

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 October 2006 (19.10.2006)

PCT

(10) International Publication Number  
**WO 2006/109988 A1**

(51) International Patent Classification:  
**H04N 7/32** (2006.01)

135-110 (KR). **YOON, Doe Hyun** [KR/KR]; 101-801 Dongbu Centreville, Garak-dong, Songpa-gu, Seoul 138-160 (KR).

(21) International Application Number:

PCT/KR2006/001341

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

(22) International Filing Date: 12 April 2006 (12.04.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/670,676 13 April 2005 (13.04.2005) US  
10-2005-0066622 22 July 2005 (22.07.2005) KR

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

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

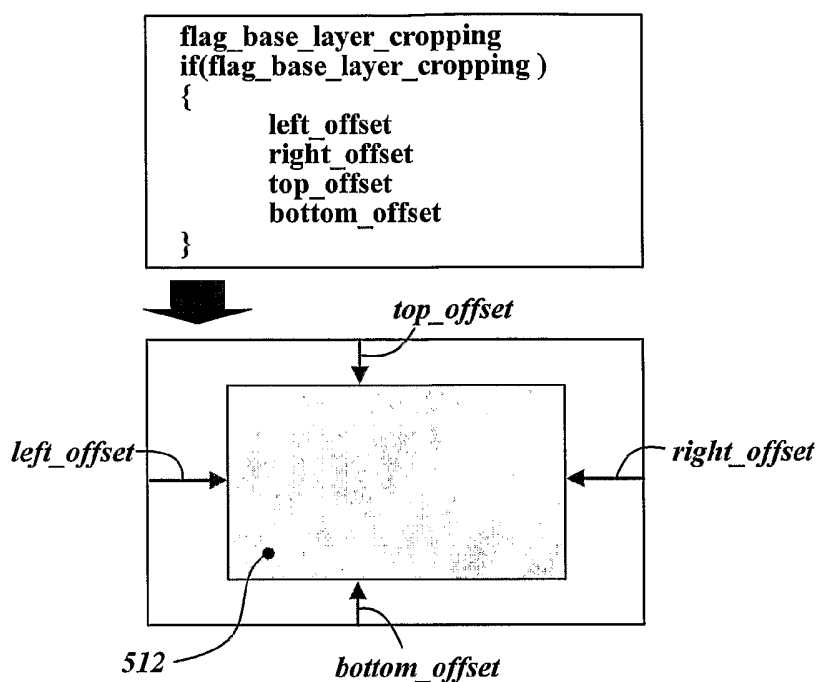
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT,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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **JEON, Byeong Moon** [KR/KR]; 306-1005 Hyundai Apt., Gwangjang-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

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR DECODING VIDEO SIGNAL USING REFERENCE PICTURES



(57) Abstract: In the method for decoding a video signal, at least a portion of a current image in a current layer is predicted based on at least a portion of a reference image, offset information and dimension information. The offset information may indicate a position offset between at least one boundary pixel of the reference image and at least one boundary pixel of the current image, and the dimension information may indicate at least one dimension of the current image.



**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 AND APPARATUS FOR DECODING VIDEO SIGNAL USING REFERENCE PICTURES

### 1. Technical Field

5       The present invention relates to scalable encoding and decoding of a video signal, and more particularly to a method and apparatus for encoding a video signal, wherein a base layer in the video signal is additionally used to code an enhanced layer in the video signal, and a method and apparatus for decoding such  
10 encoded video data.

### 2. Background Art

Scalable Video Codec (SVC) is a method which encodes video into a sequence of pictures with the highest image quality while  
15 ensuring that part of the encoded picture sequence (specifically, a partial sequence of frames intermittently selected from the total sequence of frames) can also be decoded and used to represent the video with a low image quality. Motion Compensated Temporal Filtering (MCTF) is an encoding scheme that has been suggested  
20 for use in the scalable video codec.

Although it is possible to represent low image-quality video by receiving and processing part of the sequence of pictures encoded in a scalable fashion as described above, there is still a problem in that the image quality is significantly reduced if  
25 the bitrate is lowered. One solution to this problem is to hierarchically 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, so that each decoder can select and decode a sequence suitable for its capabilities and

characteristics. One example is to encode and transmit not only a main picture sequence of 4CIF (Common Intermediate Format) but also an auxiliary picture sequence of CIF and an auxiliary picture sequence of QCIF (Quarter CIF) to decoders. Each sequence is referred to as a layer, and the higher of two given layers is referred to as an enhanced layer and the lower is referred to as a base layer.

Such picture sequences have redundancy since the same video signal source is encoded into the sequences. To increase the coding efficiency of each sequence, there is a need to reduce the amount of coded information of the higher sequence by performing inter-sequence picture prediction of video frames in the higher sequence from video frames in the lower sequence temporally coincident with the video frames in the higher sequence.

However, video frames in sequences of different layers may have different aspect ratios. For example, video frames of the higher sequence (i.e., the enhanced layer) may have a wide aspect ratio of 16:9, whereas video frames of the lower sequence (i.e., the base layer) may have a narrow aspect ratio of 4:3. In this case, there is a need to determine which part of a base layer picture is to be used for an enhanced layer picture or for which part of the enhanced layer picture the base layer picture is to be used when performing prediction of the enhanced layer picture.

### 3. Disclosure of Invention

The present invention relates to decoding a video signal.

In one embodiment of the method for decoding a video signal, In the method for decoding a video signal, at least a portion of a current image in a current layer is predicted based on at least a portion of a reference image, offset information and dimension information. The offset information may indicate a position offset between at least one boundary pixel of the reference image and at least one boundary pixel of the current image, and the

dimension information may indicate at least one dimension of the current image.

In one embodiment, the reference image is based on a base image in a base layer. For example, the reference image may be  
5 at least an up-sampled portion of the base image.

In one embodiment, the offset information includes left offset information indicating a position offset between at least one left side pixel of the reference image and at least one left side pixel of the current image.

10 In another embodiment, the offset information includes top offset information indicating a position offset between at least one top side pixel of the reference image and at least one top side pixel of the current image.

In one embodiment, the dimension information includes width  
15 information indicating a width of the current image.

In another embodiment, the dimension information includes height information indicating a height of the current image.

In one embodiment, the offset information may be obtained from a header for at least a portion of a picture (e.g., a slice, frame, etc.) in the current layer. Also, it may be determined that  
20 the offset information is present based on an indicator in the header.

Other embodiments include apparatuses for decoding a video signal.

25

#### **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  
30 accompanying drawings, in which:

FIG. 1 is a block diagram of a video signal encoding apparatus to which a scalable video signal coding method according to the present invention is applied;

FIG. 2 is a block diagram of part of an MCTF encoder shown in FIG. 1 responsible for carrying out image estimation/prediction and update operations;

FIGS. 3a and 3b illustrate the relationship between enhanced layer frames and base layer frames which can be used as reference frames for converting an enhanced layer frame to an H frame having a predictive image;

FIG. 4 illustrates how part of a base layer picture is selected and enlarged to be used for a prediction operation of an enhanced layer picture according to an embodiment of the present invention;

FIGS. 5a and 5b illustrate embodiments of the structure of information regarding a positional relationship of a base layer picture to an enhanced layer picture, which is transmitted to the decoder, according to the present invention;

FIG. 6 illustrates how an area including a base layer picture is enlarged to be used for a prediction operation of an enhanced layer picture according to another embodiment of the present invention;

FIG. 7 illustrates how a base layer picture is enlarged to a larger area than an enhanced layer picture so as to be used for a prediction operation of the enhanced layer picture according to yet another embodiment of the present invention;

FIG. 8 is a block diagram of an apparatus for decoding a data stream encoded by the apparatus of FIG. 1; and

FIG. 9 is a block diagram of part of an MCTF decoder shown in FIG. 8 responsible for carrying out inverse prediction and update operations.

## 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. 1 is a block diagram of a video signal encoding

apparatus to which a scalable video signal coding method according to the present invention is applied. Although the apparatus of FIG. 1 is implemented to code an input video signal in two layers, principles of the present invention described below can also be applied when a video signal is coded in three or more layers. The present invention can also be applied to any scalable video coding scheme, without being limited to an MCTF scheme which is described below as an example.

The video signal encoding apparatus shown in FIG. 1 comprises an MCTF encoder 100 to which the present invention is applied, a texture coding unit 110, a motion coding unit 120, a base layer encoder 150, and a muxer (or multiplexer) 130. The MCTF encoder 100 is an enhanced layer encoder which encodes an input video signal on a per macroblock basis according to an MCTF scheme and generates suitable management information. The texture coding unit 110 converts information of encoded macroblocks into a compressed bitstream. The motion coding unit 120 codes motion vectors of image blocks obtained by the MCTF encoder 100 into a compressed bitstream according to a specified scheme. The base layer encoder 150 encodes an input video signal according to a specified scheme, for example, according to the MPEG-1, 2 or 4 standard or the H.261, H.263 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. The muxer 130 encapsulates the output data of the texture coding unit 110, the small-screen picture sequence output from the base layer encoder 150, and the output vector data of the motion coding unit 120 into a desired format. The muxer 130 then multiplexes and outputs the encapsulated data into a desired transmission format. The base layer encoder 150 can provide a low-bitrate data stream not only by encoding an input video signal into a sequence of pictures having a smaller screen size than pictures of the enhanced layer, but also by encoding an input video signal into

a sequence of pictures having the same screen size as pictures of the enhanced layer at a lower frame rate than the enhanced layer. In the embodiments of the present invention described below, the base layer is encoded into a small-screen picture sequence, and  
5 the small-screen picture sequence is referred to as a base layer sequence and the frame sequence output from the MCTF encoder 100 is referred to as an enhanced layer sequence.

The MCTF encoder 100 performs motion estimation and prediction operations on each target macroblock in a video frame.  
10 The MCTF encoder 100 also performs an update operation for each target macroblock by adding an image difference of the target macroblock from a corresponding macroblock in a neighbor frame to the corresponding macroblock in the neighbor frame. FIG. 2 illustrates some elements of the MCTF encoder 100 for carrying  
15 out these operations.

The elements of the MCTF encoder 100 shown in FIG. 2 include an estimator/predictor 102, an updater 103, and a decoder 105. The decoder 105 decodes an encoded stream received from the base layer encoder 150, and enlarges decoded small-screen frames to  
20 the size of frames in the enhanced layer using an internal scaler 105a. The estimator/predictor 102 searches for a reference block of each macroblock in a current frame, which is to be coded into residual data, in adjacent frames prior to or subsequent to the current frame and in frames enlarged by the scaler 105a. The  
25 estimator/predictor 102 then obtains an image difference (i.e., a pixel-to-pixel difference) of each macroblock in the current frame from the reference block or from a corresponding block in a temporally coincident frame enlarged by the scaler 105a, and codes the image difference into the macroblock. The  
30 estimator/predictor 102 also obtains a motion vector originating from the macroblock and extending to the reference block. The updater 103 performs an update operation for a macroblock in the current frame, whose reference block has been found in frames



prior to or subsequent to the current frame, by multiplying the image difference of the macroblock by an appropriate constant (for example,  $1/2$  or  $1/4$ ) and adding the resulting value to the reference block. The operation carried out by the updater 103 is referred to as a 'U' operation, and a frame produced by the 'U' operation is referred to as an 'L' frame.

The estimator/predictor 102 and the updater 103 of FIG. 2 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 on the video frame. A frame (or slice) having an image difference, which is produced by the estimator/predictor 102, is referred to as an 'H' frame (or slice). The 'H' frame (or slice) contains data having high frequency components of the video signal. In the following description of the embodiments, the term 'picture' is used to indicate a slice or a frame, provided that the use of the term is technically feasible.

The estimator/predictor 102 divides each of the input video frames (or L frames obtained at the previous level) into macroblocks of a desired size. For each divided macroblock, the estimator/predictor 102 searches for a block, whose image is most similar to that of each divided macroblock, in previous/next neighbor frames of the enhanced layer and/or in base layer frames enlarged by the scaler 105a. That is, the estimator/predictor 102 searches for a macroblock temporally correlated with each divided macroblock. A block having the most similar image to a target image block has the smallest image difference from the target image block. The image difference of two image blocks is defined, for example, as the sum or average of pixel-to-pixel differences of the two image blocks. Of blocks having a threshold image difference or less from a target macroblock in the current frame, a block having the smallest image difference from the target macroblock is referred to as a reference block. A picture

including the reference block is referred to as a reference picture. For each macroblock of the current frame, two reference blocks (or two reference pictures) may be present in a frame (including a base layer frame) prior to the current frame, in a frame (including a base layer frame) subsequent thereto, or one in a prior frame and one in a subsequent frame.

If the reference block is found, the estimator/predictor 102 calculates and outputs a motion vector from the current block to the reference block. The estimator/predictor 102 also calculates and outputs pixel error values (i.e., pixel difference values) of the current block from pixel values of the reference block, which is present in either the prior frame or the subsequent frame, or from average pixel values of the two reference blocks, which are present in the prior and subsequent frames. The image or pixel difference values are also referred to as residual data.

If no macroblock having a desired threshold image difference or less from the current macroblock is found in the two neighbor frames (including base layer frames) via the motion estimation operation, the estimator/predictor 102 determines whether or not a frame in the same time zone as the current frame (hereinafter also referred to as a "temporally coincident frame") or a frame in a close time zone to the current frame (hereinafter also referred to as a "temporally close frame") is present in the base layer sequence. If such a frame is present in the base layer sequence, the estimator/predictor 102 obtains the image difference (i.e., residual data) of the current macroblock from a corresponding macroblock in the temporally coincident or close frame based on pixel values of the two macroblocks, and does not obtain a motion vector of the current macroblock with respect to the corresponding macroblock. The close time zone to the current frame corresponds to a time interval including frames that can be regarded as having the same image as the current frame. Information of this time interval is carried within an encoded

stream.

The above operation of the estimator/predictor 102 is referred to as a 'P' operation. When the estimator/predictor 102 performs the 'P' operation to produce an H frame by searching for  
5 a reference block of each macroblock in the current frame and coding each macroblock into residual data, the estimator/predictor 102 can selectively use, as reference pictures, enlarged pictures of the base layer received from the scaler 105a, in addition to neighbor L frames of the enhanced layer  
10 prior to and subsequent to the current frame, as shown in FIG. 3a.

In an example embodiment of the present invention, five frames are used to produce each H frame. FIG. 3b shows five frames that can be used to produce an H frame. As shown, a current L  
15 frame 400L has L frames 401 prior to and L frames 402 subsequent to the current L frame 400L. The current L frame 400L also has a base layer frame 405 in the same time zone. One or two frames from among the L frames 401 and 402 in the same MCTF level as a current L frame 400L, the frame 405 of the base layer in the same  
20 time zone as the L frame 400L, and base layer frames 403 and 404 prior to and subsequent to the frame 405 are used as reference pictures to produce an H frame 400H from the current L frame 400L. As will be appreciated from the above discussion, there are various reference block selection modes. To inform the decoder  
25 of which mode is employed, the MCTF encoder 100 transmits 'reference block selection mode' information to the texture coding unit 110 after inserting/writing it into a field at a specified position of a header area of a corresponding macroblock.

When a picture of the base layer is selected as a reference  
30 picture for prediction of a picture of the enhanced layer in the reference picture selection method as shown in FIG. 3b, all or part of the base layer picture can be used for prediction of the enhanced layer picture. For example, as shown in Fig. 4, when

a base layer picture has an aspect ratio of 4:3, an actual image portion 502 of the base layer picture has an aspect ratio of 16:9, and an enhanced layer picture 500 has an aspect ratio of 16:9, upper and lower horizontal portions 501a and 501b of the base layer picture contain invalid data. In this case, only the image portion 502 of the base layer picture is used for prediction of the enhanced layer picture 500. To accomplish this, the scaler 105a selects (or crops) the image portion 502 of the base layer picture (S41), up-samples the selected image portion 502 to enlarge it to the size of the enhanced layer picture 500 (S42), and provides the enlarged image portion to the estimator/predictor 102.

The MCTF encoder 100 incorporates position information of the selected portion of the base layer picture into a header of the current picture coded into residual data. The MCTF encoder 100 also sets and inserts a flag "flag\_base\_layer\_cropping", which indicates that part of the base layer picture has been selected and used, in the picture header at an appropriate position so that the flag is delivered to the decoder. The position information is not transmitted when the flag "flag\_base\_layer\_cropping" is reset.

FIGS. 5a and 5b illustrate embodiments of the structure of information regarding a selected portion 512 of a base layer picture. In the embodiment of FIG. 5a, the selected portion 512 of the base layer picture is specified by offsets (left\_offset, right\_offset, top\_offset, and bottom\_offset) from the left, right, top and bottom boundaries of the base layer picture. The left offset indicates a position offset between left side pixels (or, for example, at least one pixel) in the base layer image and left side pixels in the selected portion 512. The top offset indicates a position offset between top side pixels (or, for example, at least one pixel) in the base layer image and top side pixels in the selected portion 512. The right offset indicates a position offset between right side pixels (or, for example, at least one

pixel) in the base layer image and right side pixels in the selected portion 512. The bottom side offset indicates a position offset between bottom side pixels (or, for example, at least one pixel) in the base layer image and bottom side pixels in the selected portion 512. In the embodiment of FIG. 5b, the selected portion 512 of the base layer picture is specified by offsets (left\_offset and top\_offset) from the left and top boundaries of the base layer picture and by the width and height (crop\_width and crop\_height) of the selected portion 512. Various other specifying methods are also possible.

The offsets in the information of the selected portion shown in FIGS. 5a and 5b may have negative values. For example, as shown in Fig. 6, when a base layer picture has an aspect ratio of 4:3, an enhanced layer picture 600 has an aspect ratio of 16:9, and an actual image portion of the picture has an aspect ratio of 4:3, the left and right offset values (left\_offset and right\_offset) have negative values  $-d_L$  and  $-d_R$ . Portions 601a and 601b extended from the base layer picture are specified by the negative values  $-d_L$  and  $-d_R$ . The extended portions 601a and 601b are padded with offscreen data, and a picture 610 including the extended portions 601a and 601b is upsampled to have the same size as that of the enhanced layer picture 600. Accordingly, data of an area 611 in the enlarged base layer picture, which corresponds to an actual image portion of the enhanced layer picture 600, can be used for prediction of the actual image portion of the enhanced layer picture 600.

Since the offset fields of the information illustrated in FIGS. 5a and 5b may have negative values, the same advantages as described above in the example of FIG. 4 can be achieved by using the information of FIGS. 5a and 5b as position information of an area overlapping with the enhanced layer picture, which is to be associated with the enlarged base layer picture, instead of using the information of FIGS. 5a and 5b for specifying the selected

area in the base layer picture.

Specifically, with reference to Fig. 7, when a base layer picture 702 is upsampled so that an actual image area 701 of the base layer picture 702 is enlarged to the size of an enhanced layer picture 700, the enlarged (e.g., up-sampled) picture corresponds to an area larger than the enhanced layer picture 700. In this example, top and bottom offsets  $top\_offset$  and  $bottom\_offset$  are included in the position information of an area overlapping with the enhanced layer picture 700. These offsets correspond to the enlarged base layer picture, and are assigned negative values  $-d_T$  and  $-d_B$  so that only an actual image area of the enlarged base layer picture is used for prediction of the enhanced layer picture 700. In the example of FIG. 7, left and right offsets of the position information of the area corresponding to the enlarged base layer picture are zero. However, it will be understood that the left and right offsets may be non-zero, and also correspond to the enlarged base layer picture. It will also be appreciated that a portion of the image in the enlarged base layer picture may not be used in determining the enhanced layer picture. Similarly, when the offset information corresponds to the base layer picture, as opposed to the up-sample base layer picture, a portion of the image in the base layer picture may not be used in determining the enhanced layer picture.

Furthermore, in this embodiment, the left offset indicates a position offset between left side pixels (or, for example, at least one pixel) in the up-sampled base layer image and left side pixels in the enhanced layer image. The top offset indicates a position offset between top side pixels (or, for example, at least one pixel) in the up-sampled base layer image and top side pixels in the enhanced layer image. The right offset indicates a position offset between right side pixels (or, for example, at least one pixel) in the up-sampled base layer image and right side pixels in the enhanced layer image. The bottom side offset indicates a

position offset between bottom side pixels (or, for example, at least one pixel) in the up-sampled base layer image and bottom side pixels in the enhanced layer image.

As described above, the information of FIGS. 5a and 5b can be used as information for selection of a portion of a base layer picture, which is to be used for prediction of an enhanced layer picture, or can be used as position information of an area overlapping with an enhanced layer picture, which is to be associated with a base layer picture for use in prediction of the enhanced layer picture.

Information of the size and aspect ratio of the base layer picture, mode information of an actual image of the base layer picture, etc., can be determined by decoding, for example, from a sequence header of the encoded base layer stream. Namely, the information may be recorded in the sequence header of the encoded base layer stream. Accordingly, the position of an area overlapping with the enhanced layer picture, which corresponds to the base layer picture or the selected area in the base layer picture described above, are determined based on position or offset information, and all or part of the base layer picture is used to suit this determination.

Returning to Figs. 1 and 2, the MCTF encoder 100 generates a sequence of H frames and a sequence of L frames, respectively, by performing the 'P' and 'U' operations described above on a certain-length sequence of pictures, for example, on a group of pictures (GOP). Then, an estimator/predictor and an updater at a next serially-connected stage (not shown) generates a sequence of H frames and a sequence of L frames by repeating the 'P' and 'U' operations on the generated L frame sequence. The 'P' and 'U' operations are performed an appropriate number of times (for example, until one L frame is produced per GOP) to produce a final enhanced layer sequence.

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 in the enhanced and/or base layer according to the method described below.

5        FIG. 8 is a block diagram of an apparatus for decoding a data stream encoded by the apparatus of FIG. 1. The decoding apparatus of FIG. 8 includes a demuxer (or demultiplexer) 200, a texture decoding unit 210, a motion decoding unit 220, an MCTF decoder 230, and a base layer decoder 240. The demuxer 200 separates a  
10 received data stream into a compressed motion vector stream, a compressed macroblock information stream, and a base layer 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  
15 vector stream to its original uncompressed state. The MCTF decoder 230 is an enhanced layer decoder which converts the uncompressed macroblock information stream and the uncompressed motion vector stream back to an original video signal according to an MCTF scheme. The base layer decoder 240 decodes the base  
20 layer stream according to a specified scheme, for example, according to the MPEG-4 or H.264 standard.

The MCTF decoder 230 includes, as an internal element, an inverse filter that has a structure as shown in FIG. 9 for reconstructing an input stream to its original frame sequence.

25        FIG. 9 shows some elements of the inverse filter for reconstructing a sequence of H and L frames of MCTF level N to a sequence of L frames of level N-1. The elements of the inverse filter of FIG. 9 include an inverse updater 231, an inverse predictor 232, a motion vector decoder 235, an arranger 234, and  
30 a scaler 230a. The inverse updater 231 subtracts pixel difference values of input H frames from corresponding pixel values of input L frames. The inverse predictor 232 reconstructs input H frames to frames having original images with reference to the L frames,



from which the image differences of the H frames have been subtracted, and/or with reference to enlarged pictures output from the scaler 240a. The motion vector decoder 235 decodes an input motion vector stream into motion vector information of each block and provides the motion vector information to an inverse predictor (for example, the inverse predictor 232) of each stage. The arranger 234 interleaves the frames completed by the inverse predictor 232 between the L frames output from the inverse updater 231, thereby producing a normal video frame sequence. The scaler 230a enlarges small-screen pictures of the base layer to the enhanced layer picture size, for example, according to the information as shown in FIGS. 5a and 5b.

The L frames output from the arranger 234 constitute an L frame sequence 601 of level N-1. A next-stage inverse updater and predictor of level N-1 reconstructs the L frame sequence 601 and an input H frame sequence 602 of level N-1 to an L frame sequence. This decoding process is performed the same number of times as the number of MCTF levels employed in the encoding procedure, thereby reconstructing an original video frame sequence. With reference to 'reference\_selection\_code' information carried in a header of each macroblock of an input H frame, the inverse predictor 232 specifies an L frame of the enhanced layer and/or an enlarged frame of the base layer which has been used as a reference frame to code the macroblock to residual data. The inverse predictor 232 determines a reference block in the specified frame based on a motion vector provided from the motion vector decoder 235, and then adds pixel values of the reference block (or average pixel values of two macroblocks used as reference blocks of the macroblock) to pixel difference values of the macroblock of the H frame; thereby reconstructing the original image of the macroblock of the H frame.

When a base layer picture has been used as a reference frame of a current H frame, the scaler 230a selects and enlarges an area

in the base layer picture (in the example of FIG. 4) or enlarges a larger area than the base layer picture (in the example of FIG. 6) based on positional relationship information as shown in FIGS. 5a and 5b included in a header analyzed by the MCTF decoder 230 so that the enlarged area of the base layer picture is used for reconstructing macroblocks containing residual data in the current H frame to original image blocks as described above. The positional relationship information is extracted from the header and is then referred to when information indicating whether or not the positional relationship information is included (specifically, the flag "flag\_base\_layer\_cropping" in the example of FIGS. 5a and 5b) indicates that the positional relationship information is included.

In the case where the information of FIGS. 5a and 5b has been used as information indicating the position of an area overlapping with an enhanced layer picture, to use in prediction of the enhanced layer picture, the inverse predictor 232 uses an enlarged one of the base layer picture received from the scaler 230a for prediction of the enhanced layer picture by associating the entirety of the enlarged base layer picture with all or part of the current H frame or with a larger area than the current H frame according to the values (positive or negative) of the offset information. In the case of FIG. 7 where the enlarged base layer picture is associated with a larger area than the current H frame, the predictor 232 uses only an area in the enlarged base layer picture, which corresponds to the H frame, for reconstructing macroblocks in the current H frame to their original images. In this example, the offset information included negative values.

For one H frame, the MCTF decoding is performed in specified units, for example, in units of slices in a parallel fashion, so that the macroblocks in the frame have their original images reconstructed and the reconstructed macroblocks are then combined to constitute a complete video frame.

The above decoding method reconstructs an MCTF-encoded data stream to a complete video frame sequence. The decoding apparatus decodes and outputs a base layer sequence or decodes and outputs an enhanced layer sequence using the base layer depending on its  
5 processing and presentation capabilities.

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

As is apparent from the above description, a method and apparatus for encoding/decoding a video signal according to the  
10 present invention uses pictures of a base layer provided for low-performance decoders, in addition to pictures of an enhanced layer, when encoding a video signal in a scalable fashion, so that the total amount of coded data is reduced, thereby increasing coding efficiency. In addition, part of a base layer  
15 picture, which can be used for a prediction operation of an enhanced layer picture, is specified so that the prediction operation can be performed normally without performance degradation even when a picture enlarged from the base layer picture cannot be directly used for the prediction operation of  
20 the enhanced layer picture.

Although this invention has been described with reference to the example embodiments, it will be apparent to those skilled in the art that various improvements, modifications, replacements, and additions can be made in the invention without departing from  
25 the scope and spirit of the invention. Thus, it is intended that the invention cover the improvements, modifications, replacements, and additions 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 reference image, offset  
5 information and dimension information, the offset information  
indicating a position offset between at least one boundary pixel  
of the reference image and at least one boundary pixel of the  
current image, and the dimension information indicating at least  
one dimension of the current image.  
10
2. The method of claim 1, wherein the reference image is  
based on a base image in a base layer.
3. The method of claim 2, wherein the reference image is at  
15 least an up-sampled portion of the base image.
4. The method of claim 1, wherein the offset information  
includes left offset information indicating a position offset  
between at least one left side pixel of the reference image and  
20 at least one left side pixel of the current image.
5. The method of claim 4, wherein the reference image is  
based on a base image in a base layer.
- 25 6. The method of claim 5, wherein the reference image is at  
least an up-sampled portion of the base image.
7. The method of claim 4, wherein the offset information  
includes top offset information indicating a position offset  
30 between at least one top side pixel of the reference image and  
at least one top side pixel of the current image.

8. The method of claim 7, wherein the reference image is based on a base image in a base layer.

5        9. The method of claim 8, wherein the reference image is at least an up-sampled portion of the base image.

10       10. The method of claim 7, wherein the dimension information includes width information indicating a width of the current image.

15       11. The method of claim 7, wherein the dimension information includes height information indicating a height of the current image.

12. The method of claim 7, wherein the dimension information includes width information indicating a width of the current image and height information indicating a height of the current image.

20       13. The method of claim 1, wherein the offset information includes top offset information indicating a top position offset between at least one top side pixel of the reference image and at least one top side pixel of the current image.

25       14. The method of claim 13, wherein the reference image is based on a base image in a base layer.

15       15. The method of claim 14, wherein the reference image is at least an up-sampled portion of the base image.

30       16. The method of claim 1, wherein the dimension information includes width information indicating a width of the current image.

17. The method of claim 1, wherein the dimension information includes height information indicating a height of the current image.

5

18. The method of claim 1, wherein the dimension information includes width information indicating a width of the current image and height information indicating a height of the current image.

10 19. The method of claim 1, wherein the predicting step obtains the offset information from a header for at least a portion of a picture in the current layer.

20. The method of claim 19, wherein the predicting step  
15 determines that the offset information is present based on an indicator in the header.

21. The method of claim 20, wherein the picture is a slice.

20 22. The method of claim 20, wherein the picture is a frame.

23. The method of claim 1, further comprising:  
up-sampling at least a portion of a base image in a base layer  
to produce an up-sampled image as the reference image.

25

24. The method of claim 23, wherein  
the offset information includes left offset information  
indicating a position offset between at least one left side pixel  
of the reference image and at least one left side pixel of the  
30 current image, and top offset information indicating a top  
position offset between at least one top side pixel of the  
reference image and at least one top side pixel of the current  
image; and

the dimension information includes width information indicating a width of the current image and height information indicating a height of the current image.

5           25. The method of claim 24, wherein the predicting step obtains the offset information from a header for at least a portion of a picture in the current layer.

26. An apparatus for decoding a video signal, comprising:  
10           a decoder predicting at least a portion of a current image in a current layer based on at least a portion of a reference image, offset information and dimension information, the offset information indicating a position offset between at least one boundary pixel of the reference image and at least one boundary  
15 pixel of the current image, and the dimension information indicating at least one dimension of the current image.

FIG. 1

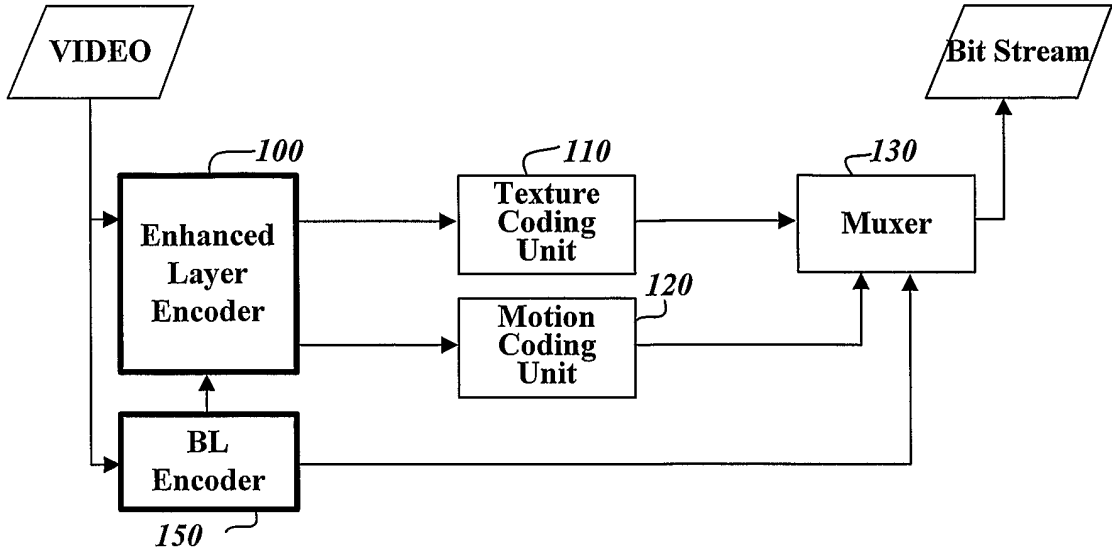
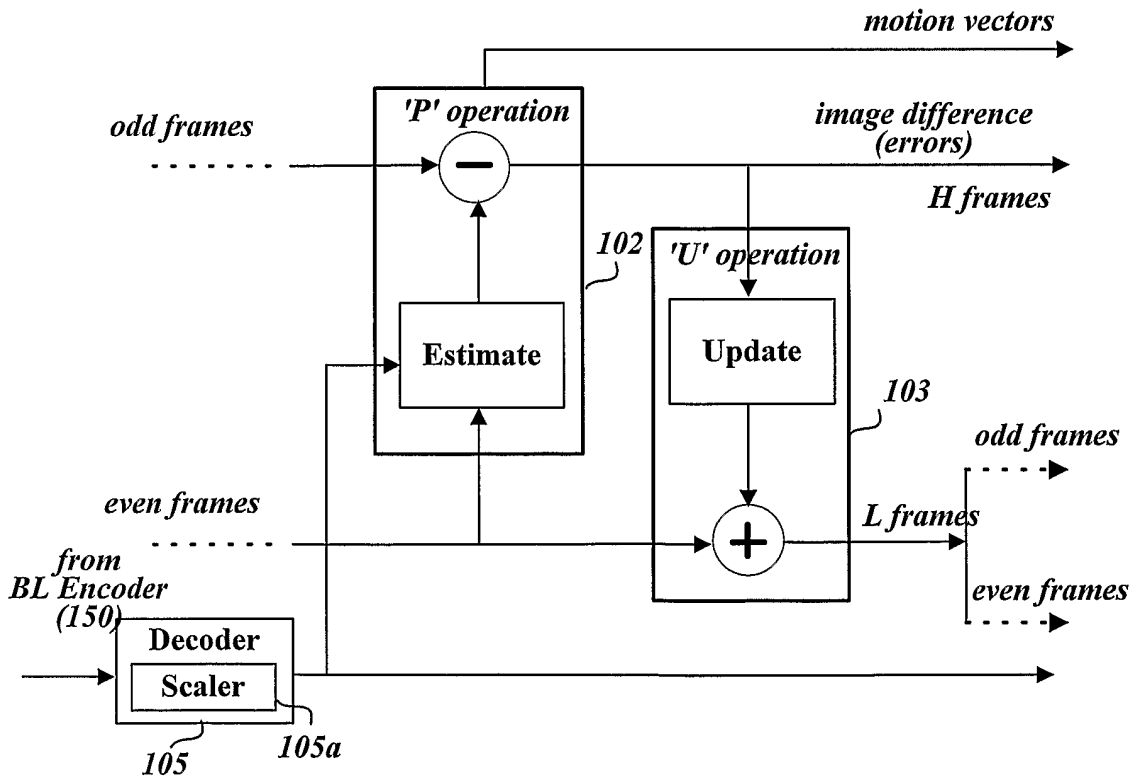


FIG. 2





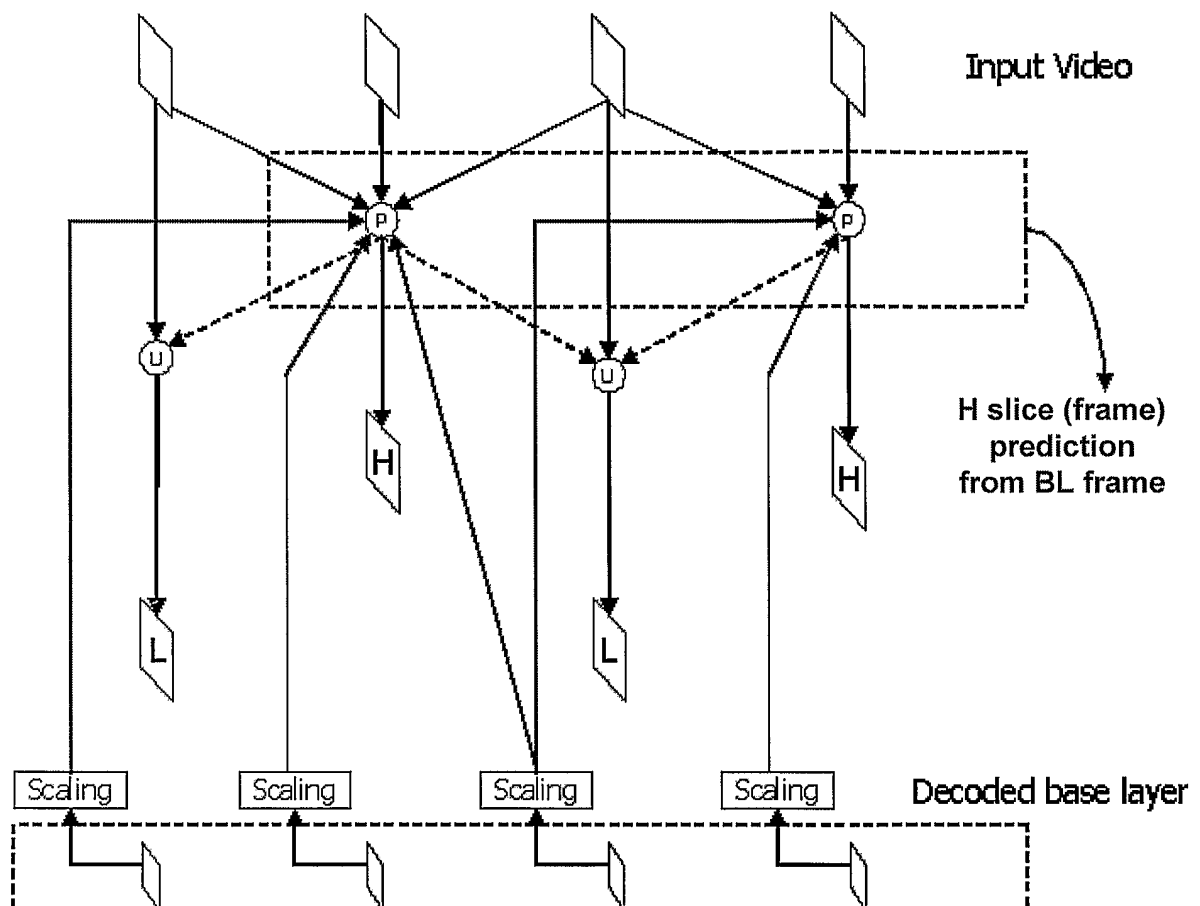
**FIG. 3a**

FIG. 3b

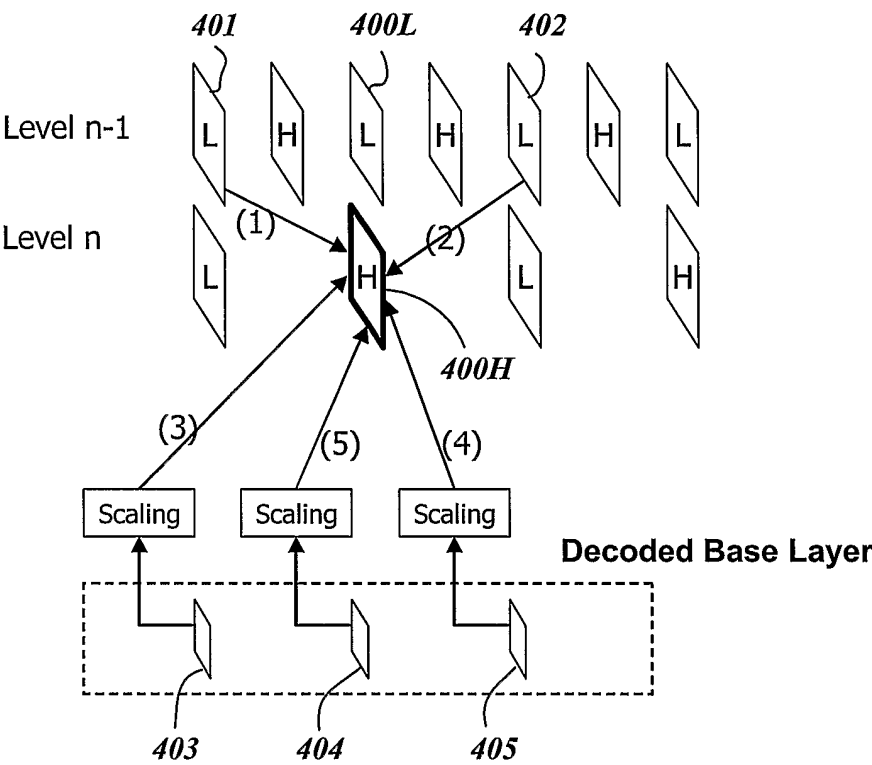


FIG. 4

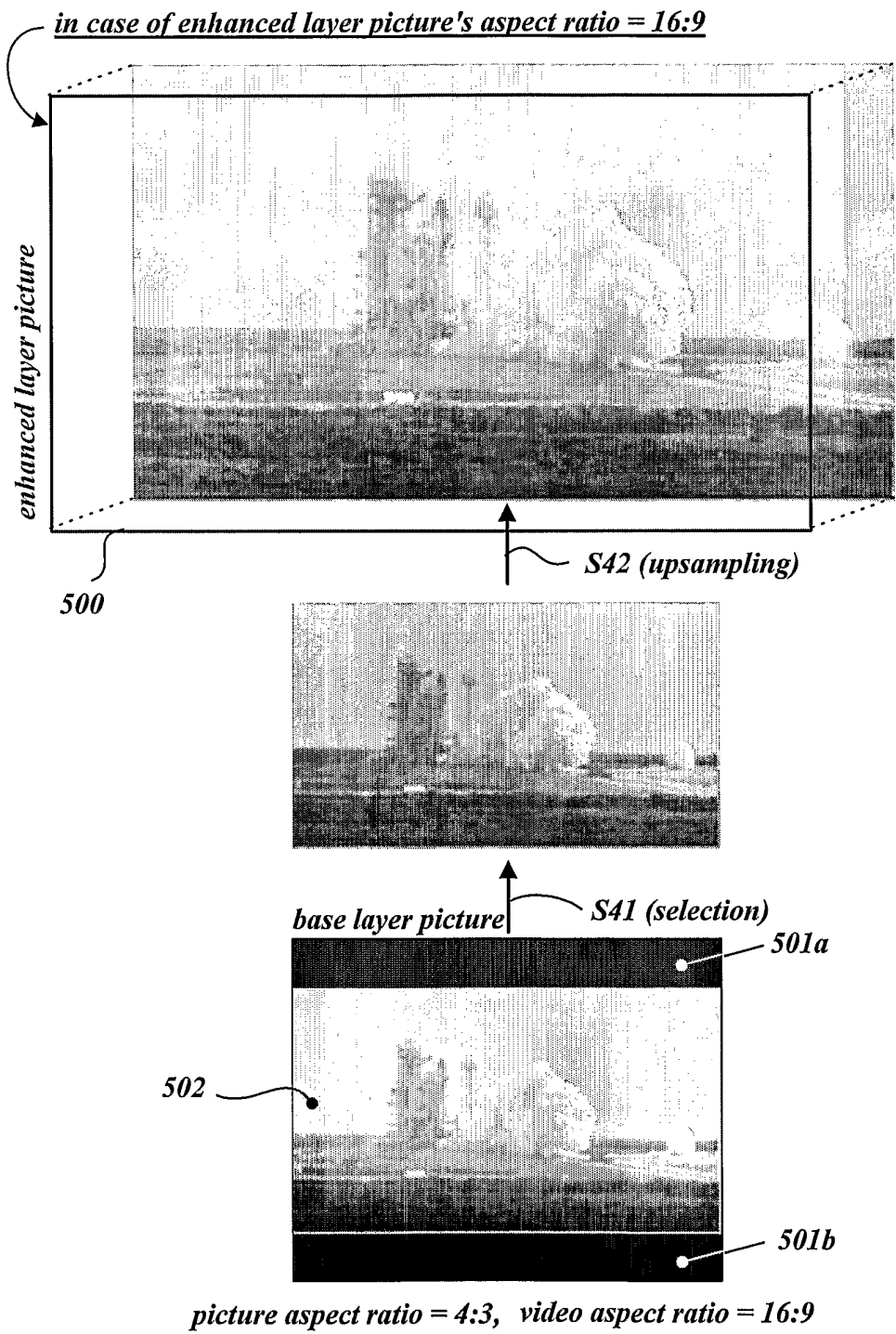


FIG. 5a

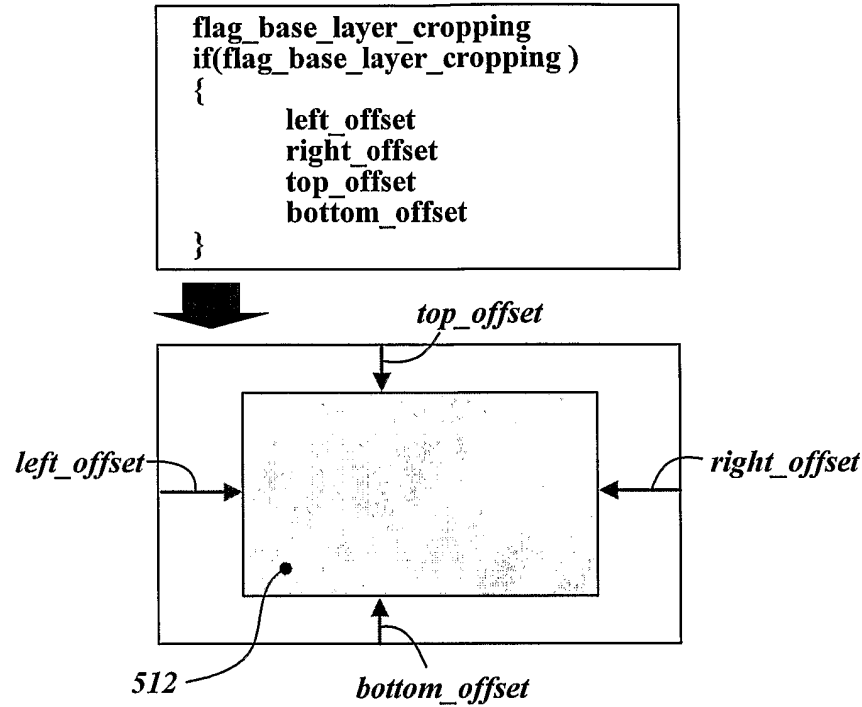


FIG. 5b

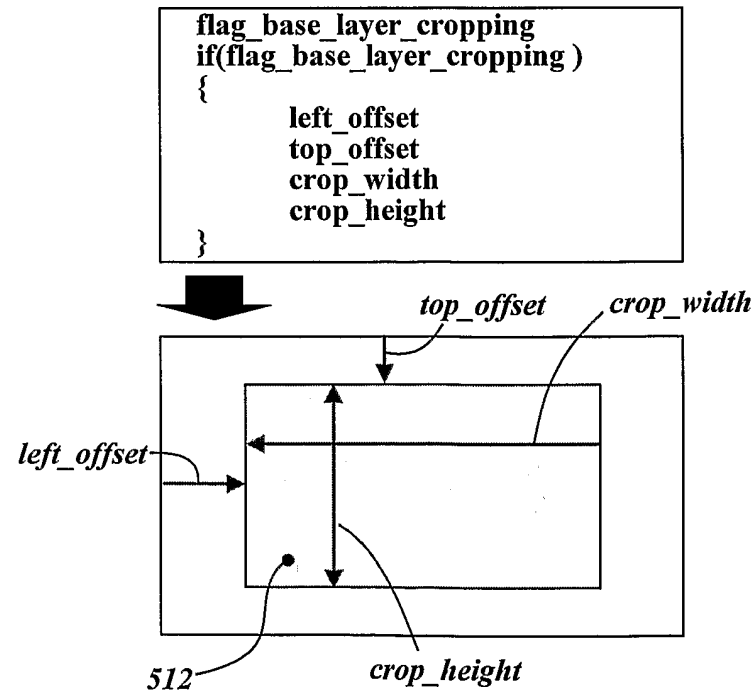
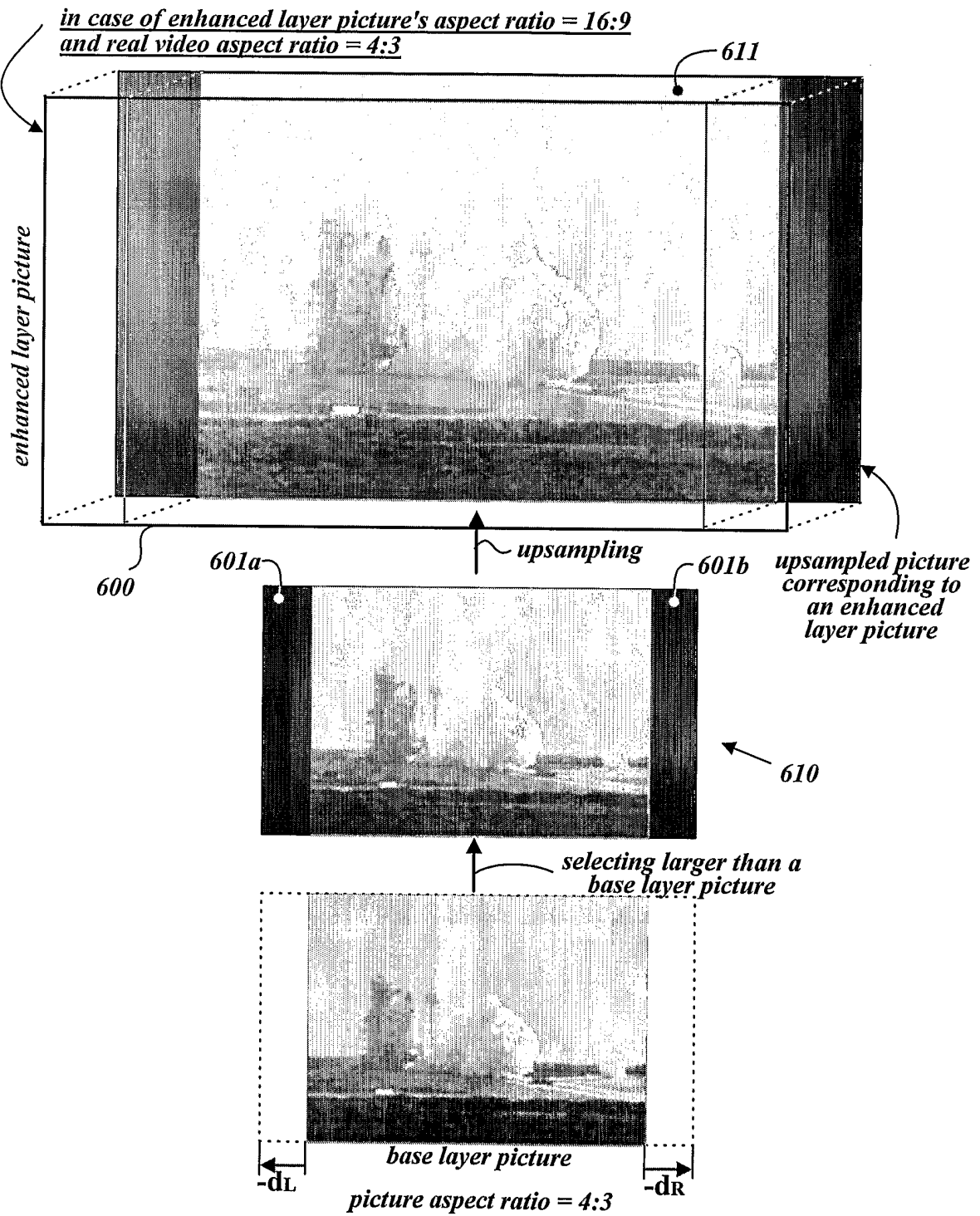
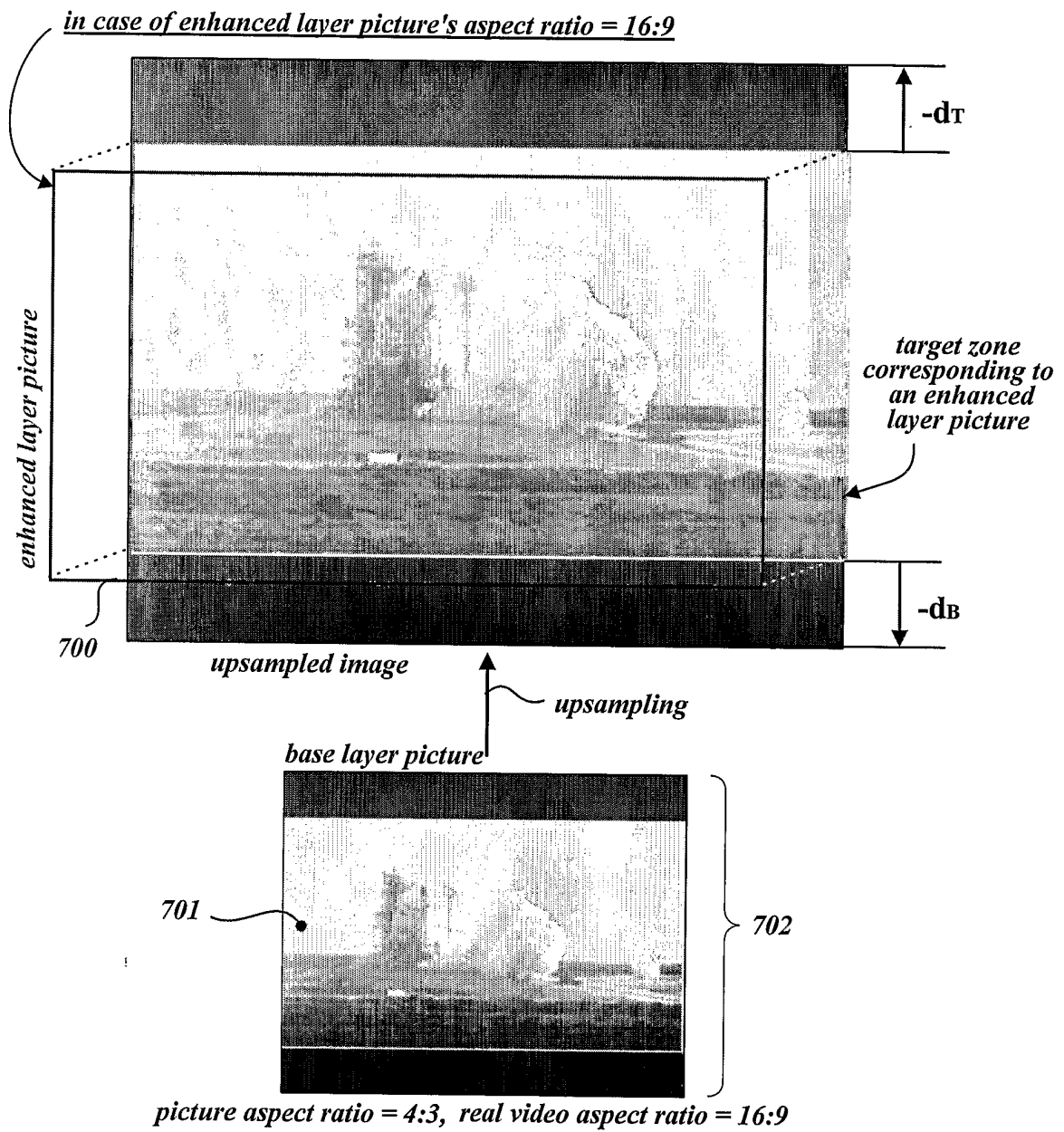


FIG. 6



**FIG. 7**

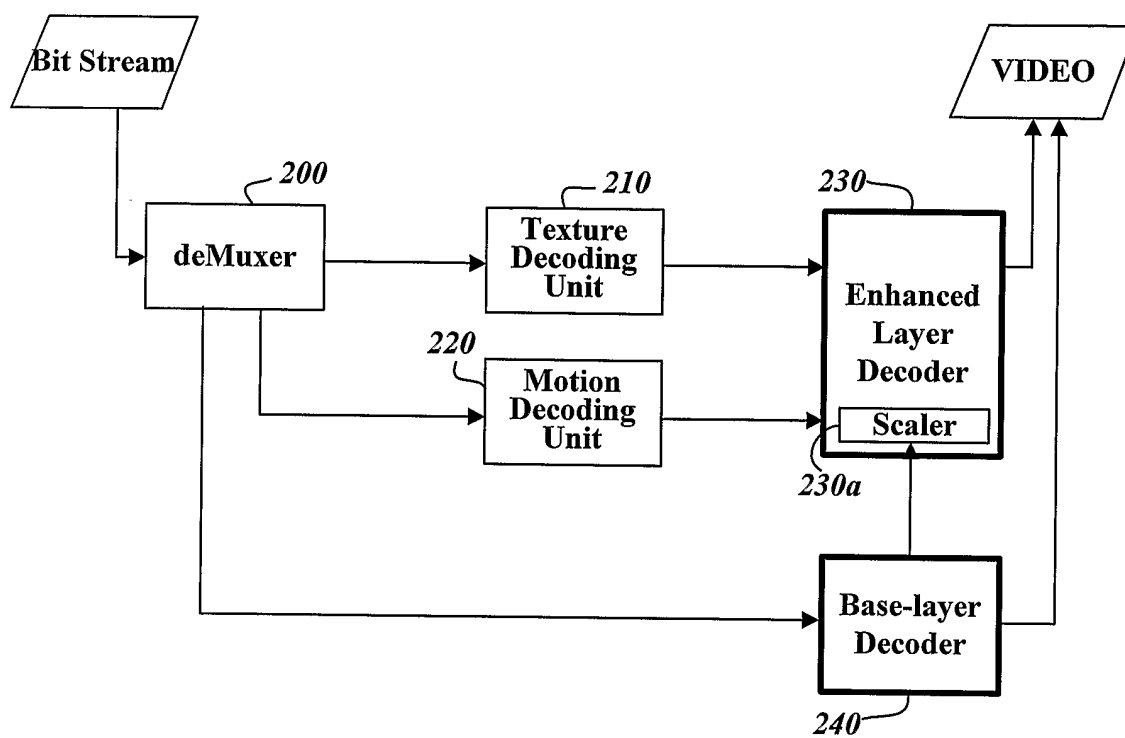
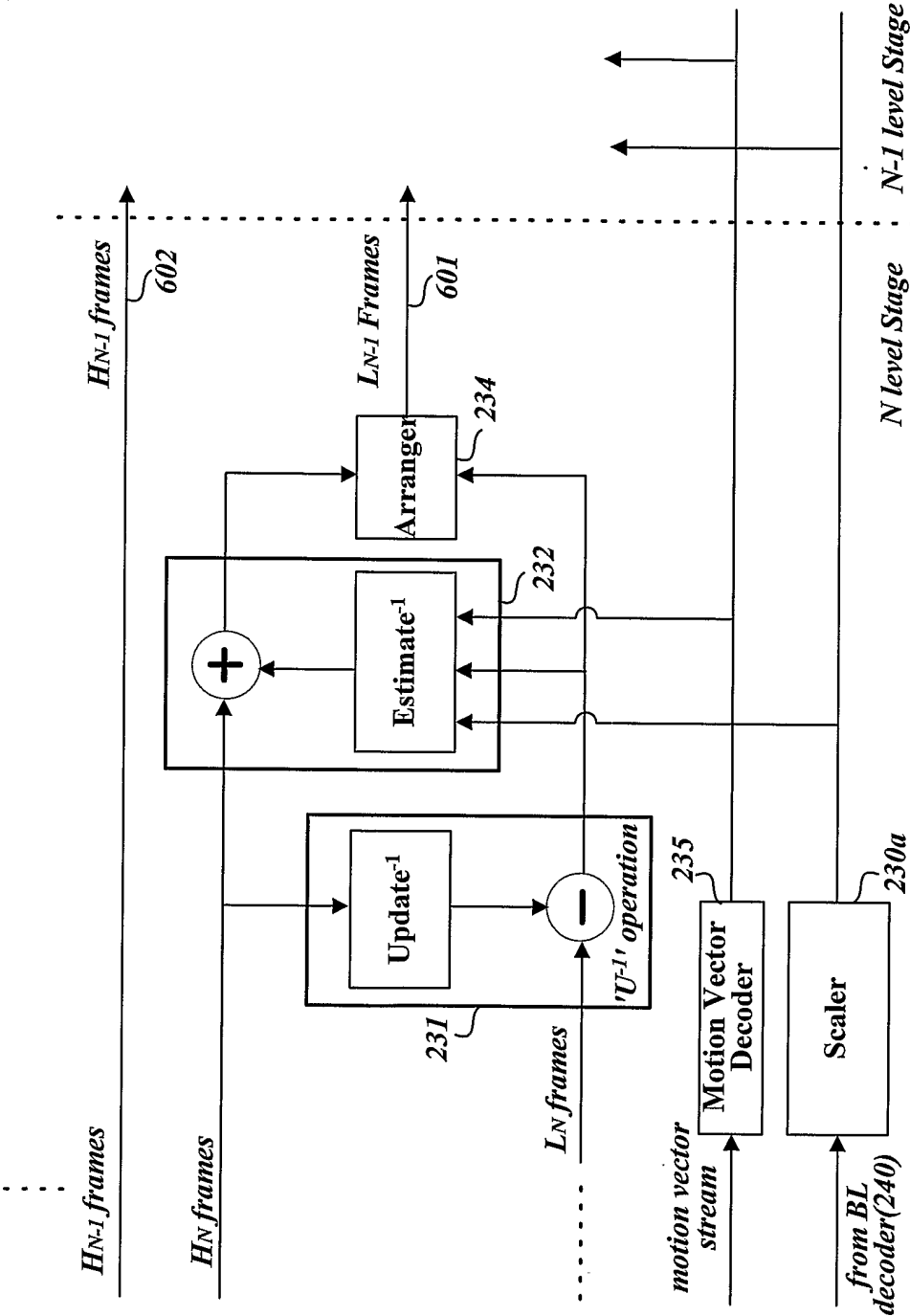
**FIG. 8**

FIG. 9





## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/KR2006/001341**A. CLASSIFICATION OF SUBJECT MATTER****H04N 7/32(2006.01);**

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 applications for Utility models since 1975

Japanese Utility models and application for Utility models since 1975

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

eKIPASS (KIPO internal), IEEE Xplore: "video, coding, scalable, layer, base, enhancement, offset, position, MCTF"

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 6535559 B2 (YAGASAKI and SUZUKI) 18 Mar. 2003 See column 5,10-13; figures 3-12.	1-26
A	US 6510177 B1 (BONET et al.) 21 Jan. 2003 See abstract; claims 1-54; figures 3-11B.	1-26
A	US 5973739 A (NILSSON) 26 Oct. 1999 See abstract; claim 1.	1-26
A	US 2004/0114689 A1 (ZHANG and BOSSEN) 17 Jun. 2004 See abstract.	1-26



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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

21 JULY 2006 (21.07.2006)

Date of mailing of the international search report

**24 JULY 2006 (24.07.2006)**

Name and mailing address of the ISA/KR

Korean Intellectual Property Office  
920 Dunsan-dong, Seo-gu, Daejeon 302-701,  
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

KIM, Heung Soo

Telephone No. 82-42-481-5764



**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

PCT/KR2006/001341

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US06535559	18.03.2003	AU199865202B2	22.10.1998
		AU6520298A1	22.10.1998
		AU732452B2	26.04.2001
		CA2255923AA	08.10.1998
		CA2255923C	07.06.2005
		CA2255923C	08.10.1998
		CA2421090AA	08.10.1998
		CA2421090C	08.02.2005
		CA2421090A1	08.10.1998
		CA2421090C	08.10.1998
		CN1185876C	19.01.2005
		CN1186944C	26.01.2005
		CN1220804	23.06.1999
		CN1312655	12.09.2001
		CN1312655A	12.09.2001
		CN1630375A	22.06.2005
		EP00914007A1	06.05.1999
		EP01152622A1	07.11.2001
		EP0914007A1	06.05.1999
		EP1152622A1	07.11.2001
		EP914007A1	06.05.1999
		EP914007A4	07.11.2001
		HK1043461A1	19.08.2005
		ID20680A	11.02.1999
		IL127274A0	22.09.1999
		JP10336669A2	18.12.1998
		JP10336669	18.12.1998
		JP13036911	09.02.2001
		JP2001036911A2	09.02.2001
		JP3380980B2	24.02.2003
		JP3380983B2	24.02.2003
		KR1020000016220	25.03.2000
		KR1020017007032	05.06.2001
		KR2000016220A	25.03.2000
		TW398150B	11.07.2000
		TW398150A	11.07.2000
		US06414991	02.07.2002
		US06643328	04.11.2003
		US20020114391A1	22.08.2002
		US20020118750A1	29.08.2002
		US20020122486A1	05.09.2002
		US2002114391A1	22.08.2002
		US2002114391AA	22.08.2002
		US2002118750A1	29.08.2002
		US2002118750AA	29.08.2002
		US2002122486A1	05.09.2002
		US2002122486AA	05.09.2002
		US20030133502A1	17.07.2003
		US2003133502A1	17.07.2003
		US2003133502AA	17.07.2003

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

PCT/KR2006/001341

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US06510177	21.01.2003	US6414991BA	02.07.2002
		US6535559BB	18.03.2003
		US6643328BB	04.11.2003
		W09844742A1	08.10.1998
US06510177	21.01.2003	US6510177B1	21.01.2003
		US6510177BA	21.01.2003
US05973739	26.10.1999	AU3765693A1	08.11.1993
		AU667349B2	21.03.1996
		CA2132350C	28.07.1998
		DE69322079C0	17.12.1998
		DE69322079T2	29.04.1999
		EP00636301B1	11.11.1998
		EP0636301A1	01.02.1995
		EP636301A1	01.02.1995
		EP636301B1	11.11.1998
		FI944406A	23.09.1994
		FI944406A0	23.09.1994
		GB9206860A0	13.05.1992
		JP7505273T2	08.06.1995
		KR1019950701179	20.02.1995
		N0943572A	23.11.1994
		N0943572A0	26.09.1994
		NZ249959A	26.07.1995
		SG47573A1	17.04.1998
		US5973739A	26.10.1999
		W09320650A1	14.10.1993
US20040114689A1	17.06.2004	JP16201315	15.07.2004
		JP2004201315A2	15.07.2004
		US2004114689A1	17.06.2004
		US2004114689AA	17.06.2004