(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 884 118 B9**

(12) **CORRECTED EUROPEAN PATENT SPECIFICATION**

(15) Correction information:
Corrected version no 1 (W1 B1)
Corrections, see
Description Paragraph(s) 12, 16, 20-22
Claims EN 4-7, 9-10

(48) Corrigendum issued on:
12.06.2013 Bulletin 2013/24

(45) Date of publication and mention of the grant of the patent:
15.07.2009 Bulletin 2009/29

(21) Application number: **06743398.7**

(22) Date of filing: **03.05.2006**

(51) Int Cl.:
H04N 7/26 *(2006.01)* ***H04N 7/46*** *(2006.01)*

(86) International application number:
PCT/EP2006/061994

(87) International publication number:
WO 2006/125713 (30.11.2006 Gazette 2006/48)

(54) METHOD AND APPARATUS FOR ENCODING VIDEO DATA, AND METHOD AND APPARATUS FOR DECODING VIDEO DATA

VERFAHREN UND VORRICHTUNG ZUR KODIERUNG VON VIDEODATEN SOWIE VERFAHREN UND VORRICHTUNG ZUR DEKODIERUNG VON VIDEODATEN

MÉTHODE ET APPAREIL DE CODAGE ET DÉCODAGE DE DONNÉES VIDEO

(84) Designated Contracting States:
DE FR GB

(30) Priority: **27.05.2005 EP 05300426**

(43) Date of publication of application:
06.02.2008 Bulletin 2008/06

(73) Proprietor: **Thomson Licensing**
92130 Issy-les-Moulineaux (FR)

(72) Inventors:
- **CHEN, Ying**
  Hai Dian District, Beijing 100083 (CN)
- **ZHAI, Jiefu**
  Hai Dian District,
  Beijing 100083 (CN)
- **GAO, Kui**
  Zhongguancun,Beijing 100080 (CN)
- **XIE, Kai**
  Hai Dian District
  Beijing 100086 (CN)

(74) Representative: **Rittner, Karsten**
Deutsche Thomson OHG
European Patent Operations
Karl-Wiechert-Allee 74
30625 Hannover (DE)

(56) References cited:
- **SCHWARZ H ET AL: "Scalable Extension of H. 264/AVC" ISO/IEC JTC1/CS29/WG11 MPEG04/M10569/S03, March 2004 (2004-03), pages 1-39, XP002340402**
- **REICHEL J ET AL: "Scalable Video Model 3.0" ISO/IEC JTC1/SC29/WG11 N6716,, October 2004 (2004-10), pages 1-85, XP002341767**
- **SUN S & FRANCOIS E: "Extended Spatial Scalability with picture-level adaptation" JOINT VIDEO TEAM (JVT) OF ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 AND ITU-T SG16 Q.6), JVT-O008, 16 April 2005 (2005-04-16), pages 1-20, XP002347764**
- **"Recommendation H.263: Video coding for low bit rate communication" ITU-T DRAFT RECOMMENDATION H.263, February 1998 (1998-02), pages 1-167, XP002176560**

EP 1 884 118 B9

Description

Field of the invention

[0001] This invention relates to a method and an apparatus for encoding video data, and to a method and an apparatus for decoding video data.

Background

[0002] The H.264/AVC standard provides excellent coding efficiency but it does not consider scalable video coding (SVC). SVC provides different layers, usually a base-layer (BL) and an enhancement-layer (EL). The Motion Picture Expert Group (MPEG) works on enhanced functionality of the video codec. Various techniques were proposed, and the Joint Video Team (JVT) started a standard called JSVC, with corresponding reference software (JSVM). SVC provides temporal, SNR and spatial scalability for applications. The BL 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, since it is hard to use the redundancy between the two spatial scalable layers.

[0003] 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.

[0004] 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 BL motion vector (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.

[0005] For Inter coded pictures, including both P pictures and B pictures of SVC, residual prediction is used to decrease the energy of the residual for improving coding efficiency. The basic idea is to first get the predicted residual by upsampling the residual signal of the corresponding BL picture, wherein a 2-tap bilinear filter is used. Then the predicted residual is subtracted from the real residual which is obtained from the motion estimation in the EL, and the difference is coded by DCT, entropy coding etc.

[0006] Residual upsampling is commonly done MB by MB, and for each MB by 4x4, 8x8 or 16x16 subblocks, based on MC accuracy. If the MC accuracy is e.g. 16x16, the whole 16x16 MB uses just one motion vector; if the MC accuracy is 8x8, each four 8x8 sub-blocks may have different motion vectors. The residuals for different 8x8 sub-blocks have low correlation, so the upsampling process is done for four different sub-blocks. SVC utilizes a simple 2-tap bilinear filter, performing the upsampling process first in the horizontal and then in the vertical direction. The respective filter works on MB level, and thus cannot cross the boundary of an 8x8 block.

[0007] An option for the described procedure is whether to use residual prediction or not for a particular MB. A mode decision process tries different modes, all with or without residual prediction. This is called adaptive residual prediction.

[0008] The typical frame structure employed by H.264/SVC contains two intra-coded reference frames that are used at the receiver for Instantaneous Decoder Refresh (IDR), and then a number of intra-coded or inter-coded frames, which make several GOPs (group-of-pictures). Inter-coded frames can be interpolated or predicted. In wavelet decomposition, the EL of a GOP typically consists of several high-pass frames followed by a low-pass frame. A low-pass frame is used for both the preceding and the following high-pass frames, i.e. for two GOPs .

[0009] Document REICHEL J ET AL: "Scalable Video Model 3.0" ISO/IEC JTC1/SC29/WG11 N6716" October 2004 (2004-10), pages 1-85, XP002341767, discloses a method for encoding video data containing high-pass frames and low-pass frames into at least two spatially scalable layers obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the encoding is based on prediction and update steps, and wherein inter-layer residual prediction on macroblock level is used.

Summary of the Invention

[0010] Using the predicted residual is a very complex process, both for the encoder and the decoder. Thus, a simplified processing is desirable that enables usage of less complex encoders and/or decoders. Disabling the residual prediction will reduce the decoder complexity by a factor of about two, but it reduces coding efficiency. On the other hand, adaptive residual prediction is a very potential technique that may improve coding efficiency of about 5-10% for the same objective quality (based on PSNR). Generally, it is desirable to substantially maintain or even improve the level of efficiency of the encoding and decoding processes. So it is advisable to enable residual prediction. But if e.g. for real-time applications residual prediction is enabled for every picture, known decoders are too slow for real-time decoding of e.g. a dual-layer (QCIF/CIF) spatial scalable bit-stream.

[0011] The present invention provides simplified residual prediction techniques, focusing on reducing the encoding and/or decoding complexity of the spatial scalable EL, with the coding efficiency being only slightly reduced or even improved.

[0012] According to one aspect of the invention, a method for encoding video data into at least two spatially

scalable layers containing high-pass frames and low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the encoding is based on prediction and update steps, comprises the steps of encoding the low-pass frames, wherein inter-layer residual prediction may be used,

splitting the high-pass frames into two (preferably interleaving) frame groups, e.g. by assigning each of the high-pass frames to one group, e.g. according to their sequence number i.e. using a fixed raster,

encoding the frames of the first of said frame groups, wherein inter-layer residual prediction may be used, and encoding the frames of the second of the frame groups using an encoding method without inter-layer residual prediction, i.e. residual prediction is prohibited for these frames.

[0013] In a preferred embodiment of this aspect of the invention, encoding the frames of the first of the frame groups comprises mode selection, wherein none of the possible encoding modes uses inter-layer residual prediction on macroblock level. In one embodiment however at least one of the encoding modes for the first frame group uses inter-layer residual prediction on frame level, which is called "Simplified Residual Prediction" herein. For simplified residual prediction, the EL residual is preferably generated from the information of the EL, and the BL residual is not used.

[0014] In a particularly preferred embodiment, the first frame group comprises the even high-pass frames and the second frame group comprises the odd high-pass frames.

[0015] In particular, the invention is advantageous for those high-pass and low-pass frames that belong to an enhancement-layer of a scalable video signal. Thus, an improved and simplified residual prediction scheme for SVC is provided.

[0016] The resulting video signal generated by an encoder according to one embodiment of the invention comprises at least two spatially scalable layers, a BL and an EL, wherein the EL contains encoded low-pass frame data and encoded high-pass frame data as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), and wherein the encoded high-pass frame data contain an encoding mode indication and can be split into two types or groups, with the high-pass frame data of one of these groups containing an indication indicating if it was encoded using inter-layer residual prediction, and at least some of them being encoded using inter-layer residual prediction on frame level but not on MB level, and the high-pass frame data of the other of these groups being encoded without using inter-layer residual prediction. Thus, these frames need not contain such inter-layer residual prediction indication. The frame data of the second group contain fully encoded residual pictures, without using any residual prediction.

[0017] Thus inter-layer, residual prediction can be skipped for the second group of high-pass frames, and it can be performed on frame level instead of MB level for the first group of high-pass frames. This leads to reduced complexity of encoders as well as decoders, since e.g. the frame needs not be split into blocks during encoding neither decoding.

[0018] As an example, if the size of a GOP is 16, its structure is (previous IDR frames not shown, with P being predicted frames and Bx being bilinear predicted frames):
P1 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 B13 B14 B15 P2

According to a preferred embodiment of the invention, there is no residual prediction done for the odd frames: B1, B3, B5, B7, B9, B11, B13, B15

Simplified adaptive residual prediction is done for the even frames: B2, B4, B6, B8, B10, B12, B14

Original conventional Adaptive residual prediction: P1 P2

[0019] Advantageously, for the encoding method according to the invention there is no need to split the frames into MBs.

[0020] A corresponding device for encoding video data into at least two spatially scalable layers containing high-pass frames and low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the encoding is based on prediction and update steps, comprises means for encoding the low-pass frames, wherein inter-layer residual prediction may be used,

means for splitting the high-pass frames into two interleaving frame groups,

means for encoding the frames of the first of said frame groups, wherein inter-layer residual prediction may be used, and

means for encoding the frames of the second of the frame groups using an encoding method without residual prediction. Further, the device may comprise means for inserting a residual prediction indication flag into the frames of the first frame group.

[0021] According to another aspect of the invention, a method for decoding video data containing at least two spatially scalable layers, and encoded high-pass frames and encoded low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the decoding of at least the high-pass frames is based on inverse prediction and inverse update steps, comprises the steps of

decoding the low-pass frames according to their encoding mode,

determining from the sequential order of a high-pass frame whether it belongs to a first or a second group of frames (B_{even},B_{odd}), and

decoding the high-pass frame, wherein if the high-pass frame belongs to the first group of frames (B_{even}) the decoding uses inter-layer prediction of the residual that is used for the inverse prediction and inverse update steps, and if the high-pass frame belongs to the second group of frames the residual that is used for the inverse prediction and inverse update steps is obtained without inter-layer prediction. Said prediction may use upsam-

pling of the corresponding BL residual, but in principle also another prediction technique.

[0022] A corresponding device for decoding video data, the video data containing at least two spatially scalable layers, and encoded high-pass frames and encoded low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the decoding of at least the encoded high-pass frames is based on inverse prediction and inverse update steps, comprises at least means for decoding the low-pass frames according to their encoding mode,
means for determining from the sequential order of a high-pass frame whether it belongs to a first or a second group of frames (B_{even}, B_{odd}), and
means for decoding the high-pass frame, wherein if the high-pass frame belongs to the first group of frames (B_{even}) the means for decoding performs inter-layer prediction of the residual that is used for the inverse prediction and inverse update steps, and if the high-pass frame belongs to the second group of frames the residual that is used for the inverse prediction and inverse update steps is obtained without inter-layer prediction.

[0023] Advantageous embodiments of the invention are disclosed in the dependent claims, the following description and the figures.

Brief description of the drawings

[0024] Exemplary embodiments of the invention are described with reference to the accompanying drawings, which show in

Fig.1 the principle of residual upsampling in horizontal direction;

Fig.2 a residual prediction scheme according to the invention for a GOP with GopSize=8;

Fig.3 block boundaries for conventional residual prediction; and

Fig.4 simplified residual prediction.

Detailed description of the invention

[0025] The following text refers to frames as well as to pictures. When frames are mentioned, the same applies to pictures and vice versa.

[0026] Fig.1 shows the principle of residual upsampling in horizontal direction, using 4x4 motion estimation accuracy. SVC utilizes e.g. simple 2-tap bilinear filters, performing first an upsampling process on the horizontal direction and then on the vertical direction. An upsampled pixel value is generated by averaging two nearby original pixels, and the location of the upsampled pixel is just in the middle of the two original pixels, so the coefficients are [1/2, 1/2]. The 2-tap filter can't cross the MB boundary. While there is no problem at the left boundary, the upsampled pixel at the right boundary is just a copy of its nearest neighbor pixel. This is called "nearest neighboring method".

[0027] For the conventional upsampling, the whole residual picture must first be split into MBs and then into sub-blocks. This means a lot of memory copy operations and upsampling steps for small blocks. That is a main reason for the high complexity of the decoder when residual prediction is enabled.

[0028] The present invention discloses methods to reduce the complexity of the encoder and the complexity of the decoder, by adaptively using the residual prediction, partly skipping residual prediction and modifying the residual prediction method.

[0029] Typically, for each GOP there is one of the P-frames that is a low-pass frame. E.g. if the GOP size is 16, the parameter picture_id_inside_gop associated with the low-pass frame is 16. The other 15 frames being P- or B-frames are high-pass frames with different decomposition levels. The frames with the highest decomposition levels are those with odd numbers for picture_id_inside_gop: 1, 3, 5, 7, 9, 11, 13, 15. We call those pictures odd pictures. In this invention, we propose two solutions to substitute the conventional residual upsampling process.

[0030] One solution is doing the residual prediction using the conventional upsampling method. However, the process is not performed on blocks, but on the whole frame. This means that the 2-tap filter ignores any boundary within the frame until it reaches the boundary of the whole frame. Thus, there is no need to split the whole residual frame into MBs or sub-blocks.

[0031] The second solution is not to use any residual prediction at all for some frames, since once residual prediction is used (to improve the coding efficiency), adaptive type will be chosen in the mode decision process. That is, during mode decision all modes related to motion compensation will try two different sub-modes, the mode with residual prediction or the mode without residual prediction. A flag indicating which mode was chosen (residual prediction_flag) will be written into each MB of a picture.

[0032] As experiments show, low-pass pictures have a high probability of using residual prediction. Typically about 30% of the MBs of a low-pass picture will enable the residual_prediction_flag. However, it was found that the higher the decomposition stage of the high-pass picture is, the less MBs use residual prediction. For the highest decomposition stage pictures (the odd pictures), only very few MBs have the residual_prediction_flag enabled.

[0033] According to the present invention, the high-pass frames of a GOP are split into two interleaving groups, and for the frames or pictures of one of these groups the residual prediction is done on frame level, while for all the frames of the other group the residual prediction is prohibited and can thus be skipped during mode decision.

[0034] In one embodiment of the present invention, the residual prediction is prohibited for all the odd pictures.

[0035] Advantageously, even if no residual prediction is used at all, the viewer can usually not notice the decrease in rate distortion (RD) performance, because the mode of only few MBs is changed. Another important reason is that when residual prediction is not used at all for any MB in the whole picture, then each MB will save one bit for the residual_prediction_flag. Even after entropy coding some bits will be saved, so that the coding efficiency for odd pictures is improved.

[0036] Actually, from the complexity point of view, if residual prediction is disabled, about half of the computation effort during the decoding process will be saved because the upsampling process can be skipped for each MB. This is advantageous for real-time decoders.

[0037] Another method to reduce the complexity of the decoder is for the other high-pass pictures (i.e. even high-pass pictures), we do the residual upsampling based on the whole frame. The advantage is that we don't actually need to detect the boundary of a motion estimation (ME) block, and the 2-tap filter will be implemented in the same way throughout the whole frame, until it encounters the boundary of the frame. So, we don't need to split the whole residual frame into blocks before residual upsampling. So we save the splitting time and the memory moving time for the small blocks.

[0038] As shown in Fig.2, the original adaptive residual prediction (ORP) is utilized for low-pass pictures, in order to preserve high coding efficiency. For the highest decomposition stage, e.g. frames that have odd values of picture_id_inside_gop, no residual predication (NRP) is used and the residual_prediction_flag is saved for each MB in these frames. For the other high-pass frames (with picture_id_inside_gop = 2,4,6), we can choose to use a weak residual prediction scheme, like e.g. the above-described Simplified Residual Prediction (SRP).

[0039] Fig.3 and Fig.4 show the differences between residual prediction and simplified residual prediction. In the current JSVM residual prediction is done in blocks. When in Fig.3 the filter encounters a boundary of a ME block, it will stop and the boundary of the upsampled residual will be predicted using the nearest neighboring method, as shown in Fig.1. However, to reduce the complexity, we choose to simplify the residual prediction process employed for some high-pass pictures. In those cases when residual prediction is done, it doesn't need to be based on MB level any more. So the whole residual frame does not need to be split, and many memory operations are not required. For the 2-tap filter, there is no difference until the convolve operation reaches the right or bottom boundary of the frame, as shown in Fig.4. The boundaries in Figs.3 and 4 are for the convolve operation.

[0040] In principle, the two described techniques being simplified residual prediction for at least some frames and the skipping of residual prediction for at least some of the other frames can also be used independent from each other, or in combination with other modes. E.g. the high-pass frames can be split into three groups, with one using conventional encoding, the second using simplified residual prediction and the third skipping residual prediction.

[0041] Further, other interleaving schemes can be applied than assigning even high-pass frames to one group and odd high-pass frames to the other group. E.g. the following scheme may be used:

No residual prediction for: B1-B3, B5-B7, B9-B11, B13-B15 Simplified adaptive residual prediction for: B4, B8, B12 And Original adaptive residual prediction for: P1 P2

[0042] Another possible scheme is e.g. to swap the previously described groups for "No residual prediction" and for "Simplified adaptive residual prediction".

[0043] When a decoder receives the video signal that results from the encoding method according to the invention, then it can determine from the sequence number of a particular B-frame whether residual prediction for it was skipped or not: e.g. all odd frames skip the residual prediction. Alternatively, it can evaluate if a frame contains the above-mentioned residual_prediction_flag that indicates that residual prediction might have been used. If this flag is not present, the decoder can deduce that residual prediction was skipped during encoding.

[0044] Adaptive residual prediction means that also other possible modes will be tested, e.g. inter4x4, inter16x16, inter8x8. So when we say adaptive residual prediction, all these modes may be tested with or without residual prediction. Thus, in the preferred embodiment of the invention mode selection is used for all high-pass frames (B1,...,B15), but:

- for the even frames the encoder can select between different modes, each with or without residual prediction; if residual prediction is selected, it will be on frame level, i.e. simplified residual prediction.
- for the odd frames the encoder can also select between different modes, but residual prediction is not allowed for any of the modes, i.e. residual prediction is disabled.

[0045] Thus, a high-pass frame includes an indication (e.g. a flag) that shows how it was encoded.

[0046] So, two decisions need to be made during encoding. The first is whether to use residual prediction or not for a frame. This flag indicating this option is already a part of the Picture Parameter Set (PPS), so it is an encoder issue. The second is how to do the residual prediction: the simplified or the original type. To indicate the result of this decision, one possibility is to add a flag into the PPS, which however should preferably be normative. Then the decoder can detect the corresponding decoding method from this flag.

[0047] In this invention, based on the different importance of the residual prediction for different decomposi-

tion levels of the Inter (high-pass) pictures, a simplified solution is proposed that greatly reduces the decoder complexity.

[0048] The invention can be used for video encoding and decoding, particularly when the video contains two or more spatial scalable layers and uses residuals resulting e.g. from motion estimation.

Claims

1. Method for encoding video data into at least two spatially scalable layers containing high-pass frames and low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the encoding is based on prediction and update steps, the method comprising the steps of

   - encoding the low-pass frames, wherein inter-layer residual prediction (ORP) on frame level may be used;
   - splitting the high-pass frames into two interleaving frame groups;
   - encoding the frames of the first of said frame groups, wherein inter-layer residual prediction on frame level may be used; and
   - encoding the frames of the second of the frame groups using an encoding method without inter-layer residual prediction (NoRP).

2. Method according to claim 1, wherein the step of encoding the frames of the first of the frame groups comprises selection of one of a plurality of encoding modes, wherein at least one of the possible encoding modes uses inter-layer residual prediction on frame level (SRP), but none of the possible encoding modes uses inter-layer residual prediction on macroblock level.

3. Method according to claim 1 or 2, wherein the first frame group comprises the even high-pass frames and the second frame group comprises the odd high-pass frames.

4. Method according to any of the claims 1-3, wherein the step of encoding the frames of the second of the frame groups comprises encoding mode selection.

5. Video signal comprising at least two spatially scalable layers (BL,EL), wherein the higher layer (EL) contains encoded low-pass frame data (P1,P2) and encoded high-pass frame data (B1,...,B15) as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), and wherein the encoded high-pass frame data contain an encoding mode indication and can be split into two groups, the splitting being based on the sequential order of the frames, with the frame data of one of these groups (B_{even}) containing an indication indicating if the frame was encoded using inter-layer residual prediction and at least some of the frames being encoded using inter-layer residual prediction on frame level but not on MB level, and the high-pass frame data of the other of these groups (B_{odd}) being encoded without using inter-layer residual prediction.

6. Method for decoding video data containing at least two spatially scalable layers, and coded high-pass frames and low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the decoding of at least the high-pass frames is based on inverse prediction and inverse update steps, the method comprising the steps of

   - decoding the low-pass frames according to their encoding mode;
   - determining from the sequential order of a high-pass frame whether it belongs to a first or a second (B_{even}, B_{odd}) group of frames;
   - decoding the high-pass frame, wherein if the high-pass frame belongs to the first group of frames (B_{even}) the decoding uses inter-layer prediction on frame level of the residual that is used for the inverse prediction and inverse update steps, and if the high-pass frame belongs to the second group of frames the residual that is used for the inverse inter-layer prediction and inverse update steps is obtained without prediction.

7. Apparatus for encoding video data into at least two spatially scalable layers the video data containing high-pass frames and low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the encoding is based on prediction and update steps, comprising

   - means for encoding the low-pass frames, wherein into inter-layer residual prediction (ORP) on frame level may be used;
   - means for splitting the high-pass frames into two interleaving frame groups;
   - means for encoding the frames of the first of said frame groups, wherein inter-layer residual prediction on frame level may be used; and
   - means for encoding the frames of the second of the frame groups using an encoding method without inter-layer residual prediction (NoRP).

8. Apparatus according to the previous claim, wherein the means for encoding the frames of the second of the frame groups comprises means for performing

encoding mode selection.

9. Apparatus for decoding video data, the video data containing at least two spatially scalable layers, and encoded high-pass frames and encoded low-pass frames as obtained by temporal wavelet decomposition using motion compensated temporal filtering (MCTF), wherein the decoding of at least the encoded high-pass frames is based on inverse prediction on frame level and inverse update steps, the apparatus comprising

- means for decoding the low-pass frames according to their encoding mode;
- means for determining from the sequential order of a high-pass frame whether it belongs to a first or a second (B_{even}, B_{odd}) group of frames;
- means for decoding the high-pass frame, wherein if the high-pass frame belongs to the first group of frames (B_{even}) the means for decoding performs inter-layer prediction on frame level of the residual that is used for the inverse prediction and inverse update steps, and if the high-pass frame belongs to the second group of frames the residual that is used for the inverse prediction and inverse update steps is obtained without inter-layer prediction.

10. Method or apparatus according to any of the preceding claims, wherein inter-layer residual prediction for a frame of the higher layer (EL) comprises upsampling the residual of the corresponding frame of the lower layer (BL).

Patentansprüche

1. Verfahren zum Kodieren von Videodaten in wenigstens zwei räumlich skalierbare Schichten, die Hochpass-vollbilder und Tiefpass-Vollbilder enthalten, derart wie sie durch zeitliche Wavelet-Zerlegung unter Verwendung von bewegungskompensierter zeitlicher Filterung (MCTF) erhalten werden, wobei die Kodierung auf Vorhersage- und Aktualisierungs-Schritten beruht, umfassend die Schritte:

- Kodieren der Tiefpass-Vollbilder, wobei zwischen den Schichten eine Restvorhersage (ORP) auf Vollbildebene verwendet werden kann;
- Aufspalten der Hochpass-Vollbilder in zwei verschachtelte Vollbildgruppen;
- Kodieren der Vollbilder der ersten Vollbildgruppe, wobei zwischen den Schichten eine Restvorhersage auf Vollbildebene verwendet werden kann;
- Kodieren der Vollbilder der zweiten Vollbildgruppe unter Verwendung eines Kodierverfahrens ohne Restvorhersage zwischen den Schichten (NoRP).

2. Verfahren nach Anspruch 1, bei dem der Schritt des Kodierens der Vollbilder der ersten Vollbildgruppe die Auswahl einer von mehreren Kodierbetriebsarten umfasst, wobei wenigstens eine der möglichen Kodierbetriebsarten Restvorhersage zwischen den Schichten auf Vollbildebene (SRP) verwendet, aber keine der möglichen Kodierbetriebsarten Restvorhersage zwischen den Schichten auf Makroblockebene verwendet.

3. Verfahren nach Anspruch 1 oder 2, bei dem die erste Vollbildgruppe die geradzahligen Hochpass-Vollbilder und die zweite Vollbildgruppe die ungeradzahligen Hochpass-Vollbilder umfasst.

4. verfahren nach einem der Ansprüche 1 bis 3, bei dem der Schritt des Kodierens der Vollbilder der zweiten Vollbildgruppe eine Kodierbetriebsart-Auswahl umfasst.

5. Videosignal, das wenigstens zwei räumlich skalierbare Schichten (BL, EL) umfasst, wobei die höhere Schicht (EL) kodierte Tiefpass-Vollbilddaten (P1, P2) und kodierte Hochpass-Vollbilddaten (B1, ..., B15) enthält, derart wie sie durch zeitliche Wavelet-Zerlegung unter Verwendung von bewegungskompensierter zeitlicher Filterung (MCTF) erhalten werden, und wobei die kodierten Hochpass-Vollbilddaten eine Kodierbetriebsart-Anzeige enthalten und in zwei Gruppen aufgespalten werden können, wobei die Aufspaltung auf der sequentiellen Reihenfolge der Vollbilder beruht und die Vollbilddaten einer dieser Gruppen ($B_{geradzahlig}$) eine Anzeige enthalten, die anzeigt, ob das Vollbild unter Verwendung von Restvorhersage zwischen den Schichten kodiert wurde und wenigstens einige Vollbilder, die kodiert werden, Restvorhersage zwischen den Schichten auf Vollbildebene, aber nicht auf MB-Ebene verwenden, und die Hochpass-Vollbilddaten der anderen dieser Gruppen ($B_{ungeradzahlig}$) ohne Verwendung von Restvorhersage zwischen den Schichten kodiert werden.

6. Verfahren zum Kodieren von Videodaten, die wenigstens zwei räumlich skalierbare Schichten und kodierte Hochpass-Vollbilder und Tiefpass-Vollbilder enthalten, derart wie sie durch zeitliche Wavelet-Zerlegung unter Verwendung von bewegungskompensierter zeitlicher Filterung (MCTF) erhalten werden, wobei die Kodierung wenigstens der Hochpass-Vollbilder auf inversen Vorhersage- und inversen Aktualisierungs-Schritten beruht, umfassend die Schritte:

- Dekodieren der Tiefpass-Vollbilder gemäß ihrer Kodier-Betriebsart;

- Bestimmen aus der sequentiellen Reihenfolge eines Hochpass-Vollbildes, ob es zu einer ersten oder einer zweiten ($B_{geradzahlig}$, $B_{ungeradzahlig}$) Gruppe von Vollbildern gehört;
- Dekodieren des Hochpass-Vollbildes, wobei, wenn das Hochpass-Vollbild zu der ersten Gruppe von Vollbildern ($B_{geradzahlig}$) gehört, die Dekodierung Vorhersage desjenigen Restes zwischen den Schichten auf Vollbildebene verwendet, der für die inversen Vorhersage- und die inversen Aktualisierungs-Schritte verwendet wird, und wenn das Hochpass-Vollbild zu der zweiten Gruppe von Vollbildern gehört, der Rest, der für die inversen Zwischenschicht-Vorhersage- und die inversen Aktualisierungs-Schritte verwendet wird, ohne Vorhersage erhalten wird.

7. Vorrichtung zum Kodieren von Videodaten in wenigstens zwei räumlich skalierbare Schichten, wobei die Videodaten Hochpass-Vollbilder und Tiefpass-Vollbilder enthalten, derart wie sie durch zeitliche Wavelet-Zerlegung unter Verwendung bewegungskompensierter zeitlicher Filterung (MCTF) erhalten werden, wobei die Kodierung auf Vorhersage- und Aktualisierungs-Schritten beruht, umfassend:

- Mittel zum Kodieren der Tiefpass-Vollbilder, wobei eine Restvorhersage zwischen den Schichten (ORP) auf Vollbildebene verwendet werden kann;
- Mittel zum Aufspalten der Hochpass-Vollbilder in zwei verschachtelte Vollbildgruppen;
- Mittel zum Kodieren der Vollbilder der ersten Vollbildgruppe, wobei Restvorhersage zwischen den Schichten auf Vollbildebene verwendet werden kann; und
- Mittel zum Kodieren der Vollbilder der zweiten Vollbildgruppe unter Verwendung eines Kodierverfahrens ohne Restvorhersage zwischen den Schichten (NoRP).

8. Vorrichtung nach dem vorhergehenden Anspruch, bei der die Mittel zum Kodieren der Vollbilder der zweiten Vollbildgruppe Mittel zum Ausführen einer Kodierbetriebsart-Auswahl umfassen.

9. Vorrichtung zum Dekodieren von Videodaten, die wenigstens zwei räumlich skalierbare Schichten, und kodierte Hochpass-Vollbilder und Tiefpass-Vollbilder enthalten, derart wie sie durch zeitliche Wavelet-Zerlegung unter Verwendung von bewegungskompensierter zeitlicher Filterung (MCTF) erhalten werden, wobei die Dekodierung von wenigstens den kodierten Hochpass-Vollbildern auf inverser Vorhersage auf Vollbildebene und inversen Aktualisierungs-Schritten beruht, umfassend:

- Mittel zum Dekodieren der Tiefpass-Vollbilder gemäß ihrer Kodierbetriebsart;
- Mittel, um aus der sequentiellen Reihenfolge eines Hochpass-Vollbildes zu bestimmen, ob es zu einer ersten oder zweiten ($B_{geradzahlig}$, $B_{ungeradzahlig}$) Gruppe von Vollbildern gehört;
- Mittel zum Dekodieren des Hochpass-Vollbildes, wobei, wenn das Hochpass-Vollbild zu der ersten Gruppe von Vollbildern ($B_{geradzahlig}$) gehört, das Mittel zum Dekodieren eine Vorhersage des Restes zwischen den Schichten auf Vollbildebene ausführt, der für die inversen Vorhersage- und die inversen Aktualisierungs-Schritte verwendet wird, und wenn das. Hochpass-Vollbild zu der zweiten Gruppe von Vollbildern gehört, der Rest, der für die inversen Vorhersage- und die inversen Aktualisierungs-schritte zwischen den Schichten verwendet wird, ohne Vorhersage zwischen den Schichten erhalten wird.

10. Verfahren oder Vorrichtung nach einem der vorhergehenden Ansprüche, bei dem bzw. bei der die Restvorhersage zwischen den Schichten für ein Vollbild der höheren Schicht (EL) eine Aufwärts-Abtastung (upsampling) des Restes des entsprechenden Vollbildes der niedrigeren Schicht (BL) umfasst.

Revendications

1. Procédé de codage de données vidéo en au moins deux couches extensibles spatialement contenant des trames passe-haut et des trames passe-bas comme si obtenues par une décomposition temporelle par ondelettes utilisant un filtrage temporel compensé en mouvement (MCTF), où le codage se base sur des étapes, de prédiction et de mise à jour, le procédé comprenant les étapes consistant à

- coder les trames passe-bas, où une prédiction résiduelle inter-couches (ORP) au niveau des trames peut être utilisée ;
- diviser les trames passe-haut en deux groupes de trames d'entrelacement ;
- coder les trames du premier desdits groupes de trames, où une prédiction résiduelle inter-couches au niveau des trames peut être utilisée ; et
- coder les trames du deuxième des groupes de trames à l'aide d'un procédé de codage sans prédiction résiduelle inter-couches (NoRP).

2. Procédé selon la revendication 1, où l'étape de codage des trames du premier des groupes de trames comprend la sélection d'un parmi une pluralité de modes de codage, où au moins un des modes de codage possibles utilise une prédiction résiduelle inter-couches au niveau des trames (SRP), mais

aucun des modes de codage possibles utilise une prédiction résiduelle inter-couches au niveau des macroblocs.

3. Procédé selon la revendication 1 ou 2, où le premier groupe de trames comprend les trames passe-haut paires et le deuxième groupe de trames comprend les trames passe-haut impaires.

4. Procédé selon l'une quelconque des revendications 1 à 3, où l'étape de codage des trames du deuxième des groupes de trames comprend une sélection du mode de codage.

5. Signal vidéo comprenant au moins deux couches extensibles spatialement (BL,EL), où la couche supérieure (EL) contient des données de trame passe-bas codées (P1,P2) et des données de trame passe-haut codées (B1.....B15) comme si obtenues par une décomposition temporelle par ondelettes utilisant un filtrage temporel compensé en mouvement (MCTF), et où les données de trame passe-haut codées contiennent une indication du mode de codage et peuvent être divisées en deux groupes, la division se basant sur l'ordre séquentiel des trames, avec les données de trame d'un de ces groupes (B_{even}) contenant une indication indiquant si la trame a été codée à l'aide d'une prédiction résiduelle inter-couches et si au moins quelques-unes des trames ont été codées à l'aide d'une prédiction résiduelle inter-couches au niveau des trames mais non au niveau des macroblocs, et si les données de trame passe-haut de l'autre de ces groupes (B_{odd}) ont été codées sans utiliser de prédiction résiduelle inter-couches.

6. Procédé de décodage de données vidéo contenant au moins deux couches extensibles spatialement, ainsi que des trames passe-haut et des trames passe-bas codées comme si obtenues par une décomposition temporelle par ondelettes utilisant un filtrage temporel compensé en mouvement (MCTF), où le décodage d'au moins les trames passe-haut se base sur des étapes de prédiction inverse et de mise à jour inverse, le procédé comprenant les étapes consistant à

- décoder les trames passe-bas en fonction de leur mode de codage ;
- déterminer, à partir de l'ordre séquentiel d'une trame passe-haut, si elle appartient à un premier ou un deuxième (B_{even}, B_{odd}) groupe de trames ;
- décoder la trame passe-haut, où si la trame passe-haut appartient au premier groupe de trames (B_{even}), le décodage utilise une prédiction inter-couches au niveau des trames du résidu utilisé pour les étapes de prédiction inverse et de mise à jour inverse, et si la trame passe-haut appartient au deuxième groupe de trames, le résidu utilisé pour les étapes de prédiction inter-couches inverse et de mise à jour inverse est obtenu sans prédiction.

7. Appareil de codage de données vidéo en au moins deux couches extensibles spatialement, les données vidéo contenant des trames passe-haut et des trames passe-bas comme si obtenues par une décomposition temporelle par ondelettes utilisant un filtrage temporel compensé en mouvement (MCTF), où le codage se base sur des étapes de prédiction et de mise à jour, comprenant

- moyen permettant de coder les trames passe-bas, où une prédiction résiduelle inter-couches (ORP) au niveau des trames peut être utilisée ;
- moyen permettant de diviser les trames passe-haut en deux groupes de trames d'entrelacement ;
- moyen permettant de coder les trames du premier desdits groupes de trames, où une prédiction résiduelle inter-couches au niveau des trames peut être utilisée ; et
- moyen permettant de coder les trames du deuxième des groupes de trames à l'aide d'un procédé de codage sans prédiction résiduelle inter-couches (NoRP).

8. Appareil selon la revendication précédente, où le moyen permettant de coder les trames du deuxième des groupes de trames comprend un moyen permettant de réaliser la sélection du mode de codage.

9. Appareil de décodage de données vidéo, les données vidéo contenant au moins deux couches extensibles spatialement, ainsi que des trames passe-haut codées et des trames passe-bas codées comme si obtenues par une décomposition temporelle par ondelettes utilisant un filtrage temporel compensé en mouvement (MCTF), où le décodage d'au moins les trames passe-haut codées se base sur des étapes de prédiction inverse au niveau des trames et de mise à jour inverse, l'appareil comprenant

- moyen permettant de décoder les trames passe-bas en fonction de leur mode de cordage;
- moyen permettant de déterminer, à partir de l'ordre séquentiel d'une trame passe-haut, si elle appartient à un premier ou un deuxième (B_{evan}, B_{odd}) groupe de trames ;
- moyen permettant de décoder la trame passe-haut, où si la trame passe-haut appartient au premier groupe de trames (B_{even}), le moyen de décodage réalise une prédiction inter-couches au niveau des trames du résidu utilisé pour les étapes de prédiction inverse et de mise à jour inverse, et si la trame passe-haut appartient au deuxième groupe de trames, le résidu utilisé

pour les étapes de prédiction inverse et de mise à jour inverse est obtenu sans prédiction inter-couches.

10. Procédé ou appareil selon l'une quelconque des revendications précédentes, où une prédiction résiduelle inter-couches destinée à une trame de la couche supérieure (EL) comprend le sur-échantillonnage du résidu de la trame correspondante de la couche inférieure (BL).

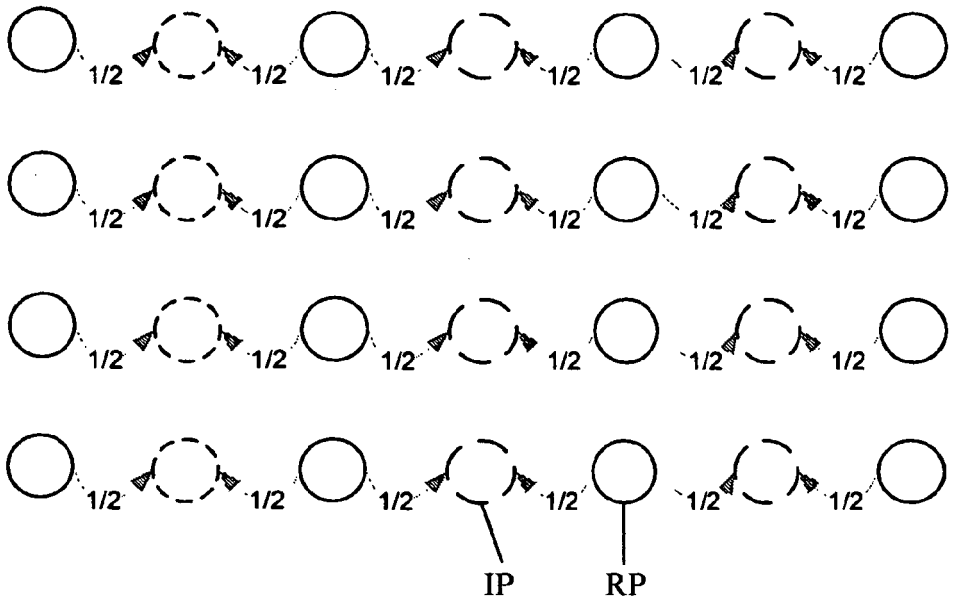
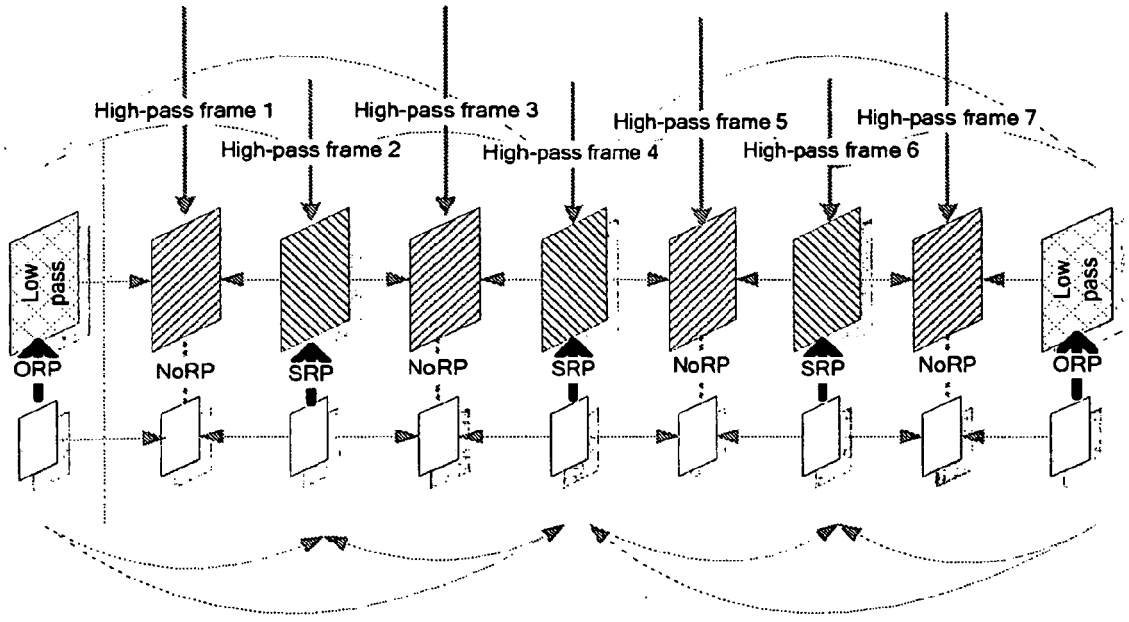


Fig.1



SRP: simplified residual prediction
ORP: original adaptive residual prediction
NoRP: No residual prediction

Fig.2

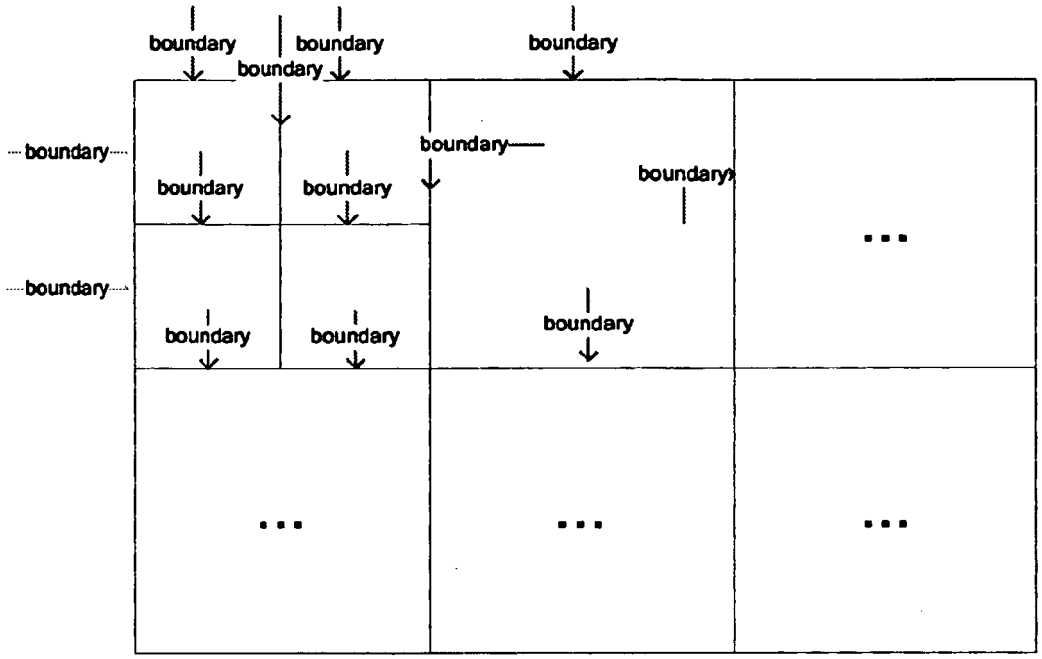

boundary
boundary
boundary
boundary
boundary
boundary
boundary
boundary
boundary
boundary
boundary
boundary
boundary
...
...
...
...

Fig.3

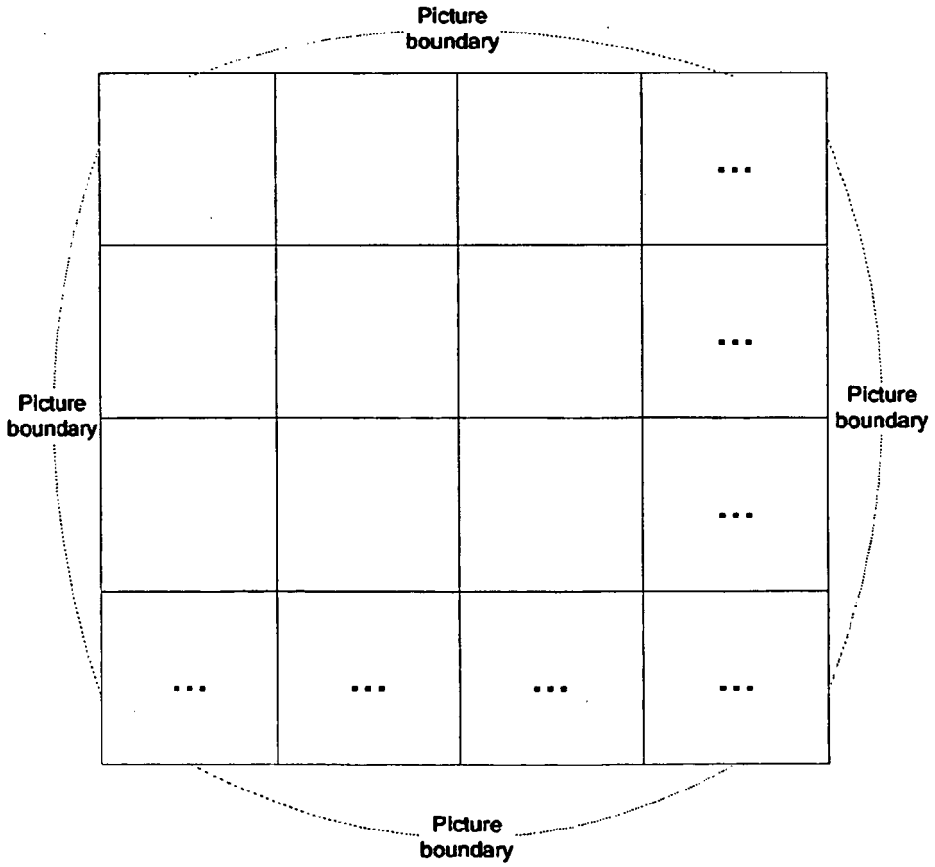

Picture boundary
Picture boundary
Picture boundary
Picture boundary
...
...
...
...
...
...
...

Fig.4

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **REICHEL J et al.** *Scalable Video Model 3.0'' ISO/IEC JTC1/SC29/WG11 N6716,* October 2004, 1-85 **[0009]**