step of performing an encoding mode decision, wherein for base layer skip mode said 4x4 motion vectors are used as predictors.

7. Apparatus for encoding spatial enhancement layer video data including
   - means for upsampling the motion field of the base layer;
   - means for assigning the motion vectors of the upsampled motion field to 16x16 macroblocks of the enhancement layer;
   - means for generating motion vectors for 8x8 blocks from the motion vectors of 16x16 macroblocks, wherein the motion vectors for intra coded 8x8 blocks (block0, ...,block3) are interpolated from the motion vectors of the four nearest inter coded 8x8 blocks (blockA,...,blockD); and
   - means for interpolating motion vectors of 4x4 blocks from the motion vectors of the 8x8 blocks (block0,..., block3).

8. Apparatus according to the previous claim, wherein for said interpolating motion vectors of 4x4 blocks

(bl4x4A,...,bl4x4D) only motion vectors of 8x8 blocks of the current 16x16 macroblock are used.

9. Apparatus according to claim 7 or 8, wherein said interpolating motion vectors of 4x4 blocks (bl4x4A, ...,bl4x4D) comprises assigning weights to the motion vectors of 8x8 blocks of the current 16x16 macroblock, wherein the weight assigned to a motion vector of an 8x8 block depends on the number of its 2x2 subblocks that are direct neighbours to the 4x4 block or included in the 4x4 block.

10. Apparatus according to claim 7, 8 or 9, further comprising means for performing an encoding mode decision, wherein for base layer skip mode said 4x4 motion vectors are used as predictors.

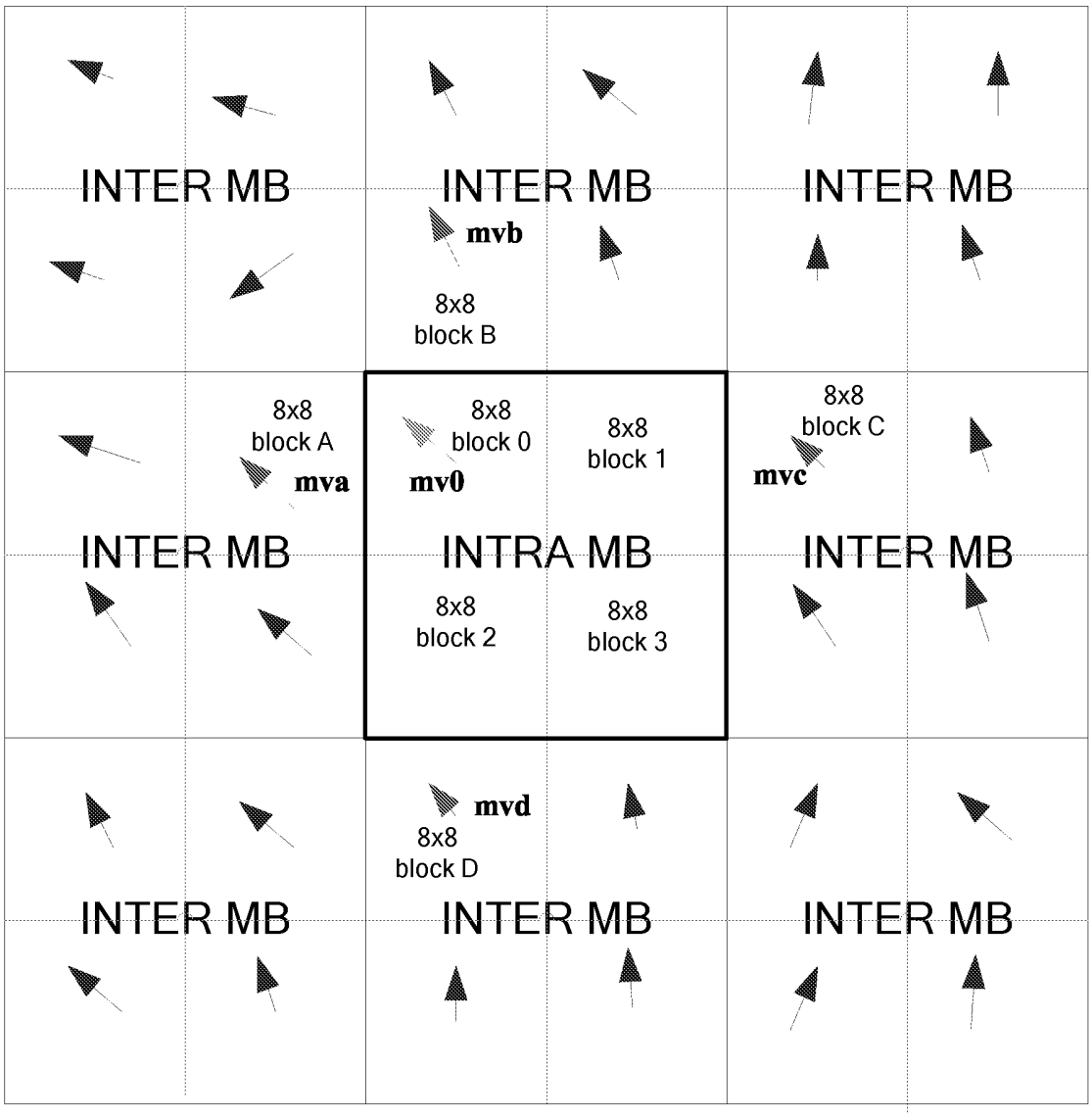

INTER MB
INTER MB
mvb
8x8
block B
INTER MB
8x8
block A
mva
8x8
block 0
mv0
8x8
block 1
8x8
block C
mvc
INTER MB
INTRA MB
INTER MB
8x8
block 2
8x8
block 3
mvd
8x8
block D
INTER MB
INTER MB
INTER MB

Fig.1

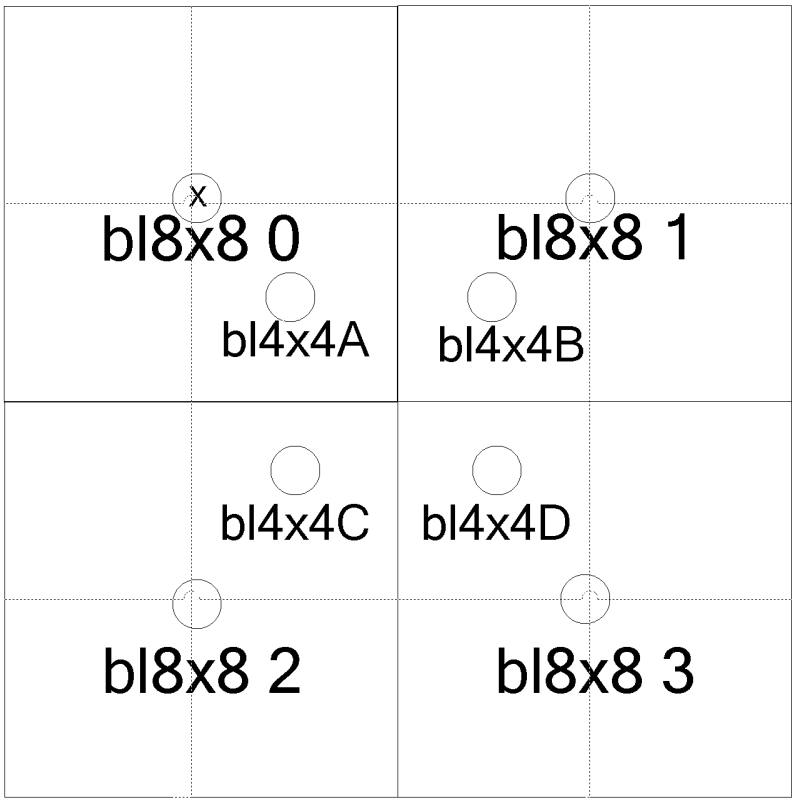

x
bl8x8 0
bl8x8 1
bl4x4A
bl4x4B
bl4x4C
bl4x4D
bl8x8 2
bl8x8 3

Fig.2

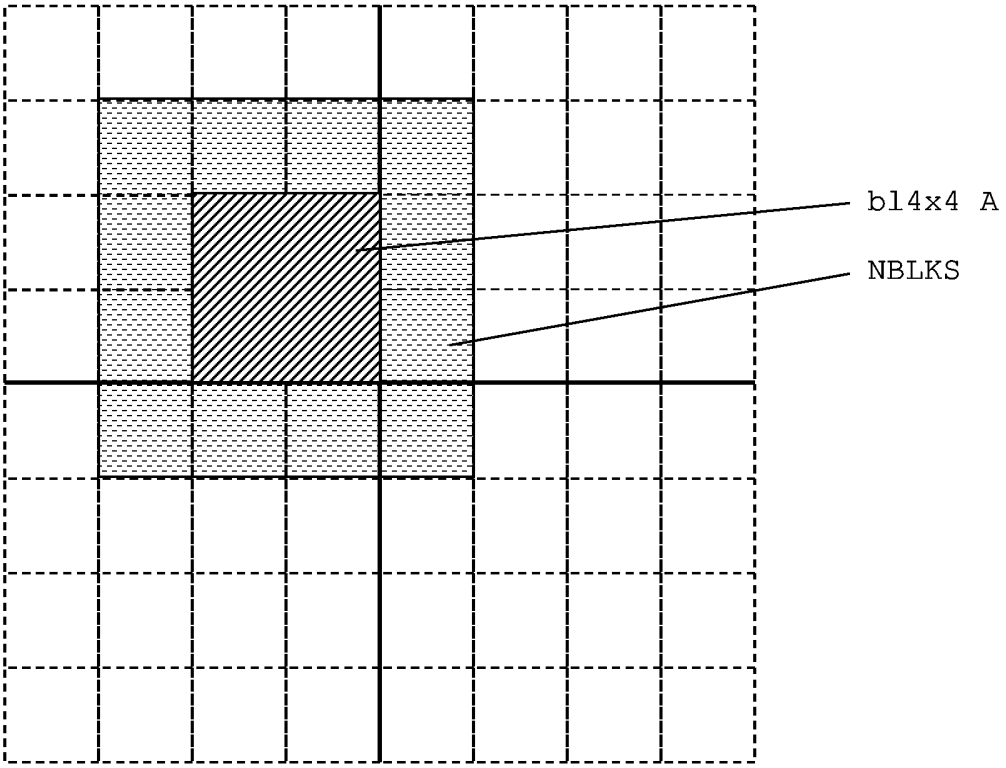

bl4x4 A
NBLKS

Fig.3

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2006/060873

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/26 H04N7/24

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	REICHEL J ET AL: "Joint Scalable Video Model JSVM 0" JOINT VIDEO TEAM (JVT) OF ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 AND ITU-T SG16 Q6), XX, XX, 17 January 2005 (2005-01-17), pages 1-73, XP002345849 the whole document paragraph [3.3.2]	1-10

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier 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
2 June 2006	13/06/2006

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	Schneiderlin, J