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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 October 2006 (05.10.2006)

PCT

(10) International Publication Number
WO 2006/104363 A1

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

(21) International Application Number:
PCT/KR2006/001196

(22) International Filing Date: 31 March 2006 (31.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/667,115	1 April 2005 (01.04.2005)	US
60/670,246	12 April 2005 (12.04.2005)	US
60/670,241	12 April 2005 (12.04.2005)	US
10-2005-0084744	12 September 2005 (12.09.2005)	KR
10-2005-0084742	12 September 2005 (12.09.2005)	KR
10-2005-0084729	12 September 2005 (12.09.2005)	KR

(71) Applicant *(for all designated States except US)*: **LG ELECTRONICS INC.** [KR/KR]; 20, Yoido-dong, Youngdungpo-gu, Seoul 150-010 (KR).

(72) Inventors; and

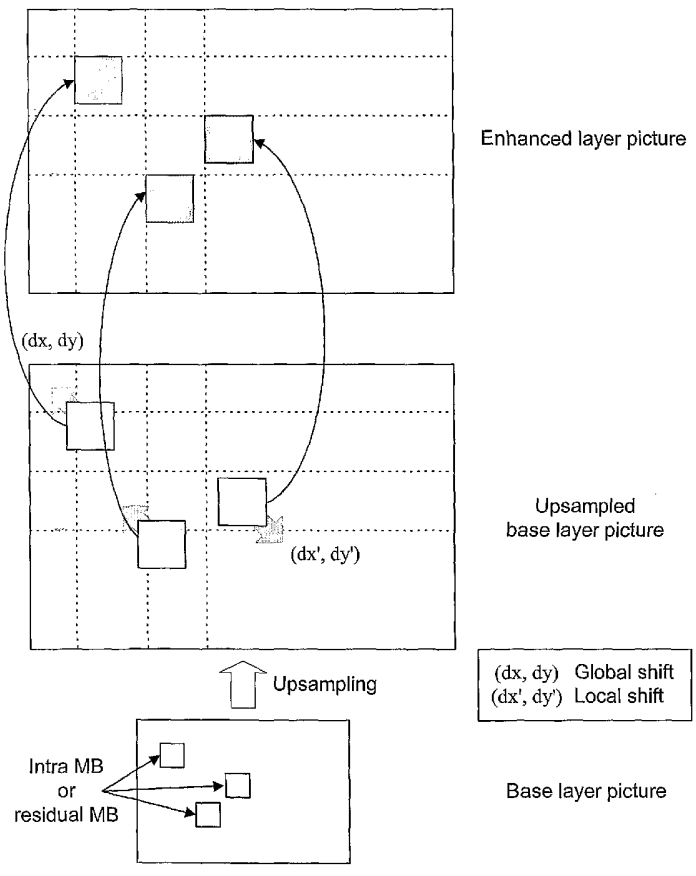
(75) Inventors/Applicants *(for US only)*: **JEON, Byeong Moon** [KR/KR]; 306-1005 Hyundai Apt., Gwang-jang-dong, Gwangjin-gu, Seoul 143-754 (KR). **PARK, Seung Wook** [KR/KR]; 1429-7, Sillim-dong, Gwanak-gu, Seoul 151-891 (KR). **PARK, Ji Ho** [KR/KR]; 53-502 Hyundai Apt., Apkujung-dong, Gangnam-gu, Seoul 135-110 (KR). **YOON, Doe Hyun** [KR/KR]; 101-801 Dongbu Centreville, Garak-dong, Songpa-gu, Seoul 138-160 (KR). **PARK, Hyun Wook** [KR/KR]; 132-1503, Hanbit Apt., Eoun-dong, Yoosung-gu, Daejun-si 305-755 (KR).

(74) Agent: **PARK, Lae Bong**; 2FL., Dongun Bldg, 413-4, Dogok 2-dong, Gangnam-gu, Seoul 135-272 (KR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM,

[Continued on next page]

(54) Title: METHOD FOR SCALABLY ENCODING AND DECODING VIDEO SIGNAL



(57) Abstract: In one embodiment, decoding of a video signal includes predicting at least a portion of a current image in a current layer based on at least a portion of a base image in a base layer and offset information. The offset information indicates an offset based on at least one pixel in the current image and a corresponding at least one pixel in the base image. For example, the offset information may represent a position offset between at least one sample in the current image and at least one sample in an up-sampled portion of the base image.

WO 2006/104363 A1

AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

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

Published:
— *with international search report*

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

DESCRIPTION

METHOD FOR SCALABLY ENCODING AND DECODING VIDEO SIGNAL

1. Technical Field

The present invention relates to scalable encoding and decoding of a video signal.

2. Background Art

It is difficult to allocate high bandwidth, required for TV signals, to digital video signals wirelessly transmitted and received by mobile phones and notebook computers. It is expected that similar difficulties will occur with mobile TVs and handheld PCs, which will come into widespread use in the future. Thus, video compression standards for use with mobile devices should have high video signal compression efficiencies.

Such mobile devices have a variety of processing and presentation capabilities so that a variety of compressed video data forms should be prepared. This means that a variety of different quality video data with different combinations of a number of variables such as the number of frames transmitted per second, resolution, and the number of bits per pixel should be provided based on a single video source. This imposes a great burden on content providers.

Because of the above, content providers prepare high-bitrate compressed video data for each source video and perform, when receiving a request from a mobile device, a process of decoding compressed video and encoding it back into video data suited to the video processing capabilities of the mobile device. However, this method entails a transcoding procedure including

decoding, scaling, and encoding processes, which causes some time delay in providing the requested data to the mobile device. The transcoding procedure also requires complex hardware and algorithms to cope with the wide variety of target encoding formats.

The Scalable Video Codec (SVC) has been developed in an attempt to overcome these problems. This scheme encodes video into a sequence of pictures with the highest image quality while ensuring that part of the encoded picture (frame) sequence (specifically, a partial sequence of frames intermittently selected from the total sequence of frames) can be decoded to produce a certain level of image quality.

Motion Compensated Temporal Filtering (MCTF) is an encoding scheme that has been suggested for use in the Scalable Video Codec. The MCTF scheme has a high compression efficiency (i.e., a high coding efficiency) for reducing the number of bits transmitted per second. The MCTF scheme is likely to be applied to transmission environments such as a mobile communication environment where bandwidth is limited.

Although it is ensured that part of a sequence of pictures encoded in the scalable MCTF coding scheme can be received and processed to video with a certain level of image quality as described above, there is still a problem in that the image quality is significantly reduced if the bitrate is lowered. One solution to this problem is to provide an auxiliary picture sequence for low bitrates, for example, a sequence of pictures that have a small screen size and/or a low frame rate.

The auxiliary picture sequence is referred to as a base layer (BL), and the main picture sequence is referred to as an enhanced or enhancement layer. Video signals of the base and enhanced layers have redundancy since the same video content is encoded into two layers with different spatial resolution or different frame rates. To increase the coding efficiency of the enhanced

layer, a video signal of the enhanced layer may be predicted using motion information and/or texture information of the base layer. This prediction method is referred to as inter-layer prediction.

FIG. 1 illustrates examples of an intra BL prediction method and an inter-layer residual prediction method, which are inter-layer prediction methods for encoding the enhanced layer using the base layer.

The intra BL prediction method uses a texture (or image data) of the base layer. Specifically, the intra BL prediction method produces predictive data of a macroblock of the enhanced layer using a corresponding block of the base layer encoded in an intra mode. The term "corresponding block" refers to a block which is located in a base layer frame temporally coincident with a frame including the macroblock and which would have an area covering the macroblock if the base layer frame were enlarged by the ratio of the screen size of the enhanced layer to the screen size of the base layer. The intra BL prediction method uses the corresponding block of the base layer after enlarging the corresponding block by the ratio of the screen size of the enhanced layer to the screen size of the base layer through upsampling.

The inter-layer residual prediction method is similar to the intra BL prediction method except that it uses a corresponding block of the base layer encoded so as to contain residual data, which is data of an image difference, rather than a corresponding block of the base layer containing image data. The inter-layer residual prediction method produces predictive data of a macroblock of the enhanced layer encoded so as to contain residual data, which is data of an image difference, using a corresponding block of the base layer encoded so as to contain residual data. Similar to the intra BL prediction method, the inter-layer residual prediction method uses the corresponding block of the base layer containing residual data after enlarging the corresponding block by the ratio of the screen size of the enhanced

layer to the screen size of the base layer through upsampling.

A base layer with lower resolution for use in the inter-layer prediction method is produced by downsampling a video source. Corresponding pictures (frames or blocks) in enhanced and base layers produced from the same video source may be out of phase since a variety of different downsampling techniques and downsampling ratios (i.e., horizontal and/or vertical size reduction ratios) may be employed.

FIG. 2 illustrates a phase relationship between enhanced and base layers. A base layer may be produced (i) by sampling a video source at lower spatial resolution separately from an enhanced layer or (ii) by downsampling an enhanced layer with higher spatial resolution. In the example of FIG. 2, the downsampling ratio between the enhanced and base layers is 2/3.

A video signal is managed as separate components, namely, a luma component and two chroma components. The luma component is associated with luminance information Y and the two chroma components are associated with chrominance information Cb and Cr. A ratio of 4:2:0 (Y:Cb:Cr) between luma and chroma signals is widely used. Samples of the chroma signal are typically located midway between samples of the luma signal. When an enhanced layer and/or a base layer are produced directly from a video source, luma and chroma signals of the enhanced layer and/or the base layer are sampled so as to satisfy the 4:2:0 ratio and a position condition according to the 4:2:0 ratio.

In the above case (i), the enhanced and base layers may be out of phase as shown in section (a) of Fig. 2 since the enhanced and base layers may have different sampling positions. In the example of section (a), luma and chroma signals of each of the enhanced and base layers satisfy the 4:2:0 ratio and a position condition according to the 4:2:0 ratio.

In the above case (ii), the base layer is produced by downsampling luma and chroma signals of the enhanced layer by a

specific ratio. If the base layer is produced such that luma and chroma signals of the base layer are in phase with luma and chroma signals of the enhanced layer, the luma and chroma signals of the base layer do not satisfy a position condition according to the 4:2:0 ratio as illustrated in section (b) of Fig. 2.

In addition, if the base layer is produced such that luma and chroma signals of the base layer satisfy a position condition according to the 4:2:0 ratio, the chroma signal of the base layer is out of phase with the chroma signal of the enhanced layer as illustrated in section (c) of Fig. 2. In this case, if the chroma signal of the base layer is upsampled by a specific ratio according to the inter-layer prediction method, the upsampled chroma signal of the base layer is out of phase with the chroma signal of the enhanced layer.

Also in case (ii), the enhanced and base layers may be out of phase as illustrated in section (a).

That is, the phase of the base layer may be changed in the downsampling procedure for producing the base layer and in the upsampling procedure of the inter-layer prediction method, so that the base layer is out of phase with the enhanced layer, thereby reducing coding efficiency.

3. Disclosure of Invention

In one embodiment, decoding of a video signal includes predicting at least a portion of a current image in a current layer based on at least a portion of a base image in a base layer and offset information. The offset information indicates an offset based on at least one pixel in the current image and a corresponding at least one pixel in the base image. For example, the offset information may indicate a position offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

In one embodiment, the offset information indicates at least

one of (i) a horizontal offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image, and (ii) a vertical offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

In an embodiment, the predicting step may obtain the offset information from a header of a slice in the base layer, and in another embodiment the offset information may be obtained from a sequence level header in the current layer.

Other related embodiments include methods of encoding a video signal, and apparatuses for encoding and decoding a video signal.

4. Brief Description of Drawings

The above and other objects, features and other advantages of the present invention will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings, in which:

FIG. 1 illustrates an example of an inter-layer prediction method for encoding an enhanced layer using a base layer;

FIG. 2 illustrates examples of phase relationships between enhanced and base layers;

FIG. 3 is a block diagram of a video signal encoding apparatus to which a scalable video signal coding method according to the present invention may be applied;

FIG. 4 illustrates elements of an EL encoder shown in FIG. 3;

FIG. 5 illustrates a method for upsampling a base layer for use in decoding an enhanced layer, encoded according to an inter-layer prediction method, taking into account a phase shift in the base layer and/or the enhanced layer, according to an embodiment of the present invention;

Fig. 6 is a block diagram of an apparatus for decoding a bit

stream encoded by the apparatus of Fig. 3; and

Fig. 7 illustrates elements of an EL decoder shown in FIG. 6.

5. Modes for Carrying out the Invention

Example embodiments of the present invention will now be described in detail with reference to the accompanying drawings.

FIG. 3 is a block diagram of a video signal encoding apparatus to which a scalable video signal coding method according to the present invention may be applied.

The video signal encoding apparatus shown in FIG. 3 comprises an enhanced layer (EL) encoder 100, a texture coding unit 110, a motion coding unit 120, a muxer (or multiplexer) 130, a downsampling unit 140, and a base layer (BL) encoder 150. The downsampling unit 140 produces an enhanced layer signal directly from an input video signal or by downsampling the input video signal, and produces a base layer signal by downsampling the input video signal or the enhanced layer signal according to a specific scheme. The specific scheme will depend on the applications or devices receiving each layer; and therefore, is a matter of design choice. The EL encoder 100 encodes the enhanced layer signal generated by the downsampling unit 140 on a per macroblock basis in a scalable fashion according to a specified encoding scheme (for example, an MCTF scheme), and generates suitable management information. The texture coding unit 110 converts data of encoded macroblocks into a compressed bitstream. The motion coding unit 120 codes motion vectors of image blocks obtained by the EL encoder 100 into a compressed bitstream according to a specified scheme. The BL encoder 150 encodes the base layer signal generated by the downsampling unit 140 according to a specified scheme, for example, according to the MPEG-1, 2 or 4 standard or the H.261 or H.264 standard, and produces a small-screen picture sequence, for example, a sequence of pictures scaled down to 25% of their

original size if needed. The muxer 130 encapsulates the output data of the texture coding unit 110, the small-screen sequence from the BL encoder 150, and the output vector data of the motion coding unit 120 into a desired format. The muxer 130 multiplexes and outputs the encapsulated data into a desired transmission format.

The downsampling unit 140 not only transmits the enhanced and base layer signals to the EL and BL encoders 100 and 150, but also transmits sampling-related information of the two layers to the EL and BL encoders 100 and 150. The sampling-related information of the two layers may include spatial resolution (or screen sizes), frame rates, the ratios between luma and chroma signals of the two layers, the positions of chroma signals of the two layers, and information regarding a phase shift between luma and chroma signals of the two layers based on the respective positions of the luma and chroma signals of the two layers.

The phase shift can be defined as the phase difference between luma signals of the two layers. Typically, luma and chroma signals of the two layers are sampled so as to satisfy a position condition according to the ratio between the luma and chroma signals, and the luma signals of the two layers are sampled so as to be in phase with each other.

The phase shift can also be defined as the phase difference between chroma signals of the two layers. The phase difference between chroma signals of the two layers can be determined based on the difference between positions of the chroma signals of the two layers after the positions of the luma signals of the two layers are matched to each other so that the luma signals of the two layers are in phase with each other.

The phase shift can also be individually defined for each layer, for example, with reference to a single virtual layer (e.g., an upsampled base layer) based on the input video signal for generating the enhanced or base layer. Here, the phase difference

is between luma and/or chroma samples (i.e., pixels) of the enhanced layer of the base layer and the virtual layer (e.g., an upsampled base layer).

The EL encoder 100 records the phase shift information transmitted from the downsampling unit 140 in a header area of a sequence layer or a slice layer. If the phase shift information has a value other than 0, the EL encoder 100 sets a global shift flag "global_shift_flag", which indicates whether or not there is a phase shift between the two layers, to, for example, "1", and records the value of the phase shift in information in fields "global_shift_x" and "global_shift_y". The "global_shift_x" value represents the horizontal phase shift. The "global_shift_y" value represents the vertical phase shift. Stated another way, the "global_shift_x" value represents the horizontal position offset between the samples (i.e., pixels), and the "global_shift_y" represents the vertical position offset between the samples (i.e., pixels).

On the other hand, if the phase shift information has a value of 0, the EL encoder 100 sets the flag "global_shift_flag" to, for example, "0", and does not record the values of the phase shift in the information fields "global_shift_x" and "global_shift_y".

The EL encoder 100 also records the sampling-related information in the header area of the sequence layer or the slice layer if needed.

The EL encoder 100 performs MCTF on the video data received from the down-sampling unit 140. Accordingly, the EL encoder 100 performs a prediction operation on each macroblock in a video frame (or picture) by subtracting a reference block, found by motion estimation, from the macroblock. Also, the EL encoder 100 selectively performs an update operation by adding an image difference between the reference block and the macroblock to the reference block.

The EL encoder 100 separates an input video frame sequence

into, for example, odd and even frames. The EL encoder 100 performs prediction and update operations on the separated frames over a number of encoding levels, for example, until the number of L frames, which are produced by the update operation, is reduced to one for a group of pictures (GOP). FIG. 4 shows elements of the EL encoder 100 associated with prediction and update operations at one of the encoding levels.

The elements of the EL encoder 100 shown in FIG. 4 include an estimator/predictor 101. Through motion estimation, the estimator/predictor 101 searches for a reference block of each macroblock of a frame (for example, an odd frame in the enhanced layer), which is to contain residual data, and then performs a prediction operation to calculate an image difference (i.e., a pixel-to-pixel difference) of the macroblock from the reference block and a motion vector from the macroblock to the reference block. The EL encoder 100 may further include an updater 102 for performing an update operation on a frame (for example, an even frame) including the reference block of the macroblock by normalizing the calculated image difference of the macroblock from the reference block and adding the normalized value to the reference block.

A block having the smallest image difference from a target block has the highest correlation with the target block. The image difference of two blocks is defined, for example, as the sum or average of pixel-to-pixel differences of the two blocks. Of blocks having a threshold pixel-to-pixel difference sum (or average) or less from the target block, a block(s) having the smallest difference sum (or average) is referred to as a reference block(s).

The operation carried out by the estimator/predictor 101 is referred to as a 'P' operation, and a frame produced by the 'P' operation is referred to as an 'H' frame. The residual data present in the 'H' frame reflects high frequency components of

the video signal. The operation carried out by the updater 102 is referred to as a 'U' operation, and a frame produced by the 'U' operation is referred to as an 'L' frame. The 'L' frame is a low-pass subband picture.

The estimator/predictor 101 and the updater 102 of FIG. 4 may perform their operations on a plurality of slices, which are produced by dividing a single frame, simultaneously and in parallel, instead of performing their operations in units of frames. In the following description of the embodiments, the term 'frame' is used in a broad sense to include a 'slice', provided that replacement of the term 'frame' with the term 'slice' is technically equivalent.

More specifically, the estimator/predictor 101 divides each input video frame or each odd one of the L frames obtained at the previous level into macroblocks of a size. The estimator/predictor 101 then searches for a block, whose image is most certain similar to that of each divided macroblock, in the current odd frame or in even frames prior to and subsequent to the current odd frame at the same temporal decomposition level, and produces a predictive image of each divided macroblock using the most similar or reference block and obtains a motion vector thereof.

As shown in Fig. 4, the EL encoder 100 may also include a BL decoder 105. The BL decoder 105 extracts encoding information such as a macroblock mode from an encoded base layer stream containing a small-screen sequence received from the BL encoder 150, and decodes the encoded base layer stream to produce frames, each composed of one or more macroblocks. The estimator/predictor 101 can also search for a reference block of the macroblock in a frame of the base layer according to the intra BL prediction method. Specifically, the estimator/predictor 101 searches for a corresponding block encoded in an intra mode in a frame of the base layer reconstructed by the BL decoder 105, which is

temporally coincident with the frame including the macroblock. The term "corresponding block" refers to a block which is located in the temporally coincident base layer frame and which would have an area covering the macroblock if the base layer frame were enlarged by the ratio of the screen size of the enhanced layer to the screen size of the base layer.

The estimator/predictor 101 reconstructs an original image of the found corresponding block by decoding the intra-coded pixel values of the corresponding block, and then upsamples the found corresponding block to enlarge it by the ratio of the screen size of the enhanced layer to the screen size of the base layer. The estimator/predictor 101 performs this upsampling taking into account the phase shift information "global_shift_x/y" transmitted from the downsampling unit 140 so that the enlarged corresponding block of the base layer is in phase with the macroblock of the enhanced layer.

The estimator/predictor 101 encodes the macroblock with reference to a corresponding area in the corresponding block of the base layer, which has been enlarged so as to be in phase with the macroblock. Here, the term "corresponding area" refers to a partial area in the corresponding block which is at the same relative position in the frame as the macroblock.

If needed, the estimator/predictor 101 searches for a reference area more highly correlated with the macroblock in the enlarged corresponding block of the base layer by performing motion estimation on the macroblock while changing the phase of the corresponding block, and encodes the macroblock using the found reference area.

If the phase of the enlarged corresponding block is further changed while the reference area is searched for, the estimator/predictor 101 sets a local shift flag "local_shift_flag", which indicates whether or not there is a phase shift, different from the global phase shift

"global_shift_x/y", between the macroblock and the corresponding upsampled block, to, for example, "1". Also, the estimator/predictor 101 records the local shift flag in a header area of the macroblock and records the local phase shift between the macroblock and the corresponding block in information fields "local_shift_x" and "local_shift_y". The local phase shift information may be replacement information, and provide the entire phase shift information as a replacement or substitute for the global phase shift information. Alternatively, the local phase shift information may be additive information, wherein the local phase shift information added to the corresponding global phase shift information provides the entire or total phase shift information.

The estimator/predictor 101 further inserts information indicating that the macroblock of the enhanced layer has been encoded in an intra BL mode in the header area of the macroblock so as to inform the decoder of the same.

The estimator/predictor 101 can also apply the inter-layer residual prediction method to a macroblock to contain residual data, which is data of an image difference, using a reference block found in other frames prior to and subsequent to the macroblock. Also in this case, the estimator/predictor 101 upsamples a corresponding block of the base layer encoded so as to contain residual data, which is data of an image difference, taking into account the phase shift information "global_shift_x/y" transmitted from the downsampling unit 140 so that the base layer is in phase with the enhanced layer. Here, the corresponding block of the base layer is a block which has been encoded so as to contain residual data, which is data of an image difference.

The estimator/predictor 101 inserts information indicating that the macroblock of the enhanced layer has been encoded according to the inter-layer residual prediction method in the header area of the macroblock so as to inform the decoder of the

same.

The estimator/predictor 101 performs the above procedure for all macroblocks in the frame to complete an H frame which is a predictive image of the frame. The estimator/predictor 101 performs the above procedure for all input video frames or all odd ones of the L frames obtained at the previous level to complete H frames which are predictive images of the input frames.

As described above, the updater 102 adds an image difference of each macroblock in an H frame produced by the estimator/predictor 101 to an L frame having its reference block, which is an input video frame or an even one of the L frames obtained at the previous level.

The data stream encoded in the method described above is transmitted by wire or wirelessly to a decoding apparatus or is delivered via recording media. The decoding apparatus reconstructs the original video signal according to the method described below.

FIG. 5 illustrates a method for upsampling a base layer for use in decoding an enhanced layer, encoded according to the inter-layer prediction method, taking into account a phase shift in the base layer and/or the enhanced layer, according to an embodiment of the present invention.

In order to decode a macroblock of the enhanced layer encoded according to the inter-layer prediction method, a block of the base layer corresponding to the macroblock is enlarged by the ratio of the screen size of the enhanced layer to the screen size of the base layer through upsampling. This upsampling is performed taking into account phase shift information "global_shift_x/y" in the enhanced layer and/or the base layer, so as to compensate for a global phase shift between the macroblock of the enhanced layer and the enlarged corresponding block of the base layer.

If there is a local phase shift "local_shift_x/y", different

from the global phase shift "global_shift_x/y", between the macroblock of the enhanced layer and the corresponding block of the base layer, the corresponding block is upsampled taking into account the local phase shift "local_shift_x/y". For example, the local phase shift information may be used instead of the global phase shift information in one embodiment, or alternatively, in addition to the global phase shift information in another embodiment.

Then, an original image of the macroblock of the enhanced layer is reconstructed using the corresponding block which has been enlarged so as to be in phase with the macroblock.

Fig. 6 is a block diagram of an apparatus for decoding a bit stream encoded by the apparatus of Fig. 3. The decoding apparatus of Fig. 6 includes a demuxer (or demultiplexer) 200, a texture decoding unit 210, a motion decoding unit 220, an EL decoder 230, and a BL decoder 240. The demuxer 200 separates a received bit stream into a compressed motion vector stream and a compressed macroblock information stream. The texture decoding unit 210 reconstructs the compressed macroblock information stream to its original uncompressed state. The motion decoding unit 220 reconstructs the compressed motion vector stream to its original uncompressed state. The EL decoder 230 converts the uncompressed macroblock information stream and the uncompressed motion vector stream back to an original video signal according to a specified scheme (for example, an MCTF scheme). The BL decoder 240 decodes a base layer stream according to a specified scheme (for example, the MPEG4 or H.264 standard).

The EL decoder 230 uses encoding information of the base layer and/or a decoded frame or macroblock of the base layer in order to decode an enhanced layer stream according to the inter-layer prediction method. To accomplish this, the EL decoder 230 reads a global shift flag "global_shift_flag" and phase shift information "global_shift_x/y" from a sequence header area or a

slice header area of the enhanced layer to determine whether or not there is a phase shift in the enhanced layer and/or the base layer and to confirm the phase shift. The EL decoder 230 upsamples the base layer taking into account the confirmed phase shift so that the base layer to be used for the inter-layer prediction method is in phase with the enhanced layer.

The EL decoder 230 reconstructs an input stream to an original frame sequence. Fig. 7 illustrates main elements of an EL decoder 230 which is implemented according to the MCTF scheme.

The elements of the EL decoder 230 of Fig. 7 perform temporal composition of H and L frame sequences of temporal decomposition level N into an L frame sequence of temporal decomposition level N-1. The elements of Fig. 7 include an inverse updater 231, an inverse predictor 232, a motion vector decoder 233, and an arranger 234. The inverse updater 231 selectively subtracts difference values of pixels of input H frames from corresponding pixel values of input L frames. The inverse predictor 232 reconstructs input H frames into L frames of original images using both the H frames and the above L frames, from which the image differences of the H frames have been subtracted. The motion vector decoder 233 decodes an input motion vector stream into motion vector information of blocks in H frames and provides the motion vector information to an inverse updater 231 and an inverse predictor 232 of each stage. The arranger 234 interleaves the L frames completed by the inverse predictor 232 between the L frames output from the inverse updater 231, thereby producing a normal L frame sequence.

The L frames output from the arranger 234 constitute an L frame sequence 701 of level N-1. A next-stage inverse updater and predictor of level N-1 reconstructs the L frame sequence 701 and an input H frame sequence 702 of level N-1 to an L frame sequence. This decoding process is performed over the same number of levels as the number of encoding levels performed in the

encoding procedure, thereby reconstructing an original video frame sequence.

A reconstruction (temporal composition) procedure at level N, in which received H frames of level N and L frames of level N produced at level N+1 are reconstructed to L frames of level N-1, will now be described in more detail.

For an input L frame of level N, the inverse updater 231 determines all corresponding H frames of level N, whose image differences have been obtained using, as reference blocks, blocks in an original L frame of level N-1 updated to the input L frame of level N at the encoding procedure, with reference to motion vectors provided from the motion vector decoder 233. The inverse updater 231 then subtracts error values of macroblocks in the corresponding H frames of level N from pixel values of corresponding blocks in the input L frame of level N, thereby reconstructing an original L frame.

Such an inverse update operation is performed for blocks in the current L frame of level N, which have been updated using error values of macroblocks in H frames in the encoding procedure, thereby reconstructing the L frame of level N to an L frame of level N-1.

For a target macroblock in an input H frame, the inverse predictor 232 determines its reference blocks in inverse-updated L frames output from the inverse updater 231 with reference to motion vectors provided from the motion vector decoder 233, and adds pixel values of the reference blocks to difference (error) values of pixels of the target macroblock, thereby reconstructing its original image.

If information indicating that a macroblock in an H frame has been encoded in an intra BL mode is included in a header area of the macroblock, the inverse predictor 232 reconstructs an original image of the macroblock using a base layer frame provided from the BL decoder 240. The following is a detailed example of

this process.

The inverse predictor 232 reconstructs an original image of an intra-coded block in the base layer, which corresponds to the macroblock in the enhanced layer, and upsamples the reconstructed corresponding block from the base layer to enlarge it by the ratio of the screen size of the enhanced layer to the screen size of the base layer. The inverse predictor 232 performs this upsampling taking into account phase shift information "global_shift_x/y" in the enhanced layer and/or the base layer so that the enlarged corresponding block of the base layer is in phase with the macroblock of the enhanced layer. Namely, if the "global_shift_flag" indicates a phase shift exists between the base layer and the enhanced layer (e.g., equals 1), then the inverse predictor 232 phase shifts the corresponding macroblock from the base layer during upsampling by the "global_shift_x" and "global_shift_y" values. The inverse predictor 232 reconstructs an original image of the macroblock by adding pixel values of a corresponding area in the enlarged corresponding block of the base layer, which has been enlarged so as to be in phase with the macroblock, to the difference values of pixels of the macroblock. Here, the term "corresponding area" refers to a partial area in the corresponding block which is at the same relative position in the frame as the macroblock.

If a local shift flag "local_shift_flag" indicates that there is a local phase shift "local_shift_x/y" different from the global phase shift "global_shift_x/y" between the macroblock and the corresponding block, the inverse predictor 232 upsamples the corresponding block taking into account the local phase shift "local_shift_x/y" (as substitute or additional phase shift information). The local phase shift information may be included in the header area of the macroblock.

If information indicating that a macroblock in an H frame has been encoded in an inter-layer residual mode is included in

a header area of the macroblock, the inverse predictor 232 upsamples a corresponding block of the base layer encoded so as to contain residual data, taking into account the global phase shift "global_shift_x/y" as discussed above to enlarge the corresponding block so as to be in phase with the macroblock of the enhanced layer. The inverse predictor 232 then reconstructs residual data of the macroblock using the corresponding block enlarged so as to be in phase with the macroblock.

The inverse predictor 232 searches for a reference block of the reconstructed macroblock containing residual data in an L frame with reference to a motion vector provided from the motion vector decoder 233, and reconstructs an original image of the macroblock by adding pixel values of the reference block to difference values of pixels (i.e., residual data) of the macroblock.

All macroblocks in the current H frame are reconstructed to their original images in the same manner as the above operation, and the reconstructed macroblocks are combined to reconstruct the current H frame to an L frame. The arranger 234 alternately arranges L frames reconstructed by the inverse predictor 232 and L frames updated by the inverse updater 231, and outputs such arranged L frames to the next stage.

The above decoding method reconstructs an MCTF-encoded data stream to a complete video frame sequence. In the case where the prediction and update operations have been performed for a group of pictures (GOP) N times in the MCTF encoding procedure described above, a video frame sequence with the original image quality is obtained if the inverse update and prediction operations are performed N times in the MCTF decoding procedure. However, a video frame sequence with a lower image quality and at a lower bitrate may be obtained if the inverse update and prediction operations are performed less than N times. Accordingly, the decoding apparatus is designed to perform inverse update and prediction

operations to the extent suitable for the performance thereof.

The decoding apparatus described above can be incorporated into a mobile communication terminal, a media player, or the like.

As is apparent from the above description, a method for encoding and decoding a video signal according to the present invention increases coding efficiency by preventing a phase shift in a base layer and/or an enhanced layer caused in downsampling and upsampling procedures when encoding/decoding the video signal according to an inter-layer prediction method.

Although the example embodiments of the present invention have been disclosed for illustrative purposes, those skilled in the art will appreciate that various improvements, modifications, substitutions, and additions are possible, without departing from the scope and spirit of the invention.

CLAIMS

1. A method for decoding a video signal, comprising:

predicting at least a portion of a current image in a current layer based on at least a portion of a base image in a base layer and offset information, the offset information indicating an offset based on at least one pixel in the current image and a corresponding at least one pixel in the base image.

2. The method of claim 1, wherein the samples are luma samples.

3. The method of claim 1, wherein the samples are chroma samples.

4. The method of claim 1, wherein the samples are luma and chroma samples.

5. The method of claim 1, wherein the predicting step predicts the portion of the current image based on at least part of an up-sampled portion of the base image and the offset information.

6. The method of claim 5, wherein the offset information indicates a position offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

7. The method of claim 6, wherein the offset information indicates a horizontal offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

8. The method of claim 7, wherein the offset information indicates a vertical offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

9. The method of claim 6, wherein the offset information indicates a vertical offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

10. The method of claim 1, wherein the predicting step comprises:

up-sampling at least the portion of the base image to produce an up-sampled image; and

predicting the portion of the current image based on at least part of the portion of up-sampled image and the offset information, the offset information indicating a position offset between at least one pixel in the portion of the current image and at least one pixel in the part of the up-sampled image.

11. The method of claim 10, wherein the offset information indicates a horizontal offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

12. The method of claim 11, wherein the offset information indicates a vertical offset between at least one sample in the current image and at least one sample in the up-sampled portion of the base image.

13. The method of claim 10, wherein the offset information indicates a vertical offset between at least one sample in the

current image and at least one sample in the up-sampled portion of the base image.

14. The method of claim 1, wherein the predicting step obtains the offset information from a header of a slice in the base layer.

15. The method of claim 14, wherein the predicting step determines that the offset information is present based on an indicator in the header of the slice.

16. The method of claim 1, wherein the predicting step obtains the offset information from a sequence level header in the current layer.

17. The method of claim 16, wherein the predicting step determines that the offset information is present based on an indicator in the sequence level header.

18. The method of claim 1, wherein the predicting step determines that the offset information is present based on an indicator in one of the base layer and the current layer.

19. A method of encoding a video signal, comprising:

encoding at least a portion of a current image in a current layer based on at least a portion of a base image in a base layer; and

recording offset information in the encoded video signal, the offset information indicating an offset based on at least one pixel in the current image and a corresponding at least one pixel in the base image.

20. An apparatus for decoding a video signal, comprising:

a decoder predicting at least a portion of a current image in a current layer based on at least a portion of a base image in a base layer and offset information, the offset information indicating an offset based on at least one pixel in the current image and a corresponding at least one pixel in the base image.

21. An apparatus for encoding a video signal, comprising:

an encoder encoding at least a portion of a current image in a current layer based on at least a portion of a base image in a base layer, and recording offset information in the encoded video signal, the offset information indicating an offset based on at least one pixel in the current image and a corresponding at least one pixel in the base image.

FIG. 1

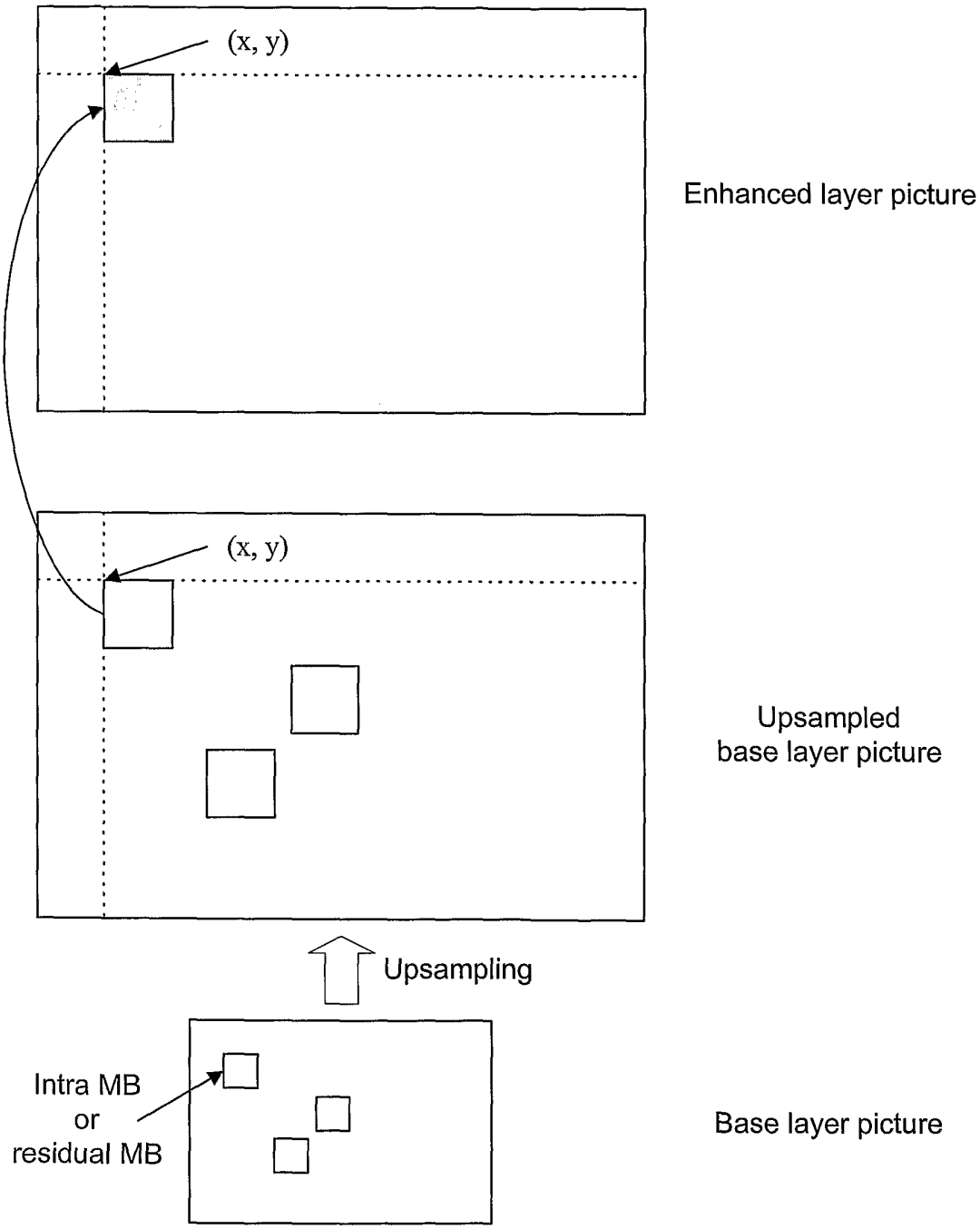

(x, y)
Enhanced layer picture
(x, y)
Upsampled
base layer picture
Upsampling
Intra MB
or
residual MB
Base layer picture

FIG. 2

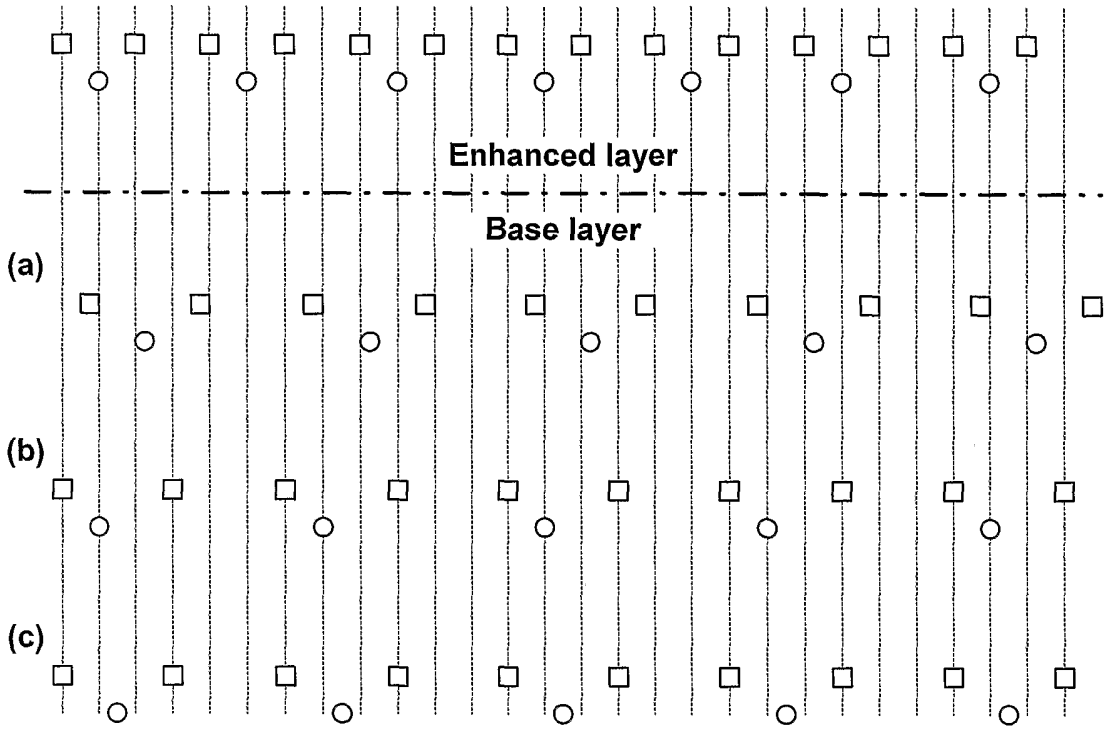

Enhanced layer
Base layer
(a)
(b)
(c)
Samples of Luma signals
Samples of Chroma signals

FIG. 3

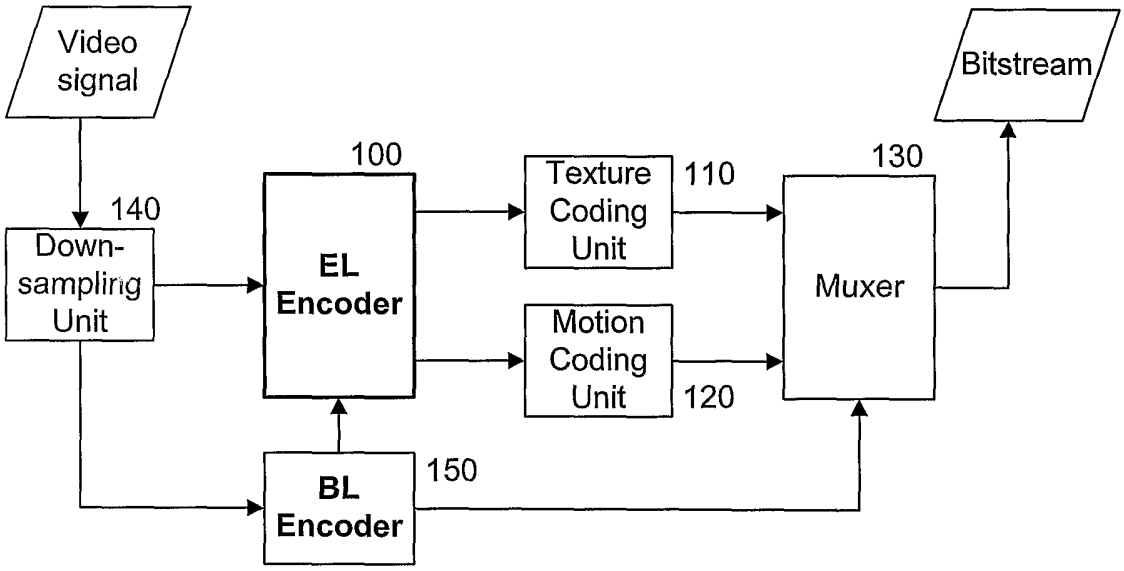

Video signal
Bitstream
100
Texture Coding Unit
110
130
140
Down-sampling Unit
EL Encoder
Muxer
Motion Coding Unit
120
BL Encoder
150

FIG. 4

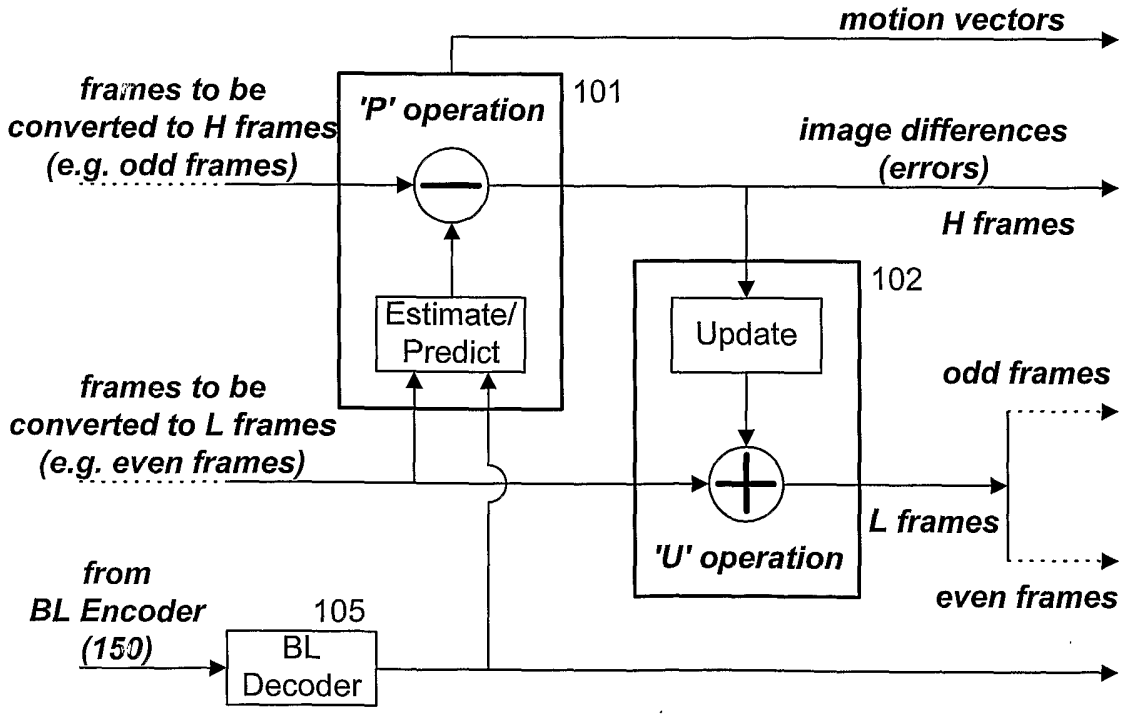

motion vectors
frames to be converted to H frames (e.g. odd frames)
'P' operation
101
image differences (errors)
H frames
102
Estimate/ Predict
Update
frames to be converted to L frames (e.g. even frames)
odd frames
L frames
'U' operation
from BL Encoder (150)
105
BL Decoder
even frames

FIG. 5

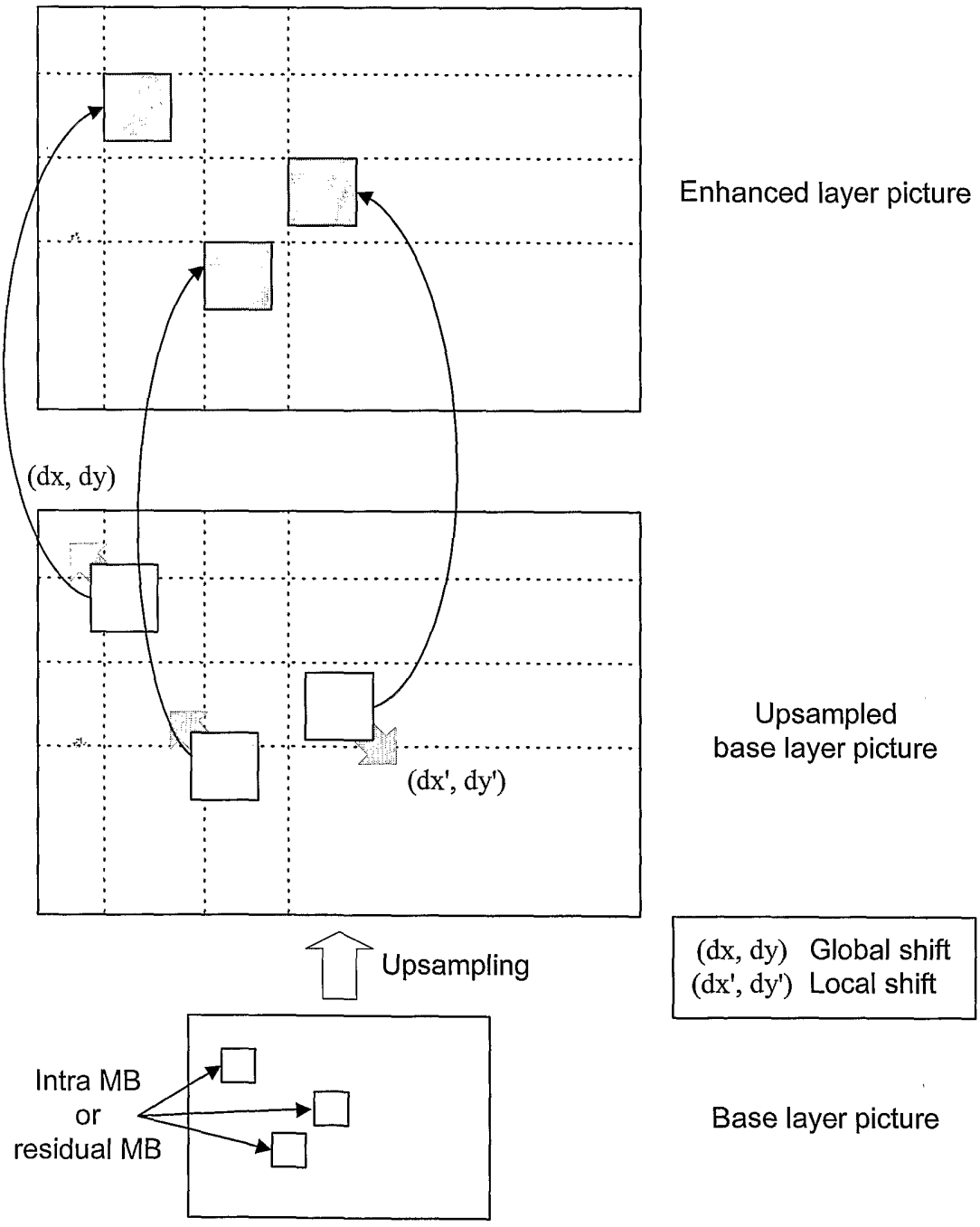

Enhanced layer picture
(dx, dy)
(dx', dy')
Upsampled
base layer picture
(dx, dy) Global shift
(dx', dy') Local shift
Upsampling
Intra MB
or
residual MB
Base layer picture

FIG. 6

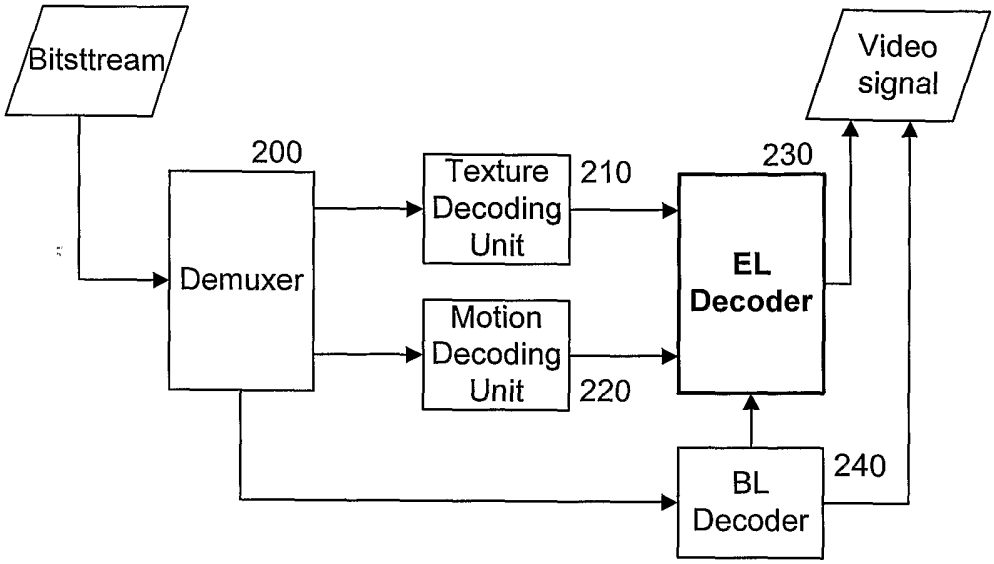
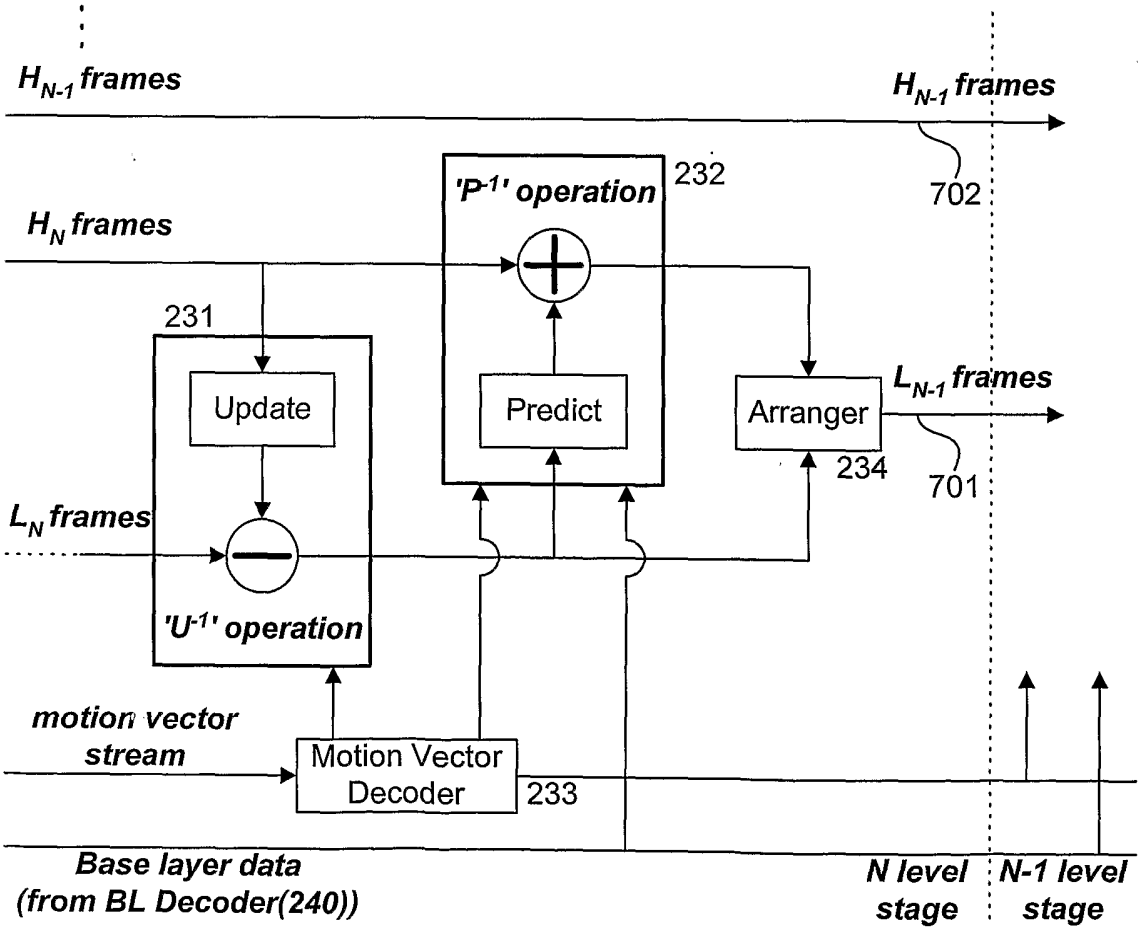

Bitsttream
Video signal
200
Demuxer
Texture Decoding Unit
210
Motion Decoding Unit
220
230
EL Decoder
BL Decoder
240

FIG. 7


H_{N-1} frames
H_{N-1} frames
702
'P^{-1}' operation
232
H_N frames
231
Update
Predict
Arranger
L_{N-1} frames
234
701
L_N frames
'U^{-1}' operation
motion vector stream
Motion Vector Decoder
233
Base layer data (from BL Decoder(240))
N level stage
N-1 level stage

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2006/001196

A. CLASSIFICATION OF SUBJECT MATTER

H04N 7/24(2006.01)i

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

Korean Patents and applications for inventions since 1975
Korean Utility models and applictions for Utility models since 1975
Japanese Utility models and application for Utility models since 1975

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

eKIPASS (KIPO internal), IEEE Xplore: "video, encoding, decoding, scalable, offset, shift, up-sample, down-sample"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6510177 B1 (BONET et al.) 21 Jan. 2003 See abstract; claims; figure 3-11B.	1-21
A	US 6836512 B2 (SCHAAR et al.) 28 Dec. 2004 See abstract; figures 3, 4A, 4B.	1-21
A	US 6728317 B1 (DEMOS) 27 Apr. 2004 See abstract; figures 8-11.	1-21
A	US 6057884 A (CHEN et al.) 02 May 2000 See abstract; figure 2.	1-21

☐ 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 application or patent 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 citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 JUNE 2006 (23.06.2006)	**26 JUNE 2006 (26.06.2006)**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office 920 Dunsan-dong, Seo-gu, Daejeon 302-701, Republic of Korea	KIM, Heung Soo
Facsimile No. 82-42-472-7140	Telephone No. 82-42-481-5764

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/KR2006/001196

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US06510177	21.01.2003	US6510177B1	21.01.2003
		US6510177BA	21.01.2003
US06836512	28.12.2004	US20020071486A1	13.06.2002
		US2002071486A1	13.06.2002
		US2002071486AA	13.06.2002
		US6836512BB	28.12.2004
US6728317B1	27.04.2004	AU200151386A1	23.10.2001
		AU200151386A5	23.10.2001
		CA2245172AA	07.08.1997
		CA2245172C	12.04.2005
		CA2245172C	07.08.1997
		CA2406459AA	18.10.2001
		CA2406459A1	18.10.2001
		CN1219255	09.06.1999
		CN1219255A	09.06.1999
		EP01012738A1	28.06.2000
		EP01279111A1	29.01.2003
		EP1012738A1	28.06.2000
		EP1012738A4	30.05.2001
		EP1279111A1	29.01.2003
		EP1279111A4	23.03.2005
		JP13500674	16.01.2001
		JP2001500674T2	16.01.2001
		JP2003531514T2	21.10.2003
		KR1019990082104	15.11.1999
		SG79277A1	20.03.2001
		US05852565	22.12.1998
		US05988863	23.11.1999
		US06728317	27.04.2004
		US2004196901A1	07.10.2004
		US2004196901AA	07.10.2004
		US2005254649AA	17.11.2005
		US5852565A	22.12.1998
		US5988863A	23.11.1999
		US6728317BA	27.04.2004
		US6957350BA	18.10.2005
		WO0177871A1	18.10.2001
		WO0200177871A1	18.10.2001
		WO9728507A1	07.08.1997

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/KR2006/001196

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US06057884	02.05.2000	AU199869934B2	10.12.1998
		AU6993498A1	10.12.1998
		AU733055B2	03.05.2001
		CA2238900AA	05.12.1998
		CA2238900C	21.10.2003
		CA2238900C	05.12.1998
		CN1209020	24.02.1999
		CN1209020A	24.02.1999
		CN1551636A	01.12.2004
		EP00883300A2	09.12.1998
		EP00883300A3	20.12.2000
		EP0883300A2	09.12.1998
		EP883300A2	09.12.1998
		EP883300A3	20.12.2000
		JP11018085A2	22.01.1999
		JP11018085	22.01.1999
		KR1019990006678	25.01.1999
		N0982508A0	02.06.1998
		TW406512B	21.09.2000
		TW406512A	21.09.2000
		US6057884A	02.05.2000