

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 August 2010 (05.08.2010)

PCT

(10) International Publication Number
WO 2010/086041 A1

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

(21) International Application Number:
PCT/EP2009/065198

(22) International Filing Date:
16 November 2009 (16.11.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/148,409 30 January 2009 (30.01.2009) US

(71) Applicant (for all designated States except US): GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER [DE/DE]; Welfengarten 1, 30167 Hannover (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KLOMP, Sven [DE/DE]; Appelstraße 9a, 30167 Hannover (DE). OSTERMANN, Jörn [DE/DE]; Appelstraße 9a, 30167 Hannover (DE). MUNDERLOH, Marco [DE/DE]; Appelstraße 9a, 30167 Hannover (DE). VATIS, Yuri [DE/DE]; Appelstraße 9a, 30167 Hannover (DE).

(74) Agent: EISENFÜHR SPEISER & PARTNER; Postfach 10 60 78, 28060 Bremen (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: METHOD AND APPARATUS FOR CODING AND DECODING A VIDEO SIGNAL

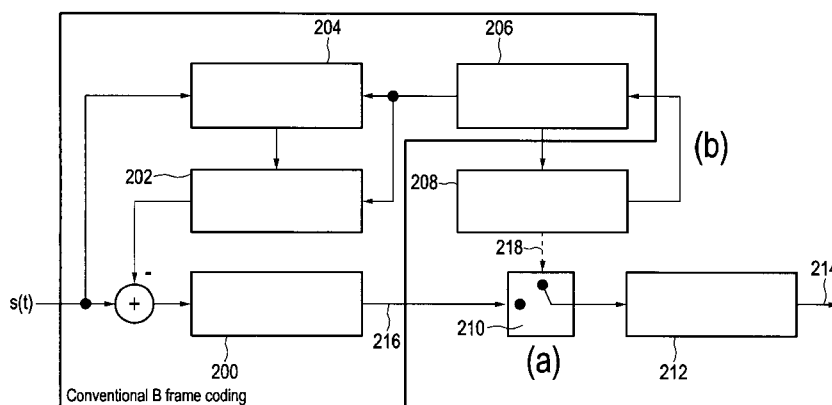


FIG. 2

(57) Abstract: A method for encoding a video signal, a method for decoding a video signal, a data signal, a coder for coding a video signal adapted to execute said encoding method, a decoder for decoding a video signal being coded by use of said encoding method adapted to execute said decoding method, and a computer program for coding and/or decoding a video signal adapted to execute said encoding method and/or adapted to execute said decoding method, when run on a computer, are provided. A decoder-side motion estimation (DSME) frame is generated at the encoder and/or at the decoder. The DSME frame can either be used as an additional prediction mode for motion-compensated prediction or as a pure DSME frame fully generated at the decoder.

WO 2010/086041 A1

Method and apparatus for coding and decoding a video signal

The present invention relates to a method for coding and decoding a video signal. It further relates to a data signal representing a coded video signal coded according to said method, a coder for coding a video signal, a decoder for decoding a video signal, and a computer program.

- 5 In current video coding solutions, such as MPEG-1, 2, 4 Video or ITU-T H.26x standards, the encoder estimates the motion between frames (P and B frames) and transmits the motion vectors and the residue to the decoder. Thus, temporal correlations between frames are exploited and compression is achieved. Due to block-based motion estimation, accurate compensation at object borders can only be achieved with small
- 10 block sizes. However, the smaller the block, the more motion vectors have to be transmitted, resulting in a discrepancy to bit rate reduction. Therefore, the block size and the corresponding motion vectors as well as the residue have a significant impact on compression performance. In H.264 / MPEG-4 AVC, the minimum block size is limited to 4x4 pixels.

An object of the present invention is to reduce at least one of the drawbacks indicated above, in particular to reduce drawbacks of block-based motion compensation it is at least an object of the present invention is to provide an alternative solution.

The present invention proposes a method for coding according to claim 1.

- 5 Accordingly, a decoder side motion estimation frame, in the following designated as DSME-frame, is used for coding a video signal. The basis for generating, i.e. calculating a DSME-frame is calculating one or a plurality of motion vectors defining the motion between the selected reference frames. The reference frames are decoded frames, each representing a frame of the video signal.
- 10 Preparing the DSME-frame as a reference frame can be performed by inserting the DSME frame in a reference picture buffer and thus the DSME frame can be selected among other reference frames in the reference picture buffer, to be selected for prediction and coding. If the DSME frame is selected for prediction, residuals can be calculated and used for coding, or the DSME frame can be selected as decoded frame, without
- 15 calculating any residuals or at least without using any calculated residuals. In any case, the DSME frame has been prepared for being used as a reference frame.

- Preferably, the method for coding a video signal using hybrid video coding comprises selecting several already coded and decoded reference frames (F_1, F_2, \dots) from the video signal as basis for the current frame (F_C), calculating one or a plurality of motion vectors
- 20 defining the motion between the reference frames (F_1, F_2, \dots) and the current frame, generating a decoder-side motion estimation (DSME) frame representing an estimation of the current frame based on said one or said plurality of motion vectors and the selected reference frames, inserting the DSME frame into a reference frame buffer of a video encoder comprising previously decoded frames, as an additional prediction signal for
- 25 predicting the current frame and/or providing a flag, indicating to use the DSME frame as decoded frame.

- Selecting the reference frames, which will also be designated as base frames, from the video signal for calculating the DSME frame can be an arbitrary choice as e.g. two frames adjacent to current frame to create an interpolation of said current frame. Other selections
- 30 are also possible, such as several frames from the past, i.e. preceding the current frame, to extrapolate the current frame.

A current frame is assumed to be that frame, on which the explained calculation is performed and for which a DSME frame is used or will be calculated. Any base frames preceding the current frame in a video signal are designated as previous frames, whereas base frames succeeding the current frame are designated as future frame.

5 According to one example, one selected reference frame is preceding the current frame and adjacent to it and another selected reference frame is succeeding the current frame and adjacent to it.

Calculating one motion vector defining the motion between the selected reference frames can be generally determined by any motion estimation algorithm operating on a frame as known in the art. If the selected reference frames are divided in blocks, common block-based motion estimation algorithms can be used. As a result, a plurality of motion vectors defining the motion between the selected reference frames are available.

10

Generating a decoder-side motion estimation (DSME) frame representing an estimation of the current frame can be carried out either using the said one or said plurality of motion vectors. Preferably, a set of motion vectors can be used to select the most appropriate motion vector for each pixel, block or frame.

15

Example:

Assuming the base frames are the two adjacent frames of the current frame, the motion between the both base frames is chosen and linear motion is assumed, and the DSME frame can be interpolated using the pixel information from one or both base frames. The position of the DSME frame between the previous and future base frame can be adjusted using weighting factors if linear motion does not occur. The DSME frame generated can now be used as a prediction signal for predicting the current frame. For that, the DSME frame is inserted into the reference picture buffer among other decoded frames to serve as an additional prediction signal to provide an additional prediction mode for the current frame.

20

25

In other words, the DSME frame inserted in the reference picture buffer is one possible frame to base the coding of the current frame on. I.e. if the DSME frame in the reference picture buffer is selected to be used for coding the current frame, the DSME frame is used as prediction for the current frame and a residual with respect to this DSME frame which might include difference vectors are calculated. These residuals are prepared for transmission and transmitted to the decoder, along with information, that the chosen

30

reference frame is the DSME frame. This information indicating, that the DSME frame is the current reference frame, can for example just be the information about the position of the DSME frame in the reference picture buffer, and that the reference frame of said position in the reference picture buffer is to be used. Accordingly, using the DSME frame as reference frame is identified by the selected position in the reference picture buffer. For this example the information which position in the reference picture buffer is used for the DSME frame can be submitted to the decoder separately and it would not be necessary to transmit this information for each frame.

E.g. it is once submitted, that the DSME frame will always be at position number 3 in the reference picture buffer and if the DSME frame is used for coding, it is only submitted - as side information - that the frame in position number 3 of the reference picture buffer was used by the coder and thus is to be used by the decoder.

Accordingly the decoder would - for this example - identify that position number 3 comprises the DSME frame and thus the DSME frame is used as the reference frame and would generate a DSME frame in the same manner as in the coder and would thus generate the same DSME frame as in the coder. Based on this DSME frame the current frame would be decoded.

Preferably, it is selected whether
to use the DSME frame as a reference frame and calculating one or a plurality of residuals, or
to use the DSME frame as a decoded frame without calculating any residuals, which is referred to as pure DSME coding. If pure DSME-coding is used, one possibility is to provide a pure-DSME-flag, indicating to the decoder that the DSME frame is used as the decoded frame without calculating any residuals.

Assuming that a DSME frame can be a quite adequate prediction, any corresponding residue could also be small. In this case it can be decided to not calculate and/or transmit the corresponding residue and thus just to use the corresponding DSME frame as the decoded frame. This is called in the present application pure DSME coding.

Pure DSME coding includes the case, when a calculated residue is zero. In case of pure DSME coding, information is provided in particular for being transmitted to the decoder, indicating, that for the current frame no residue and difference vectors will be transmitted

and thus indicating a pure DSME frame coding. This information can be indicated by a pure-DSME-flag, indicating the pure DSME-coding. Of course, a data bit corresponding to such flag can always be transmitted, whereas in case of pure-DSME-coding a 1 and in case of not pure DSME-coding a 0 is transmitted.

- 5 Alternatively, prediction and transmitting a residual of the current frame can be fully skipped, when the DSME frame can be used instead. In that case, a pure DSME slice flag, which means pure decoder-side motion estimation flag for one slice, is provided which indicates to perform the estimation of the DSME frame at the decoder. There is no need to transmit any further information, such as motion vectors or residue. The
10 calculation of the motion vectors is determined at the decoder using the decoded base frames in the same manner as explained above and a corresponding DSME frame is generated to be placed in the decoded video signal decoded frame or slice.

According to one embodiment, the method comprises the step of selecting whether the DSME frame is used as the prediction signal or a flag is provided indicating to use the
15 estimation of the DSME frame as decoded frame. Therefore, a hybrid approach is used where the encoder is deciding either to send a prediction residue or just to signal that the estimation of the DSME frame shall be performed at the decoder, which means that no additional information like prediction error or motion estimation parameters are sent to the decoder.

- 20 Pure DSME coding can be selected, when the corresponding residual is zero. However pure DSME coding can also be selected when the residual is not zero, but small. For this latter case, it must be decided whether a small decrease in quality of the current frame can be accepted with respect to the advantage of reducing the coding costs, i.e. less data has to be transmitted. Accordingly searching for a rate-distortion optimized decision is
25 proposed, in order to find the best compromise between a good video quality and low coding, decoding and/or data transmission costs. The rate-distortion optimized decision implemented in the reference encoder, which might be based on Lagrangian optimization, can be used to select the best mode.

If said corresponding flag is provided, the decoder estimates a DSME frame, which is
30 similar or identical to the DSME frame estimated at the encoder for representing the current frame, i.e. the encoder has decided that the estimated DSME frame is sufficient to

represent the current frame and thus the decoder can use the DSME frame as it is i.e. as the decoder has estimated it.

For motion estimation, either rectangular blocks or arbitrary shaped patches may be used. Furthermore, analytical approaches like optical flow based motion estimation methods can be used. Subsequently, if rectangular blocks are selected well-known block-based motion estimation such as full search, three step search, diamond search, enhanced predictive zonal search, hexagonal search may be performed to estimate the motion between the base frames. Advantageously, three consecutive frames of the video signal are selected as the selected reference frames and the current frame, whereby the current frame is between the selected reference frames, i.e. between the base frames.

According to one embodiment the reference frames are selected in dependence on the difference of the DSME frame and the current frame. Thus, an adaptive method for calculating the DSME frame is proposed. The selection can be performed by starting with one arbitrary choice of two reference frames, calculating a DSME frame and rating the quality of the calculated DSME frame by comparing the DSME frame with the current frame. Subsequently, other reference frames are selected and the calculating of the DSME frame and the rating of the result is repeated as well. This way, different DSME frames are calculated and rated and the DSME frame with the best rating is selected, i.e. the reference frames used for said DSME frame are selected.

If an adaptive method for selecting the reference frames i.e. for selecting the base frames is used, information on which reference frames were selected could be transmitted to the decoder as well. In this case, the decoder is adapted to evaluate such information in order to determine the reference frames to be selected. The selection of the reference frames can also change for every current frame.

According to an aspect of the present invention, a data signal represents a coded video signal coded according to a method described above and is used for storing and/or transmitting to a decoder. According to one embodiment, the data signal comprises a DSME flag, which indicates whether to calculate a DSME frame for the current frame or sequence or not. The DSME frame is, alternatively or additionally, also included in a reference list of decoded frames.

In a further embodiment, a position syntax element, that can be denoted position_in_reference_list syntax element, indicates position at which the current DSME frame is inserted in the reference list. Furthermore, a motion estimation syntax element, which can be denoted as motion_estimation_algorithm syntax element, indicates, which
5 one of the several motion estimation algorithms is used.

According to a further embodiment the data signal comprises a slice header for a coded slice, wherein the position_in_reference_list syntax element and/or the motion_estimation_algorithm syntax element are included. Depending on the selected motion estimation algorithm, additional data is added to the motion_estimation_algorithm
10 syntax element, such as maximum search range or weighting factor (w_{occ}).

According to one embodiment for every interprediction slice, in particular for every B-slice either using a modified reference list for generating the prediction signal or using a pure DSME slice is indicated by a DSME slice syntax element, also denoted as pure_dsme_slice syntax element, is provided in the data signal.

15 It is further proposed to provide a slice data in the data signal containing all defined standardized video coding data if the current slice is not a pure DSME slice. Otherwise, the corresponding slice data might be empty.

According to a further aspect, the data signal may comprise a macroblock flag, also denoted as dsme_mb_flag, which indicates the use of decoder-side motion estimation on
20 a macroblock level. The corresponding DSME macroblock type is incorporated into a macroblock prediction syntax.

According to the present invention, a method is provided for decoding a video signal using hybrid coding comprising decoding coded video data, wherein the coded video data has been encoded with a method described above, wherein decoder-side motion
25 estimation (DSME) is performed in accordance with the coding mechanism used for coding the video signal data, and/or wherein the coded video data is decoded depending on the information provided in the coded video data signal described above.

Preferably, a method for decoding comprises the step of evaluating if for the current frame a decoded frame or a DSME-frame is to be used as the reference frame. It also

comprises the step of generating the corresponding DSME-frame, if the DSME frame is to be used as the reference frame.

Preferably, for decoding the decoder receives the information which reference frame to use for decoding by indicating the corresponding position in the reference picture buffer.

5 According to this position the method can evaluate whether the reference frame is a DSME frame or not, as - according to this example - the reference picture buffer contains a DSME frame only at one position. This position is known by the decoder, either in general or at least for a series of some frames. Usually it only makes sense to generate a corresponding DSME frame, if it has to be used for the current frame according to the
10 information provided by the coder, as a DSME frame is usually only an adequate reference for the current frame.

According to a preferred embodiment, it is proposed, that the method for decoding comprises the step of evaluating if the DSME frame is used as the decoded frame, i.e. if pure DSME-coding is used for the current frame. In this case, the DSME frame has to be
15 generated as explained above and is then just used as the decoded frame without using any residual.

According to the present invention, a coder for coding a video signal adapted to execute the method described above is used. In particular, the coder comprises a motion-compensator for reducing the temporal redundancy by block-based motion compensated
20 prediction to provide a prediction error signal, a transformer and quantizer for transforming and quantizing the prediction error signal as well as means for inverse transforming and for dequantizing, a storage for storing reference pictures for motion compensated prediction, an estimation block for estimating a decoder-side motion estimation (DSME) frame by use of the reference pictures stored, a switch for switching
25 between using the DSME frame as an additional prediction signal for motion compensated prediction and a flag generating block for generating a flag indicating using the DSME frame as a prediction signal i.e. as a reference frame or using the DSME frame as a decoded signal, and/or an entropy coder for encoding all data in an entropy-constrained manner. These elements or a part of them can also be realized and/or
30 combined in a data processing unit, in particular in a microprocessor.

Preferably, a Lagrangian rate-distortion optimizer implements the evaluating or selecting for switching between using the DSME frame as an additional prediction signal for motion

compensated prediction and providing a flag indicating to use the DSME frame as decoded frame. The estimation block for generating the DSME frame implements DSME frame generation method as described above. For all other elements of the coder standard techniques or elements can be used.

5 According to the present invention, a decoder for decoding a video signal being coded by use of the coding method described above is proposed. The decoder is adapted to execute the decoding method described above, in particular comprising an entropy decoder for decoding the entropy-constrained coded data signal, an inverse quantizer and inverse transformer for inverse quantization and backward transformation, a storage
10 for storing decoded reference pictures, an estimation block for estimating a decoder-side motion estimation (DSME) frame by use of the decoded reference pictures stored and/or an evaluation block for evaluating a flag indicating using the DSME frame as a prediction signal i.e. as a reference frame or using the DSME frame as a decoded signal.

15 According to the present invention, a computer program is proposed for coding and/or decoding a video signal adapted to execute the coding method described above and/or adapted to execute the decoding method described above, when run on a computer.

Embodiments of the present invention will now be described, by way of example, with reference to the accompanying drawings.

20 Figure 1a and 1b provide a schematic illustration of estimating a motion vector for generating a DSME frame including refinement.

Figure 2 is a block diagram of an encoder embodiment according to the present invention.

Figure 3 is a diagram showing an example of the amount of B frames coded as pure DSME frame for various quantization parameters QP.

25 Figure 4 is a diagram showing an example of the difference of the standard coder and a coder using DSME frames bit rate for various positions of the DSME frame within the reference list.

Figure 5a – 5c provide a schematic illustration of motion compensation modes for macroblock prediction.

Figure 6 is a schematic illustration of correct compensation of accelerated motion using a factor $\omega_{acc} \in [0,1]$ according to the present invention.

5 Fig. 1 shows a schematic illustration of estimating a motion vector for generating a DSME-frame including a refinement. Fig. 1a) shows a block of a first selected reference frame 100 also referred to as reference frame 1, a DSME frame to be generated 102 and a second selected reference frame 104, also referred to as reference frame 2. Furthermore, a candidate motion vector 108 showing the estimated motion from the first
10 to the second selected reference frame but not being selected for generating a DSME-frame and a selected motion vector 106 are depicted.

As explained above, the rate-distortion performance can be improved by performing motion estimation at the decoder. The decoder estimates the motion with the aid of some reference frames and interpolates or extrapolates the current frame to be coded using
15 these motion vectors. In conventional motion estimation schemes, the motion vectors are selected by minimizing the prediction error between the current frame and a reference frame. Therefore, it might occur that the motion estimation algorithm finds motion vectors that produce the smallest residue but do not represent the true motion. Since this DSME example assumes constant motion to predict intermediate frames, those wrong motion
20 vectors would induce high interpolation errors. Therefore, the motion estimation algorithm was to be redesigned.

One possible algorithm for estimation of the motion at the decoder is described in the following. It is assumed that two reference frames are available to predict the intermediate frame.

25 First, a full-search block matching algorithm estimates the motion vectors between the two reference frames with full-pel accuracy. Since this vector field will result in overlapped and uncovered areas after frame interpolation, the following motion estimation scheme can be used: For each 16x16 block of the DSME frame, a vector is selected from the previously estimated candidates that intercept the DSME frame closest to the centre of
30 the block illustrated by motion vector 106 in Fig. 1a. This motion vector is used as the initial value for the bidirectional motion estimation in which the motion vector is refined in

sup-pel accuracy with a smaller search range. Since linear and constant motion is assumed between the reference frames, the forward and backward motion vectors are symmetrical as illustrated in Fig. 1b. In the last step, the motion vector field is smoothed by using weighted vector median (WVM) filters in order to detect and/or remove outliers.

5 Finally, the DSME frame is predicted with bilinear interpolation using the motion vector field. The same motion vectors can be used for the luminance and chrominance components. Further explanations of said standardized elements according to the state of the art can be found in: Ascenso et al., "Improving frame interpolation with spatial motion smoothing for pixel domain distributed video coding", 5th EURASIP, Slovak Republic, Juli
10 2005. However, in that architecture, the motion estimation algorithm is entirely used at the decoder and not implemented in the encoder.

Fig. 2 shows a simplified block diagram of an encoder embodiment according to the present invention. Blocks 200, 202, 204, 206 and 212 represent standardized elements of a state-of-the-art encoder environment. The transformer and the quantizer are included in
15 block 200. Block 202 and 204 represent the motion compensation 202 and the motion estimation 204 unit. The reference picture buffer - which is sometimes also designated as reference picture list or reference frame buffer - is depicted in block 206. According to the invention, the generation of a DSME frame is performed at the integrated decoder in an encoder environment and at the decoder. Block 208 represents the generation of a
20 DSME frame at the integrated decoder of the encoder environment. The DSME frame is used in two different approaches or modes as depicted in Fig. 2: (a) pure DSME frame coding and (b) reference frame insertion.

Experiments have shown that for low bit rates, it is better to use the DSME frame as decoded picture, without coding the remaining residual. Therefore, a hybrid approach is
25 used where the encoder is deciding either to send the whole frame as a B frame 216, or just to signal to use the DSME frame 218 according to position (a) at block 210. Block 210 represents the decider. In that case, the frame is called pure DSME frame, since no additional information like prediction error or motion estimation parameters are sent to the decoder in the current implementation as illustrated. The rate-distortion optimized
30 decision implemented in the reference H.264 / MPEG-4 AVC encoder can be used to select the mode with minimum Lagrangian cost.

The approach involving reference frame insertion allows the coder to use the DSME frame as reference for each macroblock. The DSME frame is fed into the reference list of the coder as shown in Fig. 2 according to position (b) at block 210. As the DSME frame is a prediction for the current frame to be encoded, the residual is smaller in many cases and thus, fewer bits have to be transmitted. Furthermore, the bit rate for transmitting the motion vector differences can also be reduced, since the motion vector predictor can assume that no motion occurred. Since coder like H.264 / MPEG-4 AVC signal the index of the selected reference with different code word sizes, coding gain is dependent on the position of the DSME frame in the reference lists as it can be seen in Fig. 4.

10 Finally, block 212 represents a standard entropy coder and the data signal 214 is transmitted to the decoder.

Fig. 3 shows the amount of DSME frames in dependence of the selected quantization parameter QP for different, generally known test sequences, i.e. how many of the calculated DSME-frames have, according to one embodiment, finally passed the criteria for being sufficient to be used as DSME frame in the decoder. Since, no prediction error is coded for pure DSME frames, the desired quality cannot be provided at higher bit rates. Thus, the encoder decides to transmit all frames as B frames with modified reference picture buffer in case of fine quantization indicated by lower quantization parameter.

20 Fig. 4 shows the bit rate reduction of a coder using DSME-frame compared to the H.264 / MPEG-4 AVC reference encoder, which is the generally known JVT reference software JM for the different positions of the DSME frame within the reference list. In the example, five positions are possible, as illustrated in bars P1 – P5 indicating the bit rate reduction when the DSME frame is inserted in position 1, position 2, position 3, position 4, or position 5, designated by P1, P2, P3, P4 or P5, respectively.

Using high quantization parameters, the rate reduction is independent of the position since all frames are encoded as pure DSME frames as mentioned above and thus, the reference lists are not used. For higher qualities, the position becomes more important.

30 If the DSME frame is inserted in front of all other reference pictures at position 1, the bit rate savings are low. This is due the fact that the encoder often selects blocks of the

temporal adjacent frame as reference. If it is moved to the second position in the list, the encoder needs more bits in signaling it to the decoder.

If inserted at position 5, the DSME frame replaces the reference frame directly following the current frame. Since that frame is often used as reference, the DSME approach is worse than the H.264 reference. Evaluations with several sequences have shown that inserting the DSME frame at the second position gives the best overall results. However, it should be configurable within the bit stream.

Fig. 5 illustrates a schematic drawing of possible motion compensation using the motion vectors available at a macroblock level. A first selected reference frame 500 comprising a macroblock 506, a second selected reference frame 504, and a DSME frame 502 are shown. The macroblock 506 and corresponding motion vectors 508, 510, and 512 illustrate the possible motion compensation modes, i.e. forward, backward and bidirectional prediction. However, motion compensation of the current macroblock 506 is not limited to these three modes. The use of a DSME macroblock type is described in the following in more detail.

An additional flag named "dsme_mb_flag" in the picture parameter set raw byte sequence payload or in the slice header can be used to allow the DSME type within the current sequence or slice. Furthermore, the new type has to be incorporate into the macroblock prediction syntax designated as "mb_pred()". Since DSME does not need motion vectors to be transmitted, the vectors can be removed for this macroblock type. Additional data for this macroblock type can be the information, how the macroblock is predicted. Due to occlusion some parts are not visible in all frames. Thus, the motion compensated pixel values of the previous frame (Fig. 5 (a)), of the next frame (Fig. 5 (b)), or of both frames (Fig. 5 (c)) can be used to predict the current macroblock. However, not only these three discrete modes are possible. A weighting factor w_{occ} can be used to compensate occlusion as illustrated in equation (1):

$$I_{DSME} = w_{occ}I_1 + (1 - w_{occ}) I_2 \text{ with } 0 \leq w_{occ} \leq 1 \quad (1)$$

where I_1 and I_2 are the motion compensated pixel values of the two reference frames and I_{DSME} the resulting prediction of the current macroblock. Thus, a weighting factor of $w_{occ} = 0.5$ corresponds to Fig. 5 (c), $w_{occ} = 1$ and $w_{occ} = 0$ match Fig. 5 (a) and (b), respectively.

Fig. 6 shows a drawing illustrating the case of non-constant motion on a macroblock level. It shows the first selected reference frame 600, the second selected reference frame 604 and the current/DSME frame 602. Furthermore, a macroblock 606 is depicted and the corresponding real motion vectors 616 and 610. An estimated motion vector 608 is also shown.

In some cases the assumption of constant motion is not fulfilled for each macroblock. An example is given in Fig. 6, where an object is moving slowly in the first half but faster in the second half. If the estimated motion is used for motion compensation, the object occurs at the wrong position in the DSME frame indicated by macroblock 606. To avoid this, the macroblock has to be virtually moved left indicated by macroblock 612 for motion compensation. Therefore, a factor $w_{acc} \in [0, 1]$, i.e. w_{acc} may take any value from 0 to 1, which represents the virtual position of the current macroblock as depicted in Fig. 6, is used for correction and should be considered and transmitted to the decoder to compensate accelerated or decelerated motion.

A modified syntax and semantics of the bit stream is needed for decoder-side motion estimation based on the H.264 / MPEG-4 AVC standard. However, the syntax elements are applicable to any video coder using DSME. Thus, an embodiment according to the present invention is described by use of the accompanying Tables.

Table 1 shows a general set raw byte sequence payload including a DSME flag syntax element and syntax.

Table 2 shows a general slice header including a pure DSME slice syntax element and syntax.

Table 3 shows a general slice data having DSME flag and pure DSME flag syntax elements and a syntax function according to one embodiment of the present invention.

Since the additional motion estimation increases the computational complexity at the encoder and decoder, a DSME flag is added in the picture parameter set raw byte sequence payload syntax to activate the DSME approach illustrated in Table 1.

This code indicates, whether the DSME scheme is applied for the current sequence (DSME flag = 1) or not (DSME flag = 0). If DSME is used, the position in the reference

list, where the DSME frame is added, is signaled with the flag "position_in_reference_list" to the decoder.

Furthermore, one of several motion estimation algorithms like block-, mesh-, or optic flow based can be selected with the flag "motion_estimation_algorithm". Dependent on this
5 flag, additional data can be transmitted. The block size as well as the matching algorithm such as full search, three step search, diamond search, enhanced predictive zonal search, and hexagonal search might be possible information needed for block-based motion estimation. Another flag can specify if either rectangular blocks or arbitrary shaped patches are used for motion estimation. The search range i.e. the maximum
10 length of a motion vector and the spatial resolution of motion vectors are also of interest and should be signaled. Since the precision can lie within sub-pel range, a filter is needed to calculate those sub-pel values required for motion compensation. Since the optimal filter depends on the sequence, the filter can also be defined in the bit stream.

For optic flow motion estimation, some additional parameters are of interest. The data
15 term, which is the main criteria to calculate the motion field, can be based on different models like constancy assumption on the luminance, the image derivatives, or multiple image features and has to be signaled to the decoder.

However, the data term might not be sufficient to compute a unique solution. Therefore, the smoothness term is used in addition to the data term to incorporate prior knowledge
20 on the motion field. Since the smoothness term depends on the sequence properties, it should be signaled as well.

Furthermore, multigrid methods are used to efficiently solve the motion field problem. A flag for the selected multigrid method such as unidirectional multigrid, unidirectional warping and bidirectional multigrid, should preferably be provided, since the methods
25 either speed up the computation or improve the quality of the result.

The optic flow method is often calculated iteratively and thus, the maximum number of iterations can also be appended to the additional data. It is also possible to add flags named "position_in_reference_list" and "motion_estimation_algorithm" with its additional data in "slice_header()" as illustrated in Table 2 to allow various parameters for each
30 slice.

The hybrid approach decides for every B slice either to use the modified reference list or the pure DSME slice. This is signaled by the flag "pure_dsme_slice". In this example it is assumed that the DSME approach is only used for B frames using motion compensated interpolation. However, DSME is also conceivable for other types like P frames where
 5 motion compensated interpolation or extrapolation can be used. Since no additional data is needed for the pure DSME slice, "slice_data()" is empty as illustrated in Table 3. However, the DSME algorithm needs to know which frames to use as base frames. This can be know a priori or "signaled" within the "base_frame_list".

This means that "slice_data()" contains all the data defined in the H.264 / MPEG-4 AVC
 10 standard if the current slice is not a pure DSME slice. Otherwise, "slice_data()" contains no data. With the previously defined syntax, it is only possible to send additional data either for each frame or for the whole sequence. To control the DSME approach also at macroblock level, a new macroblock type can be added to the existing types of B slice macroblocks. Again, the changes are only explained for B frames but are also applicable
 15 for other frames.

| seq_parameter_set_rbsp(){ | C | Descriptor |
|--|---|------------|
| : | | |
| dsme_flag | 1 | u(1) |
| if(dsme_flag){ | | |
| position_in_reference_list | 1 | u(8) |
| motion_estimation_algorithm | 1 | u(2) |
| if(motion_estimation_algorithm==0){ | | |
| <i>additional data</i> | | |
| } | | |
| else if(motion_estimation_algorithm==1){ | | |
| <i>additional data</i> | | |
| } | | |
| : | | |
| } | | |
| : | | |
| } | | |

Table 1

| Code | C | Descriptor |
|-------------------------------------|---|------------|
| slice_header(){ | | |
| : | | |
| if(dsme_flag && (slice_type == B)){ | | |
| pure_dsme_slice | 2 | u(1) |
| } | | |
| : | | |
| } | | |

Table 2

| Code | C | Descriptor |
|---|---|------------|
| slice_data(){ | | |
| if(!(dsme_flag && (slice_type==B) && pure_dsme)){ | | |
| ... | | |
| } | | |
| if(dsme_flag && (slice_type==B)){ | | |
| Base_frame_list | | |
| } | | |
| } | | |

Table 3

Claims

1. Method for coding a video signal, comprising a plurality of frames, comprising the
5 steps:
 - selecting at least two reference frames for coding a current frame,
 - calculating one or a plurality of motion vectors defining the motion between the selected reference frames,
 - generating a decoder-side motion estimation frame (DSME-frame) representing an
10 estimation of the current frame based on said one or said plurality of motion vectors and based on the selected reference frames,
 - prepare the DSME frame as a reference frame for coding and/or as a decoded frame.
- 15 2. Method according to claim 1, characterized by inserting the DSME frame into a reference picture buffer of a video encoder comprising previously decoded frames as an additional possible prediction signal for predicting the current frame.
3. Method according to claim 1 or 2, characterized in selecting whether
20 to use the DSME frame as a reference frame and calculating one or a plurality of residuals, or
to use the DSME frame as a decoded frame without calculating any residuals.
4. Method according to any of the claims 1 to 3, wherein block-based motion
25 estimation is performed to estimate the motion between the selected reference frames.
5. Method according to any of the claims 1 to 4, wherein the reference frames are selected in dependence on the difference of the DSME frame and the current frame.
6. Data signal representing a coded video signal coded according to a method of any of the preceding claims for storing and/or transmitting to a decoder.
- 30 7. Data signal according to claim 6, comprising a DSME flag indicating whether a DSME frame is to be calculated at the decoder for the current sequence or frame or not,

a position syntax element indicating the position of the DSME frame in a reference list, and/or

a motion estimation syntax element indicating which one of several motion estimation algorithms are used.

5 8. Data signal according to claim 7, comprising a slice header having the position syntax element and the motion estimation syntax element with its additional data depending on the selected motion estimation algorithm are added.

9. Data signal according to claim 8, comprising a DSME slice syntax element indicating for every inter prediction slice in particular for every B-slice either use of a
10 modified reference frame buffer for generating the prediction signal or a pure DSME slice.

10. Data signal according to any of claim 6 - 9, comprising a slice data containing all defined standardized video coding data if the current slice is not a pure DSME slice, otherwise contains no data.

11. Data signal according to any of claim 6 - 10, comprising a macroblock flag
15 indicating the use of a DSME macroblock type to be incorporated into macroblock prediction syntax.

12. Method for decoding a video signal using hybrid coding, comprising:

decoding coded video data, wherein the coded video data has been encoded with a method according to any of the claims 1 to 5,

20 performing decoder-side motion estimation (DSME) in accordance with the coding mechanism used for coding the video signal data, and/or wherein the coded video data is decoded depending on the information provided in the coded video data signal according to claims 6 to 11.

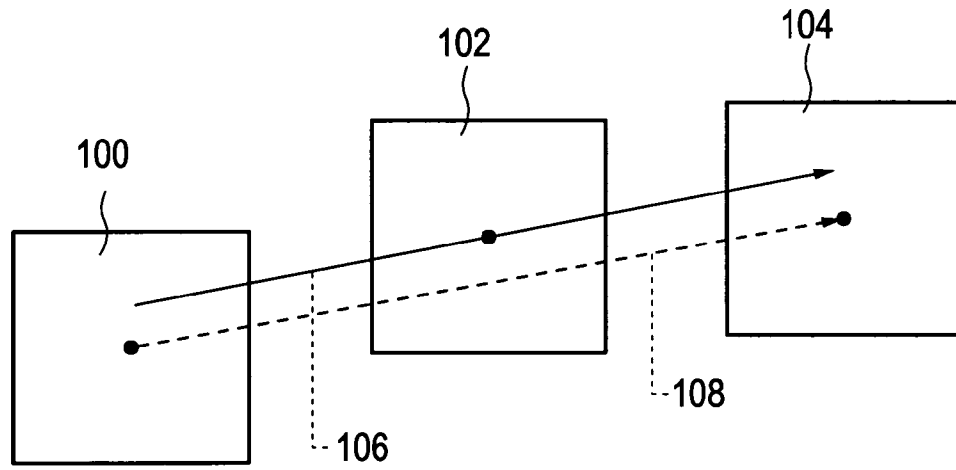
13. Method according to claim 12, comprising the steps of

25 evaluating whether for the current frame a decoded frame or a DSME frame is to be used as a reference frame, and

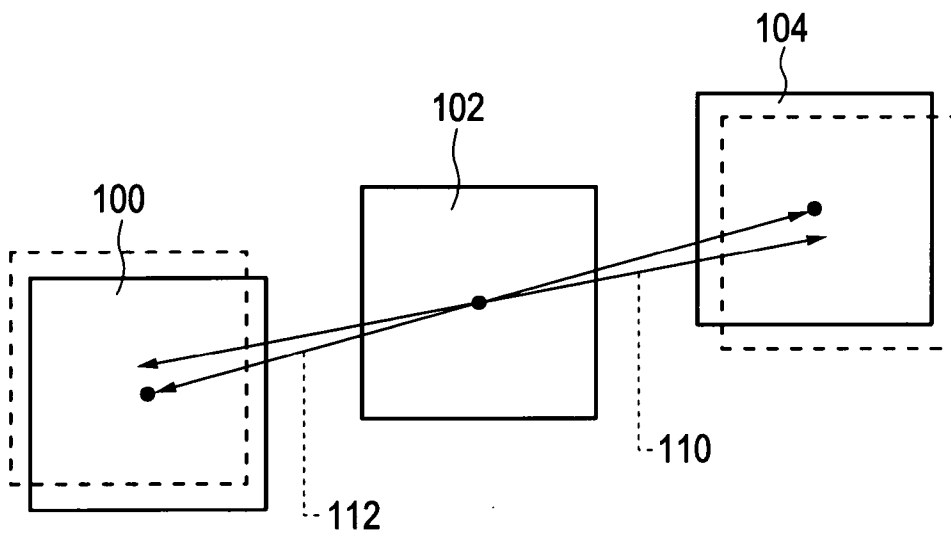
- generating the corresponding DSME frame, if the DSME frame is to be used as the reference frame, and

evaluating which reference frames were used for calculating the DSME frame.

14. Method according to claim 12 or 13, characterized by evaluating whether the DSME frame is used as the decoded frame or not.
15. Coder for coding a video signal adapted to execute a method according to any of claims 1 to 5, in particular comprising:
- a motion-compensator for reducing the temporal redundancy by block-based motion compensated prediction to provide a prediction error signal,
 - a transformer and quantizer for transforming and quantizing the prediction error signal,
 - 10 a storage for storing reference pictures for motion compensated prediction,
 - an estimation block for estimating a decoder-side motion estimation (DSME) frame by use of the reference pictures stored,
 - a switch for switching between using the DSME frame as an additional prediction signal for motion compensated prediction,
 - 15 a flag generating block for generating a flag indicating using the DSME frame as a prediction signal i.e. as a reference frame or using the DSME frame as a decoded signal, and/or
 - an entropy coder for encoding all data in an entropy-constrained manner.
- 20 16. Decoder for decoding a video signal being coded by use of a method according to any of claims 1 to 5 adapted to execute a method according to any of claims 12 to 14, in particular comprising:
- an entropy decoder for decoding the entropy-constrained coded data signal,
 - an inverse quantizer and inverse transformer for inverse quantization and
 - 25 backward transformation,
 - a storage for storing decoded reference pictures,
 - an estimation block for estimating a decoder-side motion estimation (DSME) frame by use of the decoded reference pictures stored and/or
 - an evaluation block for evaluating a flag indicating using the DSME frame as a
 - 30 prediction signal i.e. as a reference frame or using the DSME frame as a decoded signal.
17. Computer program for coding and/or decoding a video signal adapted to execute a method according to any of claims 1 to 5 and/or adapted to execute a method according to any of claims 12 to 14, when run on a computer.



(a)



(b)

FIG. 1

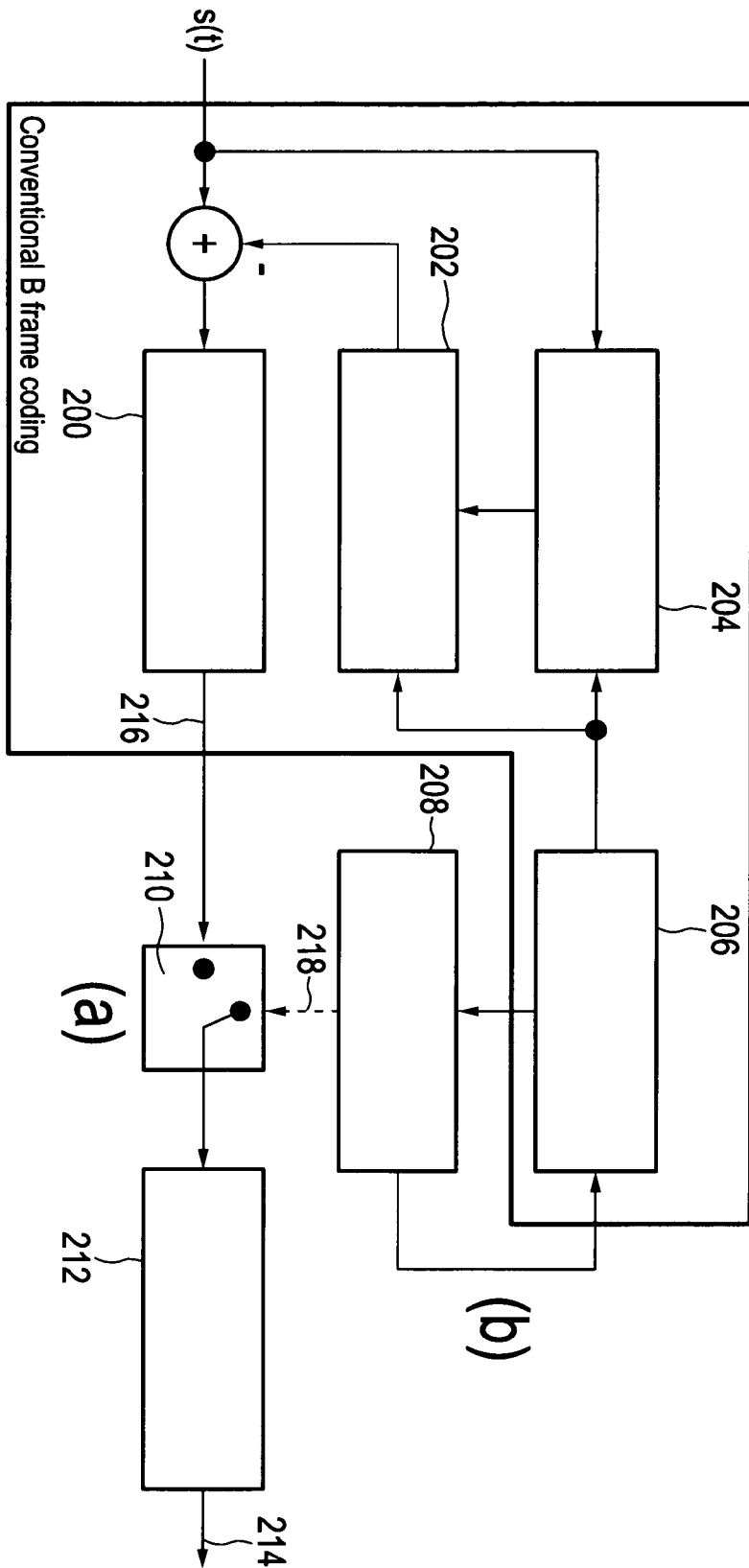


FIG. 2

FIG. 3

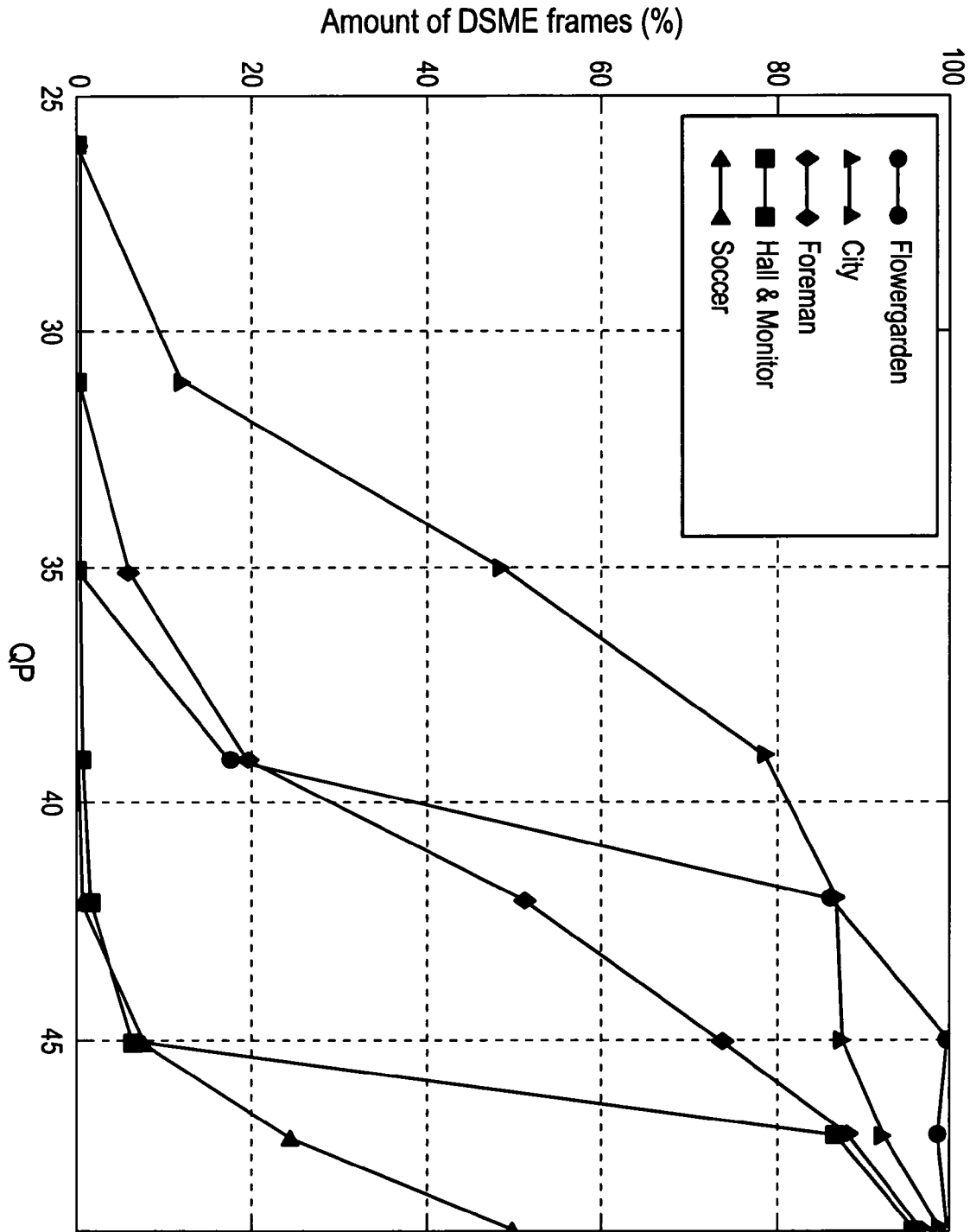
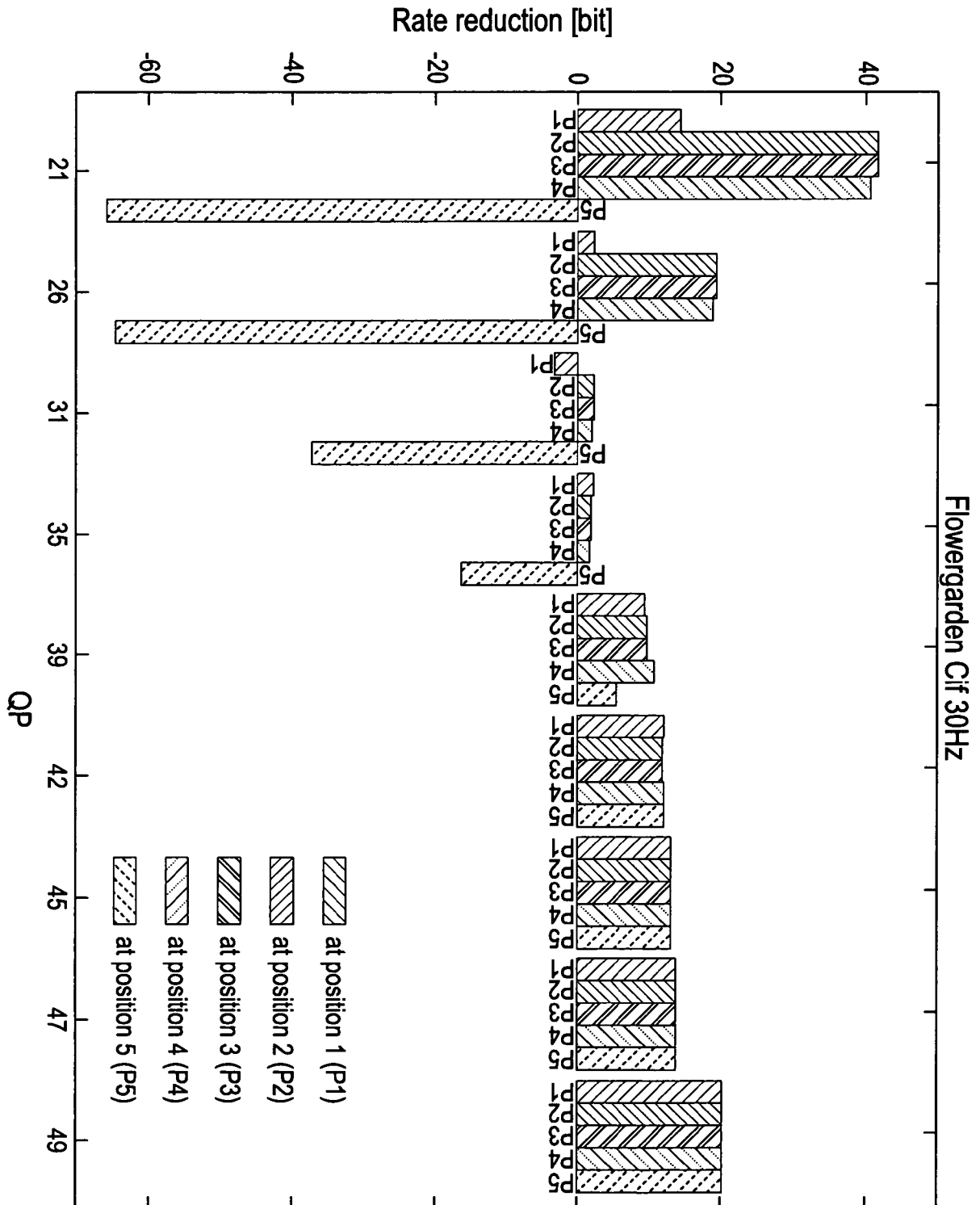
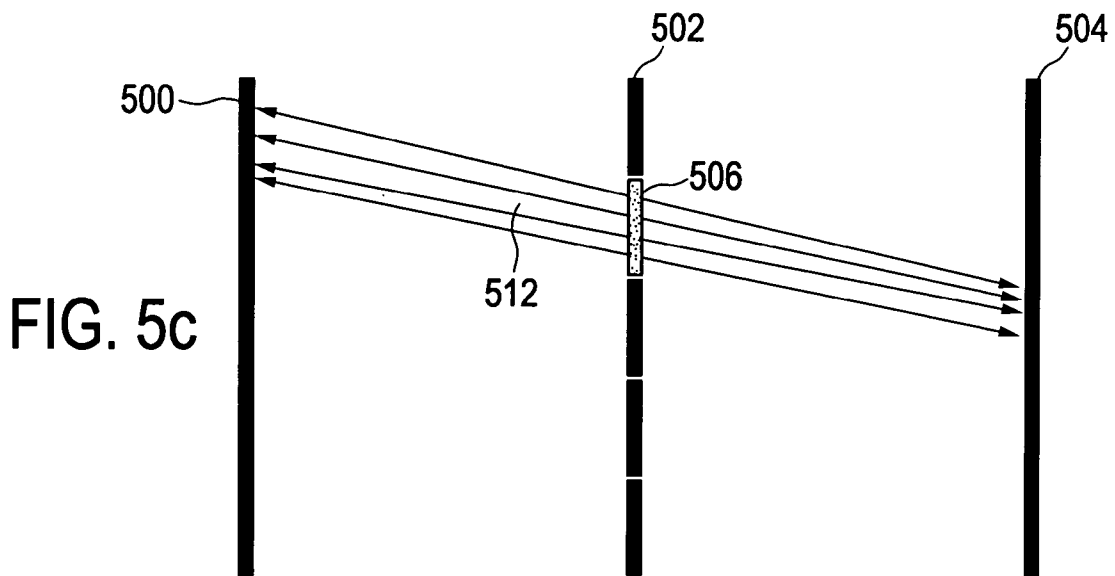
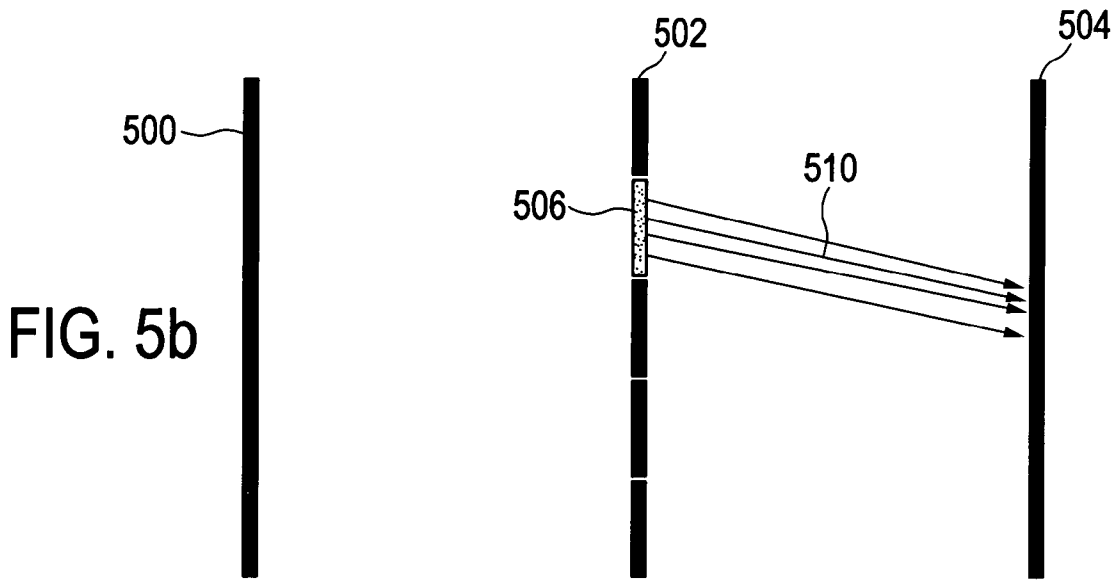
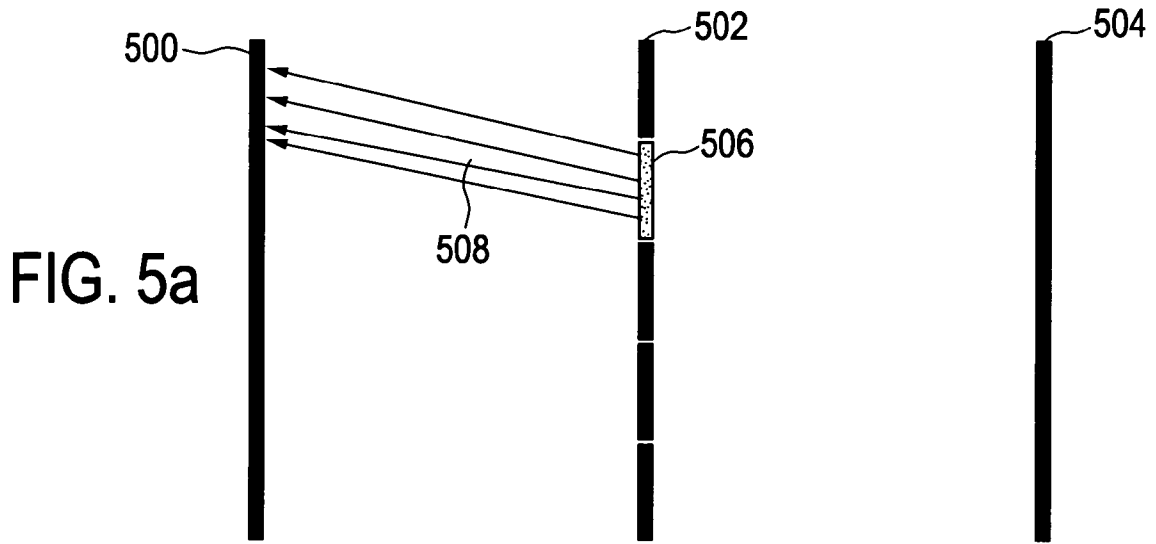


FIG. 4





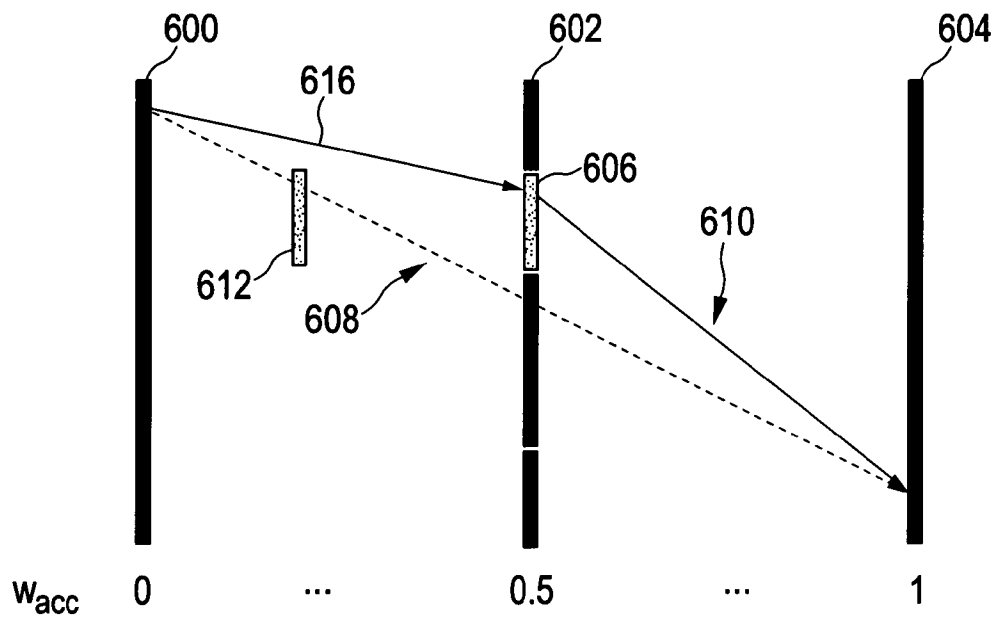


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/065198

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|-----------|---|-----------------------|
| X | EP 1 936 998 A2 (HITACHI LTD [JP]) 25 June 2008 (2008-06-25) figures 1, 2, 3, 4, 6a, 6b, 7a, 7b abstract paragraph [0020] - paragraph [0145] ----- -/-- | 1-17 |

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

18 February 2010

Date of mailing of the international search report

26/02/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040,
 Fax: (+31-70) 340-3016

Authorized officer

Cyranka, Oliver

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/065198

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|---|------------------------|
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X A | EP 1 827 025 A1 (TOSHIBA KK [JP]) 29 August 2007 (2007-08-29) figures 1, 2, 6, 10, 13, 14, 31, 32, 46 abstract paragraph [0009] - paragraph [0010] paragraph [0025] - paragraph [0030] paragraph [0039] paragraph [0044] paragraph [0068] paragraph [0071] paragraph [0082] paragraph [0085] paragraph [0088] - paragraph [0090] paragraph [0098] paragraph [0122] - paragraph [0129] paragraph [0155] paragraph [0180] - paragraph [0184] paragraph [0259] - paragraph [0272] | 1-8, 11-17 9-10 |
| X | DANE G ET AL: "Encoder-Assisted Adaptive Video Frame Interpolation" 2005 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (IEEE CAT. NO.05CH37625) IEEE PISCATAWAY, NJ, USA, IEEE, PISCATAWAY, NJ, vol. 2, 18 March 2005 (2005-03-18), pages 349-352, XP010790648 ISBN: 978-0-7803-8874-1 | 1,6,12, 15-17 |
| A | figures 1,2 pages II-349 - pages II-350 | 2-5, 7-11, 13-14 |
| A | "Frame Interpolation and Bidirectional Prediction of Video using Compactly Encoded optical-Flow Fields and Label Fields" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, vol. 9, no. 5, 1 August 1999 (1999-08-01), pages 713-726, XP11014592 DOI: 10.1109/76.780361 abstract page 714 - page 716 | 1-17 |
| | ----- -/-- ----- | |

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2009/065198

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | <p>AI-MEI HUANG ET AL: "A Multistage Motion Vector Processing Method for Motion-Compensated Frame Interpolation" IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 15, no. 5, 1 May 2008 (2008-05-01), pages 694-708, XP011225973 ISSN: 1057-7149 abstract page 694 - page 696 page 698, right-hand column - page 699, left-hand column</p> <p>-----</p> | 1-17 |
| A | <p>SU J K ET AL: "Motion-compensated interpolation of untransmitted frames in compressed video" SIGNALS, SYSTEMS AND COMPUTERS, 1996. CONFERENCE RECORD OF THE THIRTIE TH ASILOMAR CONFERENCE ON PACIFIC GROVE, CA, USA 3-6 NOV. 1996, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, vol. 1, 3 November 1996 (1996-11-03), pages 100-104, XP010231401 ISBN: 978-0-8186-7646-8 abstract page 100 - page 101</p> <p>-----</p> | 1-17 |
| X,P | <p>KLOMP S ET AL: "Decoder-side block motion estimation for H.264 / MPEG-4 AVC based video coding" CIRCUITS AND SYSTEMS, 2009. ISCAS 2009. IEEE INTERNATIONAL SYMPOSIUM ON, IEEE, PISCATAWAY, NJ, USA, 24 May 2009 (2009-05-24), pages 1641-1644, XP031479529 ISBN: 978-1-4244-3827-3 the whole document</p> <p>-----</p> | 1-17 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2009/065198

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|----------------------------|-----------------------------|
| EP 1936998 | A2 | 25-06-2008 | CN 101207819 A 25-06-2008 |
| | | | JP 2008154015 A 03-07-2008 |
| | | | US 2008159398 A1 03-07-2008 |
| | | | US 2008211968 A1 04-09-2008 |
| <hr/> | | | |
| EP 1827025 | A1 | 29-08-2007 | EP 1830576 A1 05-09-2007 |
| | | | EP 1830577 A1 05-09-2007 |
| | | | EP 1830578 A1 05-09-2007 |
| | | | EP 1833261 A1 12-09-2007 |
| | | | EP 1830579 A1 05-09-2007 |
| | | | EP 1833262 A1 12-09-2007 |
| | | | EP 1827026 A1 29-08-2007 |
| | | | EP 1827027 A1 29-08-2007 |
| | | | EP 1830580 A1 05-09-2007 |
| | | | EP 1827028 A1 29-08-2007 |
| | | | EP 1827029 A1 29-08-2007 |
| <hr/> | | | |