[19]	INTELLECTUAL PROPERTY PHILIPPINES			
[12]	INVENTION PUBLICATION			
[11]	Publication Number:	12014502472	Document Code:	B1
[22]	Publication Date:	22/12/2014		
[21]	Application Number:	12014502472	Document Code:	A
[22]	Date Filed:	6/11/2014		
[54]	Title:	VIDEO PREDICTIVE ENCODING DEVICE, VIDEO PREDICTIVE ENCODING METHOD, VIDEO PREDICTIVE ENCODING PROGRAM, VIDEO PREDICTIVE DECODING DEVICE, VIDEO PREDICTIVE DECODING METHOD, AND VIDEO PREDICTIVE DECODING PROGRAM		
[71]	Applicant(s):	NTT DOCOMO INC		
[72]	Inventor(s):	JUNYA TAKIUE SENG BOON CHOONG KENG TAN THIOW		
[30]	Priority Data:	6/7/2012 JP20120152700		
[51]	International Class 8:			

[57]	Abstract:	Conventional methods for a NAL unit header have an inefficient design in which bits are allocated to each of nal_ref_flag and nal_unit_type even if the value of nal_ref_flag is uniquely determined according to the value of nal_unit_type. The video predictive encoding device according to the present invention is equipped with an input means that inputs a plurality of images that constitute a video, and an encoding means that encodes the plurality of images using either an intra prediction method or an inter prediction method, generates compressed image data, and packetizes the compressed image data together with packet header information. The packet header information includes picture type. The encoding means determines the picture type to uniquely show whether an encoded picture data is used as a reference when decoding another picture.		

picture data is used for reference in decoding of another picture. An encoding module of a video predictive encoding program according to an aspect of the present invention determines the picture type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture in the same temporal layer.

A video predictive decoding program according to an aspect of the present invention is a video predictive decoding program comprising: an input module which inputs compressed picture data resulting from encoding of a plurality of pictures forming a video sequence by either intra prediction or inter prediction and packetization of the compressed picture data with packet header information; and a decoding module which reconstructs the packet header information and the compressed picture data, wherein the packet header information contains a picture type uniquely indicating whether reconstructed picture data is used for reference in decoding of another picture, and wherein the decoding module determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture.

In the video predictive decoding program according to an aspect of the present invention the decoding module determines whether reconstructed picture data is used for reference in decoding of another picture, based on a correspondence table in which the picture type is previously stored in association with information indicative of whether reconstructed picture data is used for reference in decoding of another picture. A decoding module of a video predictive decoding program according to an aspect of the present invention determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture in the same temporal layer.

Effects of the Invention

An effect is to save the bits used for nal_ref_flag and enable use thereof as other indication information. This is more efficient utilization of the NAL unit header. Another utilization method is to enable extension of the NAL unit types from 6 bits to 7 bits. At present the existing NAL unit types are assigned to half of 64 values of nal_unit_type available and the other 32 values of nal_unit_type are reserved, and can be used in defining new NAL unit types in the future. By using three out of these reserved values of NAL unit types and extending the bit count of the NAL unit types to 7 bits, it becomes feasible to define 93 ($128 - 32 - 3 = 93$) further NAL units in the future.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing a video predictive encoding device according to an embodiment of the present invention.

Fig. 2 is a block diagram showing a video predictive decoding device according to an embodiment of the present invention.

Fig. 3 is a flowchart showing processing of a video predictive encoding method according to an embodiment of the present invention.

Fig. 4 is a flowchart showing a detailed part of processing of the video predictive encoding method according to an embodiment of the present invention.

Fig. 5 is a flowchart showing processing of a video predictive decoding method according to an embodiment of the present invention.

Fig. 6 is a flowchart showing a detailed part of processing of the video predictive decoding method according to an embodiment of the present invention.

Fig. 7 is a hardware configuration of a computer for executing a program stored in a storage medium.

Fig. 8 is a perspective view of a computer for executing a program stored in a storage medium.

Fig. 9 is a block diagram showing a configuration example of a video predictive encoding program.

Fig. 10 is a block diagram showing a configuration example of a video predictive decoding program.

Embodiments of the Invention

Embodiments of the present invention will be described below using Figs. 1 to 10.

First, a video predictive encoding method will be described. Fig. 1 is a block diagram showing a video predictive encoding device according to an embodiment of the present invention. Reference numeral 101 denotes an input terminal, 102 a block partition unit, 103 a predicted signal generation unit, 104 a frame memory, 105 a subtraction unit, 106 a transform unit, 107 a quantization unit, 108 a de-quantization unit, 109 an inverse transform unit, 110 an addition unit, 111 an entropy encoding unit, 112 an output terminal, and 113 an input terminal. The input terminal 101 corresponds to an input means. The subtraction unit 105, transform unit 106, quantization unit 107, and entropy encoding unit 111 correspond to an encoding means. The de-quantization unit 108, inverse transform unit 109, and addition unit 110 correspond to a decoding means.

Concerning the video predictive encoding device configured as described above, the operation thereof will be described below. A video signal consisting of a plurality of pictures is fed to the input terminal 101. A picture of an encoding target is partitioned into a plurality of regions by the block partition unit 102. In the embodiment according to the present invention, the target picture is partitioned into blocks each consisting of 8×8 pixels, but it may be partitioned into blocks of any size or shape other than the foregoing. A predicted signal is then generated for a region as a target of an encoding process (which will be referred to hereinafter as a target block). The embodiment according to the present invention employs two types of prediction methods. Namely, they are inter prediction and intra prediction.

In the inter prediction, reconstructed pictures having been encoded and thereafter reconstructed are used as reference pictures and motion information to provide the predicted signal with the smallest error from the target block is determined from the reference pictures. This process is called motion estimation. Depending upon situation, the target block may be further partitioned into sub-regions and an inter prediction method can be determined for each of the sub-regions. In this case, the most efficient partition method for the entire target block and motion information of each sub-region are determined out of various partition methods. In embodiments according to the present invention, the operation is carried out in the predicted signal generation unit 103, the target block is fed via line L102, and the reference pictures are fed via line L104. The reference pictures to be used herein are a plurality of pictures which have been encoded and reconstructed in the past. The details are the same as in the methods of MPEG-2 or 4 and H.264 which are the conventional technologies. The motion information and sub-region partition method determined as described above are fed via line L112 to the entropy encoding unit 111 to be encoded thereby and then the encoded data is output from the output terminal 112. Information (reference index) indicating from which reference picture the predicted signal is derived out of the plurality of reference pictures is also sent via line L112 to the entropy encoding unit 111. The predicted signal generation unit 103 derives reference picture signals from the frame memory 104, based on the reference pictures and motion information, corresponding to the sub-region partition method and each sub-region, and generates the predicted signal. The inter-predicted signal generated in this manner is fed via line L103 to the subtraction unit 105.

In the intra prediction, an intra-predicted signal is generated using reconstructed pixel values spatially adjacent to the target block. Specifically, the predicted signal generation unit 103 derives reconstructed pixel signals in the same frame from the frame memory 104 and extrapolates these signals to generate the intra-predicted signal. The information about the method of extrapolation is fed via line L112 to the entropy encoding unit 111 to be encoded thereby, and then the encoded data is output from the output terminal 112. The intra-predicted signal generated in this manner is fed to the subtraction unit 105. The method of generating the intra-predicted signal in the predicted signal generation unit 103 is the same as the method of H.264

being the conventional technology. The predicted signal with the smallest error is selected from the inter-predicted signals obtained as described above, and the selected predicted signal is fed to the subtraction unit 105.

The subtraction unit 105 subtracts the predicted signal (fed via line L103) from the signal of the target block (fed via line L102) to generate a residual signal. This residual signal is transformed by a discrete cosine transform by the transform unit 106 to form transform coefficients, which are quantized by the quantization unit 107. Finally, the entropy encoding unit 111 encodes the quantized transform coefficients and the encoded data is output along with the information about the prediction method from the output terminal 112.

For the intra prediction or the inter prediction of the subsequent target block, the compressed signal of the target block is subjected to inverse processing to be reconstructed. Namely, the quantized transform coefficients are inversely quantized by the de-quantization unit 108 and then transformed by an inverse discrete cosine transform by the inverse transform unit 109, to reconstruct a residual signal. The addition unit 110 adds the reconstructed residual signal to the predicted signal fed via line L103 to reconstruct a signal of the target block and the reconstructed signal is stored in the frame memory 104. The present embodiment employs the transform unit 106 and the inverse transform unit 109, but it is also possible in other embodiments to use other transform processing instead of these transform units. Depending upon the situation, in some embodiments the transform unit 106 and the inverse transform unit 109 may be omitted.

Input data from the input terminal 113 includes display order information of each picture, a type of encoding of each picture (intra predictive encoding, inter predictive encoding, or bidirectional predictive encoding), and information about the NAL unit type, and the predicted signal generation unit 103 operates based on these pieces of information. These pieces of information are also fed via line L113 to the entropy encoding unit 111 to be encoded thereby, and the encoded data is output from the output terminal 112. The operation of the entropy encoding unit 111 for encoding of the NAL unit type will be described later.

Next, a video predictive decoding method will be described. Fig. 2 is a block diagram showing a video predictive decoding device according to an embodiment of the present invention. Reference numeral 201 denotes an input terminal, 202 a data analysis unit, 203 a de-quantization unit, 204 an inverse transform unit, 205 an addition unit, 206 an output terminal, 207 a frame memory, 208 a predicted signal generation unit, and 209 a frame memory management unit. The input terminal 201 corresponds to an input means. The data analysis unit 202, de-quantization unit 203, inverse transform unit 204, and addition unit 205 correspond to a decoding means. In other embodiments, the decoding means may be means other than the foregoing. Furthermore, in embodiments the decoding means may be configured without the inverse transform unit 204.

Concerning the video predictive decoding device configured as described above, the operation thereof will be described below. Compressed data resulting from compression encoding by the video predictive encoding device is input through the input terminal 201. This compressed data contains the residual signal resulting from predictive encoding of each target block obtained by partitioning of a picture into a plurality of blocks, and the information related to the generation of the predicted signal. The information related to the generation of the predicted signal includes, in addition to the NAL unit type, the information about block partitioning (size of block), the motion information, and the aforementioned reference index in the case of the inter prediction, and includes the information about the extrapolation method from reconstructed surrounding pixels in the case of the intra prediction.

The data analysis unit 202 extracts the residual signal of the target block, the information related to the generation of the predicted signal including the NAL unit type, the quantization parameter, and the display order information of the picture from the compressed data. The operation for extraction of the NAL unit type in the data analysis unit 202 will be described later. The residual signal of the target block is inversely quantized on the basis of the quantization parameter (fed via line L202) by the de-quantization unit 203. The result is transformed by an inverse discrete cosine transform by the inverse transform unit 204.

Next, the information related to the generation of the predicted signal such as the display order information of the target picture, the encoding type of the picture, the NAL unit type, and the reference index is fed via line L206 to the predicted signal generation unit 208. The predicted signal generation unit 208 accesses the frame memory 207, based on the information related to the generation of the predicted signal, to derive a reference signal from a plurality of reference pictures (via line L207) and generate a predicted signal. This predicted signal is fed via line L208 to the addition unit 205, the addition unit 205 adds this predicted signal to the reconstructed residual signal to reconstruct a target block signal, and the signal is output via line L205 from the output terminal 206 and simultaneously stored into the frame memory 207.

Reconstructed pictures to be used for decoding and reconstruction of the subsequent picture are stored in the frame memory 207.

Table 2 and Table 3 are tables indicating choices of two types of syntaxes concerning use modes of two bytes of the NAL unit header.

Table 2

nal_unit(NumBytesInNALunit) {	Descriptor
forbidden_zero_bit	f(1)
reserved	u(1)
nal_unit_type	u(6)
temporal_id	u(3)
reserved_one_5bits	u(5)
.... (The rest of the NAL unit)	

Table 3

nal_unit(NumBytesInNALunit) {	Descriptor
forbidden_zero_bit	f(1)
nal_unit_type	u(7)
temporal_id	u(3)
reserved_one_5bits	u(5)
.... (The rest of the NAL unit)	

In Tables 2 and 3, numbers in parentheses in the Descriptor column indicate bit counts of corresponding items.

In the NAL unit header syntax of Table 2, nal_ref_flag is replaced by a reserved bit (reserved). This bit is ignored by currently existing decoding devices, but it can be assigned a new meaning or semantics for future decoding devices. It is noted that the bit arrangement in Table 2 is just for description and the reserved bit may be located at another place in the 2-byte header.

In the NAL unit header syntax of Table 3, nal_unit_type is assigned 7 bits and at most 128 different kinds of nal_unit_type can be defined thereby. In the present embodiment the assignment of 7 bits to nal_unit_type was selected, but the bit saved in nal_ref_flag may be assigned to temporal_id.

Table 4 shows the NAL unit types in the present embodiment.

Table 4

nal_unit_type	Category	Content of NAL unit and RBSP syntax structure	nal_ref_flag
0		Unspecified	-
1	Other slice	Coded slice of a non-RAP, non-TFD and non-TLA picture slice_layer_rbsp( )	0
2	TFD slice	Coded slice of a TFD picture slice_layer_rbsp( )	0
3	TLA slice	Coded slice of a non-TFD TLA picture slice_layer_rbsp( )	0
4	RAP slice	Coded slice of a CRAT picture slice_layer_rbsp( )	1
5	RAP slice	Coded slice of an CRANT picture slice_layer_rbsp( )	1
6	RAP slice	Coded slice of a BLCT picture slice_layer_rbsp( )	1
7	RAP slice	Coded slice of a BLCNT picture slice_layer_rbsp( )	1
8	RAP slice	Coded slice of an IDR picture slice_layer_rbsp( )	1
9	Other slice	Coded slice of a non-RAP, non-TFD and non-TLA picture slice_layer_rbsp( )	1
10	TFD slice	Coded slice of a TFD picture slice_layer_rbsp( )	1
11	TLA slice	Coded slice of a non-TFD TLA picture slice_layer_rbsp( )	1
12..24		Reserved	-
25	Parameter Set	Video parameter set video_parameter_set_rbsp( )	1
26	Parameter Set	Sequence parameter set seq_parameter_set_rbsp( )	1
27	Parameter Set	Picture parameter set pic_parameter_set_rbsp( )	1
28	Parameter Set	Adaptation parameter set aps_rbsp( )	1
29	Information	Access unit delimiter access_unit_delimiter_rbsp( )	0
30	Information	Filler data filler_data_rbsp( )	0
31	Information	Supplemental enhancement information (SEI) sei_rbsp( )	0
32..47		Reserved	-
48..63		Unspecified	-

Table 4 is a table showing values of nal_ref_flag estimated from the values of nal_unit_type. The NAL unit types can be grouped into a plurality of categories, as shown in the second column of Table 4. The categories are as described below.

1) RAP slice: NAL unit including a coded slice of a random access picture.

2) TLA slice: NAL unit including a coded slice of temporal layer access.

3) TFD slice: NAL unit including a coded slice of a picture tagged for discard.

4) Other slice: NAL unit including a coded slice except for the above slices.

5) Parameter set: NAL unit including a video, sequence, picture, or adaptation parameter set.

6) Information: NAL unit including an access delimiter, filler data, or supplemental enhancement information (SEI).

In the present embodiment, three new kinds of NAL unit types corresponding to 9, 10, and 11 as values of nal_unit_type (picture types) are added to nal_unit_type in the conventional technology. The NAL units with these values of nal_unit_type include the same slice types as the NAL units with the respective values of nal_unit_type of 1, 2, and 3. nal_unit_type: 1 includes a coded slice of a non-RAP, non-TFD, and non-TLA picture, nal_unit_type: 2 includes a coded slice of a TFD picture, and nal_unit_type: 3 includes a coded slice of a non-TFT picture and a TLA picture.

The present embodiment is different from the conventional technology in that the values 1, 2, and 3 are the coded slices belonging to non-reference pictures and the values 9, 10, and 11 are the coded slices belonging to reference pictures.

The values assigned to the respective categories are not limited to those described above. Furthermore, each category may be extended to some sub-categories and these sub-categories may be assigned new values, using the reserved values in Table 4.

Fig. 3 shows the operation of the video predictive encoding device for encoding of the NAL unit header in the present embodiment. In step 118, the video predictive encoding device derives video data to be packetized. In step 120, the device encodes the first bit of the NAL unit always fixed to 0. In step 130, the device determines nal_unit_type and encodes it. In step 140 the device encodes temporal_id and in step 150 the device encodes reserved five bits (reserved_one_5 bits), completing the NAL unit header. In step 160, the device packetizes the remaining payload (payload) and terminates the processing.

Fig. 4 shows the details of the process in the determination and encoding of nal_unit_type in step 130 above.

In step 210, the video predictive encoding device determines whether the data to be packetized is a coded slice belonging to any one of random access pictures (RAPs); when the data is a coded slice belonging to any one of RAPs (YES), the device goes to step 220. If not (NO) the device goes to step 230.

In step 220, the video predictive encoding device encodes nal_unit_type by a number from 4 to 8 to infer that nal_ref_flag is 1, according to the RAP type, and then moves to step 140.

In step 230, the video predictive encoding device determines whether the data to be packetized is a parameter set, and when the data is determined to be a parameter set (YES), the device moves to step 240. If the data is not a parameter set (NO), the device moves to step 250.

In step 240, the video predictive encoding device encodes nal_unit_type by a number from 25 to 28 to infer that nal_ref_flag is 1, according to the parameter set, and then the device moves to step 140.

In step 250, the video predictive encoding device determines whether the data to be packetized is information data, and when the data is information data (YES), the device moves to step 260. If not (NO) the device moves to step 270.

In step 260, the video predictive encoding device encodes nal_unit_type by a number from 29 to 31 to infer that nal_ref_flag is 0, according to the information type, and then moves to step 140.

In step 270, the video predictive encoding device determines whether the data to be packetized is a reference picture, and when the data is a reference picture (YES), the device moves to step 280. If the data is not a reference picture (NO), the device moves to step 290. The determination of whether or not the data is a reference picture is made based on the reference information between pictures output from the predicted signal generation unit.

The conditional branching in step 270 may be arranged as follows. In step 270 the video data must be determined as either a reference picture or a non-reference picture. In step 270 the video predictive encoding device determines whether the picture is a reference picture, and when the picture is a reference picture (YES), the device moves to step 280. If the picture is not a reference picture (NO), the device moves to step 290.

In step 280, the video predictive encoding device encodes nal_unit_type by a number from 9 to 11 to infer that nal_ref_flag is 1, according to the slice type, and then moves to step 140.

In step 290, the video predictive encoding device encodes nal_unit_type by a number from 1 to 3 to infer that nal_ref_flag is 0, according to the slice type, and then the device moves to step 140.

Fig 5 shows operation of the video predictive decoding device for decoding of the NAL unit header in the present embodiment. In step 310, the video predictive decoding device derives a next packet for decoding. In step 320, the device decodes the first bit (forbidden_zero_bit) of the NAL unit always fixed to 0. In step 330, the device decodes nal_unit_type and sets the value of nal_ref_flag. In step 340 the device decodes temporal_id and in step 350 the device decodes the reserved five bits (reserved_one_5 bits) to complete the NAL unit header. In step 360 the device reads out the remaining payload from the packet and then terminates the processing.

Fig. 6 shows the details of the process in the decoding of nal_unit_type and the setting of the value of nal_ref_flag in step 330 above.

In step 400, the video predictive decoding device decodes the NAL unit header to derive the value of nal_unit_type.

In step 410, the video predictive decoding device determines whether the value of nal_unit_type is a number from 1 to 3, and when the value is any one of 1 to 3 (YES), the NAL unit includes one of the coded slices of non-reference pictures and therefore the device moves to step 420. If value of nal_unit_type is not a number from 1 to 3 (NO), the device moves to step 430.

In step 420, the video predictive decoding device sets the value of nal_ref_flag to 0 and then moves to step 340.

In step 430, the video predictive decoding device determines whether the value of nal_unit_type is a number from 4 to 11, and when the value is any one of 4 to 11 (YES), the NAL unit includes one of the coded slices of random access pictures or coded slices of reference pictures, and therefore the device moves to step 440. If the value of nal_unit_type is not a number from 4 to 11 (NO) the device moves to step 450.

In step 440, the video predictive decoding device sets the value of nal_ref_flag to 1 and then moves to step 340.

In step 450, the video predictive decoding device determines whether the value of nal_unit_type is a number from 25 to 28, and when the value is any one of 25 to 28 (YES), the NAL unit includes a parameter set and then the device moves to step 460. If the value of nal_unit_type is not a number from 25 to 28 (NO), the device moves to step 470.

In step 460, the video predictive decoding device sets the value of nal_ref_flag to 1 and then moves to step 340.

In step 470, the video predictive decoding device determines whether the value of nal_unit_type is a number from 29 to 31, and when the value is any one of 29 to 31 (YES), the NAL unit includes information data and then the device moves to step 480. If the value of nal_unit_type is not a number from 29 to 31 (NO), nal_unit_type is an invalid value and the device moves to step 490.

In step 480, the video predictive decoding device sets the value of nal_ref_flag to 0 and then moves to step 340.

In step 490, the video predictive decoding device determines that the value of nal_ref_flag is undefined, and then the device moves to step 340.

In the present embodiment the aforementioned setting of nal_ref_flag is performed through the logical determination, but the value of nal_ref_flag may also be set using a reference table of nal_ref_flag against index of nal_unit_type in other embodiments. Table 5 is an example of the reference table of nal_ref_flag against index of nal_unit_type.

Table 5

NAL unit type range	Inferred value of nal_ref_flag
1 to 3	0
4 to 11	1
25 to 28	1
29 to 31	0

In Table 5, the thirty two entries of nal_ref_flag are set to the same values as in the last column of Table 4.

The aforementioned nal_ref_flag estimation or setting method is not limited to the video predictive decoding device but can also be applied to the MANEs.

In the present embodiment the video predictive decoding device may select not performing the setting of nal_ref_flag and may directly use the value of nal_unit_type in determining whether a decoded picture is a reference picture. This can be explained as follows by use of a logical expression. When nal_unit_type of the relevant picture is 1, 2, or 3, the relevant picture is a non-reference picture. Otherwise, the relevant picture is a reference picture and is stored for use as reference of another picture.

In the present embodiment the definition of reference picture and non-reference picture is applied to the entire video data. However, in embodiments where the video data is subjected to a selective frame drop process to discard pictures in a higher temporal layer, this definition may no longer be accurate.

Under such circumstances, some reference pictures can be pictures that are not used for reference. To avoid this situation, in some embodiments the reference pictures with nal_unit_type of 9, 10, and 11 and the non-reference pictures with nal_unit_type of 1, 2, and 3 may be defined as described below.

A reference picture is a picture to be used for inter prediction by any other picture in the same temporal layer as the foregoing picture.

A non-reference picture is a picture that is not to be used for inter prediction by any other picture in the same temporal layer as the foregoing picture.

In the conventional method described in Non Patent Literature 1, the inter prediction is instructed by a content of a reference picture set (RPS) to define which pictures can be used for inter prediction. For this reason, the foregoing definition may be described as follows.

A non-reference picture (with nal_unit_type of 1, 2, or 3) is not included in the RPS of any other picture in the same temporal layer as the foregoing picture.

A reference picture (with nal_unit_type of 9, 10, or 11) is included in the RPS of any other picture in the same temporal layer as the foregoing picture.

A video predictive encoding program and a video predictive decoding program for letting a computer function as the foregoing video predictive encoding device and video predictive decoding device can be provided as programs stored in a storage medium. Examples of such storage media include disks, CD-ROMs, DVDs, and ROMs, semiconductor memories, and so on.

Fig. 7 is a drawing showing a hardware configuration of a computer for executing a program stored in a storage medium and Fig. 8 a perspective view of a computer for executing a program stored in a storage medium. The computer can be embodied in a DVD player, a set-top box, a cell phone, etc., provided with a CPU, and be configured to perform processing and control by software.

As shown in Fig. 7, the computer 30 is provided with a reading device 12 such as a disk drive unit, a CD-ROM drive unit, or a DVD drive unit, a working memory (RAM) 14 on which an operating system is resident, a memory 16 for storing programs stored in the storage medium 10, a monitor unit 18 like a display, a mouse 20 and a keyboard 22 as input devices, a communication device 24 for transmission and reception of data or the like, and a CPU 26 for controlling execution of programs. When the storage medium 10 is put into the reading device 12, the computer 30 becomes accessible to the video predictive encoding or decoding program stored in the storage medium 10, through the reading device 12 and becomes able to operate as the video predictive encoding or decoding device, based on the video predictive encoding or decoding program.

As shown in Fig. 8, the video predictive encoding program or the video predictive decoding program may be provided in the form of computer data signal 40 superimposed on a carrier wave, through a network. In this case, the computer 30 can execute the video predictive encoding program or the video predictive decoding program after the video predictive encoding program or the video predictive decoding program is received by the communication device 24 and is stored into the memory 16.

Specifically, as shown in Fig. 9, the video predictive encoding program P100 is a video predictive encoding program provided with an input module P101 to implement input of a plurality of pictures forming a video sequence, and an encoding module P102 to encode the pictures by either the intra prediction or the inter prediction to generate compressed picture data, and to packetize the compressed picture data with packet header information, wherein the packet header information contains a picture type and wherein the encoding module P102 determines the picture type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture.

Similarly, as shown in Fig. 10, the video predictive decoding program P200 is a video predictive decoding program provided with an input module P201 to implement input of compressed picture data resulting from encoding of a plurality of pictures forming a video sequence by either the intra prediction or the inter prediction and packetization thereof along with packet header information, and a decoding module P202 to reconstruct the packet header information and the compressed picture data, wherein the packet header information contains a picture type to uniquely indicate whether reconstructed picture data is used for reference in decoding of another picture and wherein the decoding module P202 determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture.

The decoding module P202 may determine whether reconstructed picture data is used for reference in decoding of another picture, based on a correspondence table in which the picture type is previously stored in association with information indicative of whether reconstructed picture data is used for reference in decoding of another picture.

List of Reference Signs

101 input terminal; 102 block partition unit; 103 predicted signal generation unit; 104 frame memory; 105 subtraction unit; 106 transform unit; 107 quantization unit; 108 de-quantization unit; 109 inverse transform unit; 110 addition unit; 111 entropy encoding unit; 112 output terminal; 113 input terminal; 201 input terminal; 202 data analysis unit; 203 de-quantization unit; 204 inverse transform unit; 205 addition unit; 206 output terminal; 207 frame memory; 208 predicted signal generation unit.

DESCRIPTION

TITLE OF INVENTION

VIDEO PREDICTIVE ENCODING DEVICE, VIDEO PREDICTIVE ENCODING METHOD, VIDEO PREDICTIVE ENCODING PROGRAM, VIDEO PREDICTIVE DECODING DEVICE, VIDEO PREDICTIVE DECODING METHOD, AND VIDEO PREDICTIVE DECODING PROGRAM

Technical Field

The present invention relates to a video predictive encoding device, a video predictive encoding method, a video predictive encoding program, a video predictive decoding device, a video predictive decoding method, and a video predictive decoding program.

Background Art

In the conventional video compression technology, a bit stream is encapsulated in a network abstraction layer (NAL) unit. The NAL unit provides a self-contained packet and gives a video layer identity in different network environments. A header of the NAL unit contains information used in a system layer. The header of the NAL unit becomes a part of the packet header used in a packet network, and is designed to operate by media aware network elements (MANEs).

The NAL unit header in the conventional technology includes the following syntax elements: nal_ref_flag which indicates whether the NAL unit is used for reference in a decoding process of another NAL unit; nal_unit_type which indicates a type of a content transmitted by the NAL unit, where the NAL unit contains information such as a parameter set, a coded slice, or a supplemental enhancement information (SEI) message; and temporal_id which indicates a temporal identifier of the NAL unit.

The conventional technology is described in Non Patent Literature 1.

Citation List

Non Patent Literature

Non Patent Literature 1: Benjamin Bross et al., "High efficiency video coding (HEVC) text specification draft 7," Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 9th Meeting: Geneva, CH, 27th April - 7th May 2012

SUMMARY OF INVENTION

Technical Problem

As the MANEs are designed to check the minimum number of bytes at the head of a packet, the NAL unit header is a limited resource. In the conventional technology, the NAL unit header is only 2 bytes. For this reason, all syntax elements of the NAL unit header are important and should transmit as much information as possible and be unrelated to the other syntax elements.

In most of NAL unit types, nal_ref_flag needs to be set at a fixed value, and thus nal_ref_flag is not needed. In the specification described in Non Patent Literature 1, there are only three kinds of NAL unit types whose nal_ref_flag can take the value of 0 or 1. In the other NAL unit types defined in the specification, the value of nal_ref_flag is fixed. This is shown in Table 1.

Table 1

NAL unit type range	Possible nal_ref_flag	Fixed / Variable nal_ref_flag
1 to 3	0 or 1	Variable
4 to 8	1	Fixed
25 to 28	1	Fixed
29 to 31	0	Fixed

Table 1 is a table showing correspondence between the values of nal_unit_type (NAL unit type range column) and the possible values of the nal_ref_flag (Possible nal_ref_flag column). In this case, the NAL unit types of nal_unit_type can have values of 1, 2, or 3 and the value of nal_ref_flag can be 0 or 1. The remaining NAL unit types are reserved or not specified.

Although the value of nal_ref_flag is uniquely determined according to the value of nal_unit_type as described above, the conventional technique assigns respective bits to nal_ref_flag and nal_unit_type, which is an inefficient design.

Solution to Problem

A solution to the above problem is to infer the value of nal_ref_flag from the NAL unit type, without explicitly sending nal_ref_flag in the NAL unit header. Three NAL unit types in which it is inferred that nal_ref_flag is 1 are added to the three NAL unit types the content of which can be a reference picture or a non-reference picture. For the original three NAL unit types, it is inferred that nal_ref_flag is 0.

In order to solve the foregoing problem, a video predictive encoding device according to an aspect of the present invention comprises input means that inputs a plurality of pictures forming a video sequence; and encoding means which encodes the pictures by either intra prediction or inter prediction to generate compressed picture data, and which packetizes the compressed picture data along with packet header information, wherein the packet header information contains a picture type, and wherein the encoding means determines the picture type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture. An encoding means of a video predictive encoding device according to one aspect of the present invention determines the picture type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture in the same temporal layer.

A video predictive decoding device according to an aspect of the present invention comprises input means which inputs compressed picture data resulting from encoding of a plurality of pictures forming a video sequence by either intra prediction or inter prediction and packetization of the compressed picture data with packet header information; and decoding means which reconstructs the packet header information and the compressed picture data, wherein the packet header information contains a picture type uniquely indicating whether reconstructed picture data is used for reference in decoding of another picture, and wherein the decoding means determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture.

In the video predictive decoding device according to an aspect of the present invention the decoding means determines whether reconstructed picture data is used for reference in decoding of another picture, based on a correspondence table in which the picture type is previously stored in association with information indicative of whether reconstructed picture data is used for reference in decoding of another picture. A decoding means of a video predictive decoding device according to an aspect of the present invention determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture in the same temporal layer.

A video predictive encoding method according to an aspect of the present invention is a video predictive encoding method comprising: an input step of inputting a plurality of pictures forming a video sequence; and an encoding step of encoding the pictures by either intra prediction or inter prediction to generate compressed picture data, and packetizing the compressed picture data with packet header information, wherein the packet header information contains a picture type, and wherein the encoding step determines the picture type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture. An encoding step of a video predictive encoding method according to an aspect of the present invention determines the picture type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture in the same temporal layer.

A video predictive decoding method according to an aspect of the present invention is a video predictive decoding method comprising: an input step of inputting compressed picture data resulting from encoding of a plurality of pictures forming a video sequence by either intra prediction or inter prediction and packetization of the compressed picture data with packet header information; and a decoding step of reconstructing the packet header information and the compressed picture data as reconstructed picture data, wherein the packet header information contains a picture type uniquely indicating whether the reconstructed picture data is used for reference in decoding of another picture, and wherein the decoding step determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture.

In the video predictive decoding method according to an aspect of the present invention the decoding step determines whether reconstructed picture data is used for reference in decoding of another picture, based on a correspondence table in which the picture type is previously stored in association with information indicative of whether reconstructed picture data is used for reference in decoding of another picture. A decoding step of a video predictive decoding method according to an aspect of the present invention determines, based on the picture type, whether reconstructed picture data is used for reference in decoding of another picture in the same temporal layer.

A video predictive encoding program according to an aspect of the present invention is a video predictive encoding program comprising: an input module which inputs a plurality of pictures forming a video sequence; and an encoding module which encodes the pictures by either intra prediction or inter prediction to generate compressed picture data, and which packetizes the compressed picture data along with packet header information, wherein the packet header information contains a picture type, and wherein the encoding module determines the picture type so as to uniquely indicate whether encoded

'16 OCT 19 P2:28

CLAIMS

1. A video predictive encoding device comprising:

input means that inputs a plurality of pictures forming a video sequence; and

encoding means which encodes the pictures to generate compressed picture data, and encapsulates the compressed picture data in a NAL unit with NAL unit header information,

wherein the plurality of pictures forming the video sequence are separated into a plurality of temporal layers,

the NAL unit header information contains a nal_unit_type, and

wherein the encoding means determines the nal_unit_type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture in the same temporal layer.

2. A video predictive decoding device comprising:

input means that inputs compressed picture data for a plurality of pictures forming a video sequence, where the compressed picture data is encapsulated in a NAL unit with NAL unit header information; and

decoding means which decodes the NAL unit header information and reconstructs the compressed picture data as reconstructed picture data,

wherein the plurality of pictures forming the video sequence are separated into a plurality of temporal layers,

the NAL unit header information contains a nal_unit_type uniquely indicating whether reconstructed picture data is used for reference in decoding of another picture in the same temporal layer, and

wherein the decoding means reconstructs the compressed picture data based on the nal_unit_type.

3. A video predictive encoding method comprising:

an input step of inputting a plurality of pictures forming a video sequence; and

an encoding step of encoding the pictures to generate compressed picture data, and encapsulating the compressed picture data in a NAL unit with NAL unit header information,

wherein the plurality of pictures forming the video sequence are separated into a plurality of temporal layers,

the NAL unit header information contains a nal_unit_type, and

wherein the encoding step determines the nal_unit_type so as to uniquely indicate whether encoded picture data is used for reference in decoding of another picture in the same temporal layer.

4. A video predictive decoding method comprising:

an input step of inputting compressed picture data for a plurality of pictures forming a video sequence, where the compressed picture data is encapsulated in a NAL unit with NAL unit header information; and

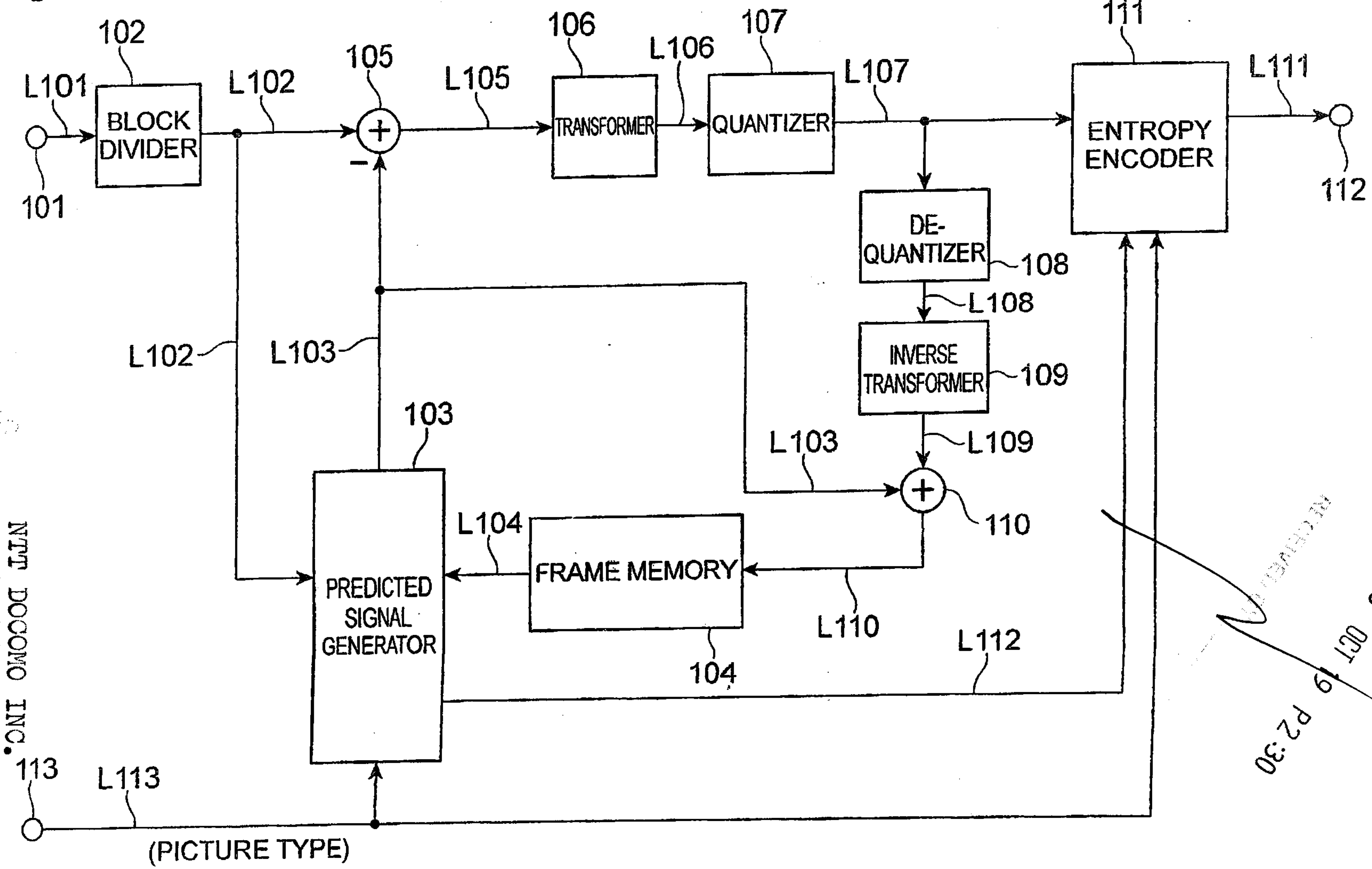
a decoding step of decoding the NAL unit header information and reconstructing the compressed picture data as reconstructed picture data,

wherein the plurality of pictures forming the video sequence are separated into a plurality of temporal layers,

the NAL unit header information contains a nal_unit_type uniquely indicating whether reconstructed picture data is used for reference in decoding of another picture in the same temporal layer, and

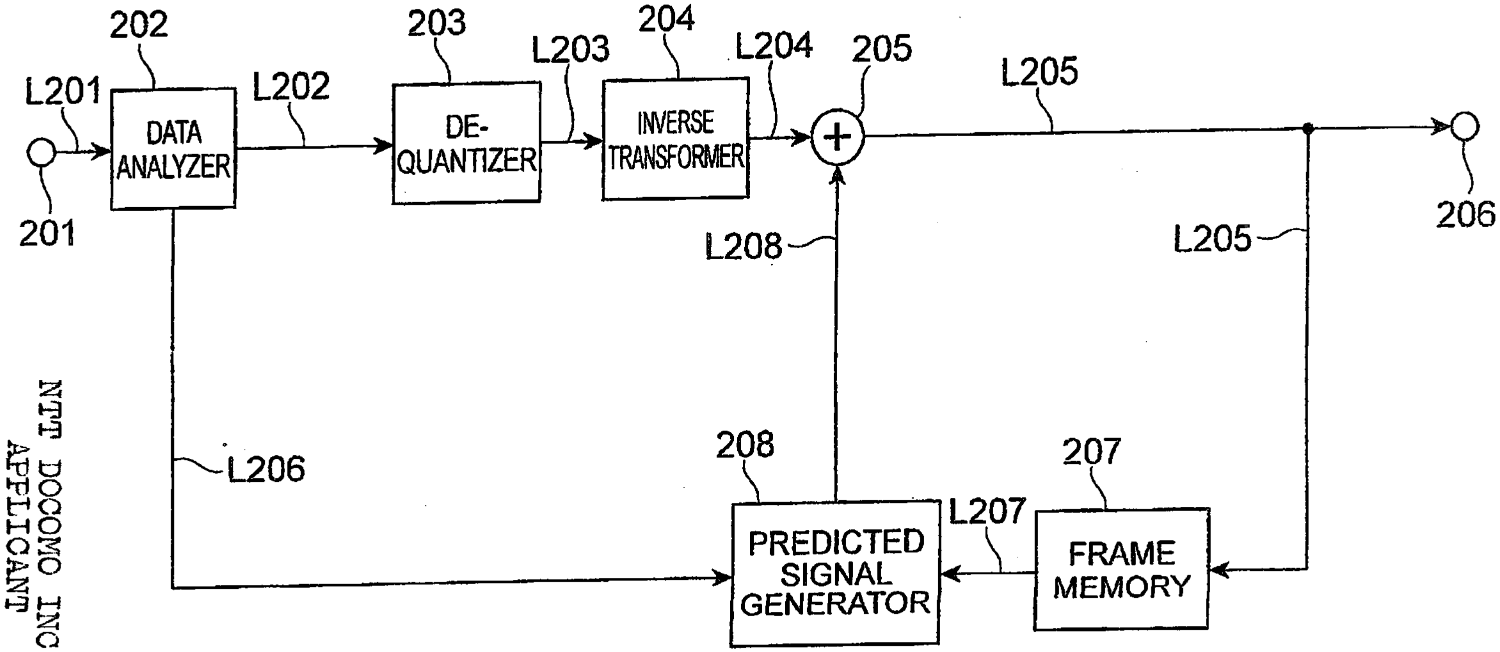
wherein the decoding step reconstructs the compressed picture data based on the nal_unit_type.

Fig.1



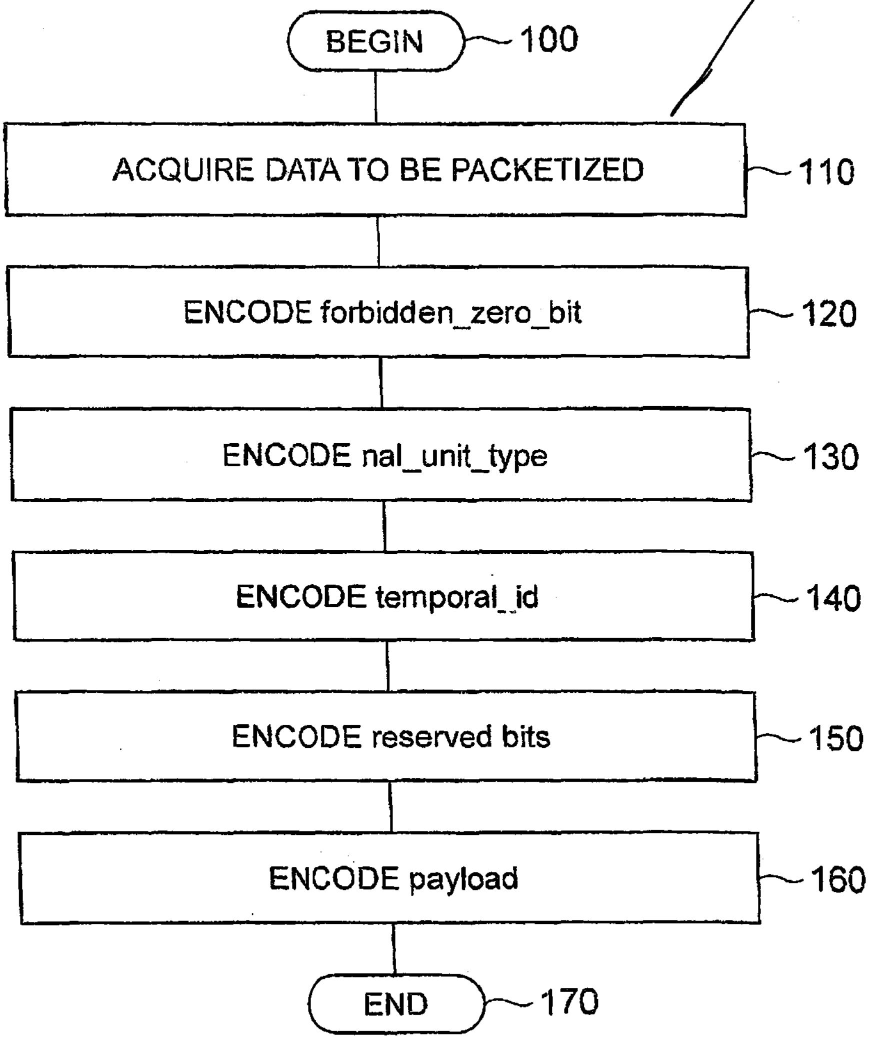
NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

Fig.2



NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

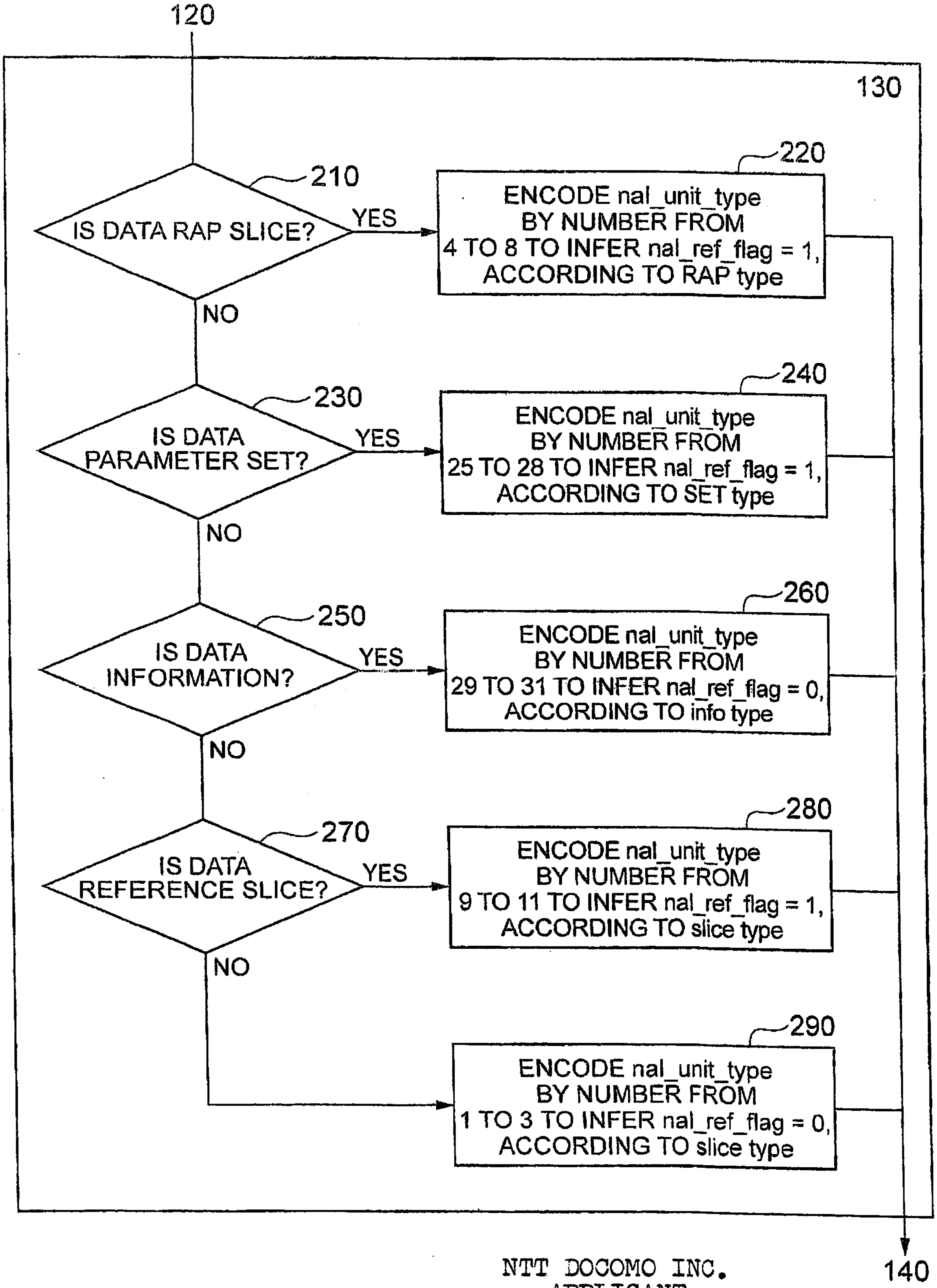
Fig.3



'16 OCT 19 P2:30

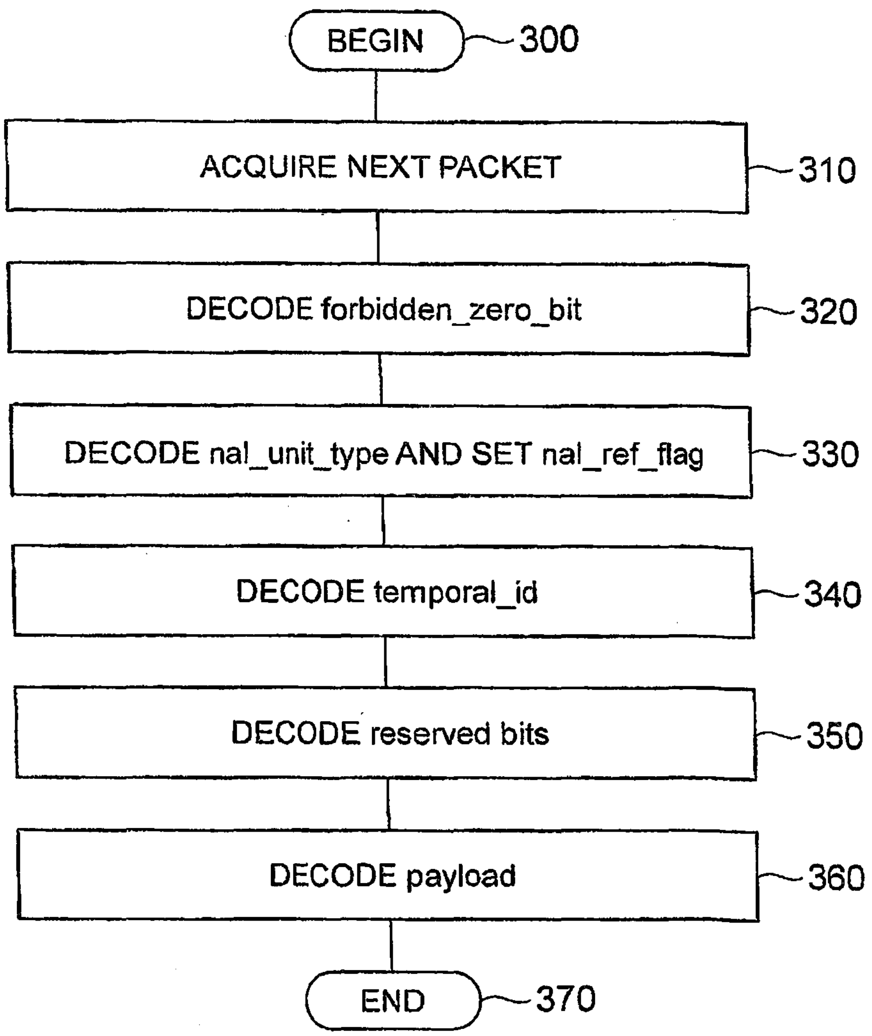
NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO & ONGSIAKO
BY:
CELSO L. CRUZ

Fig.4



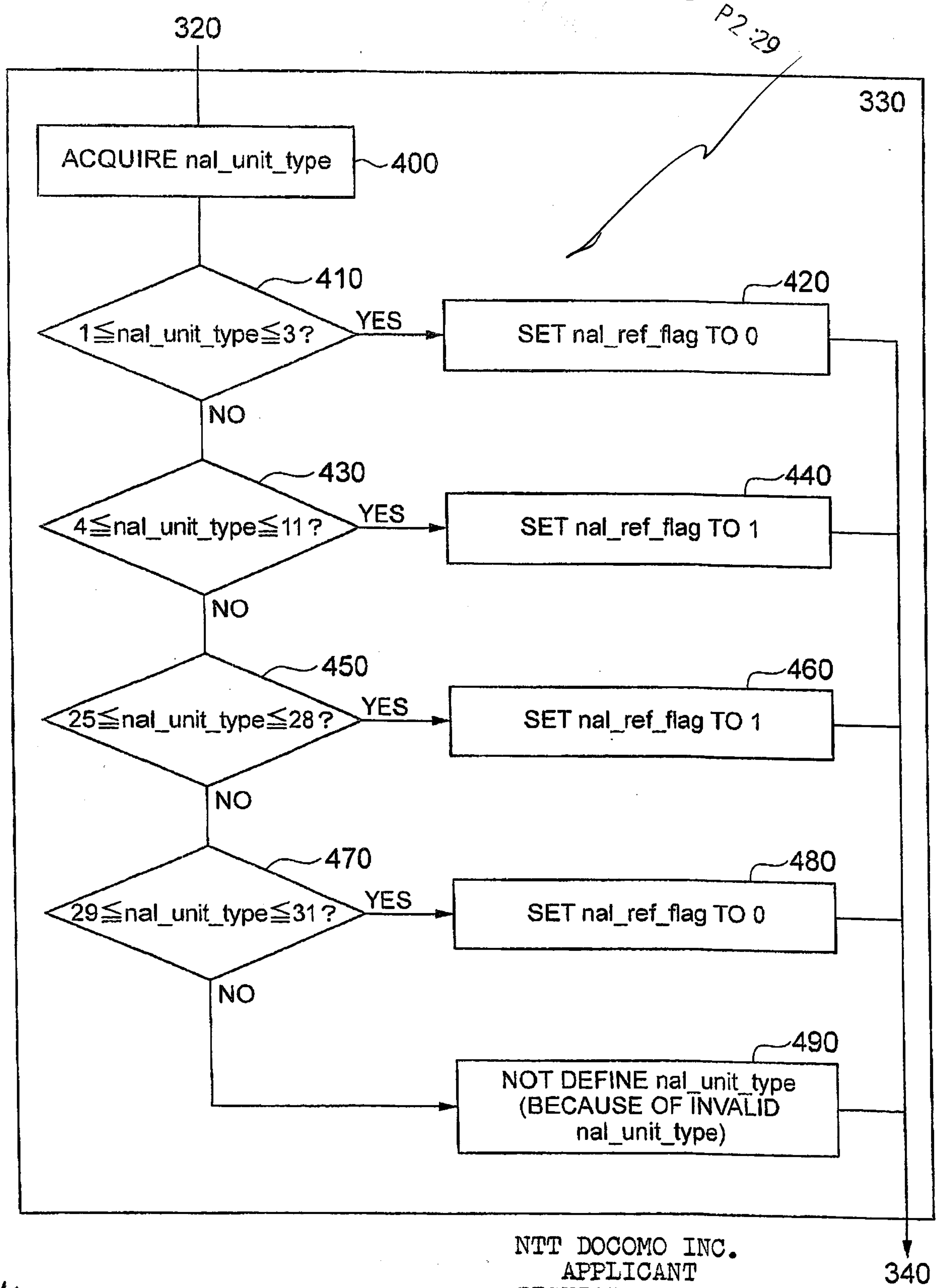
NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

Fig.5



NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

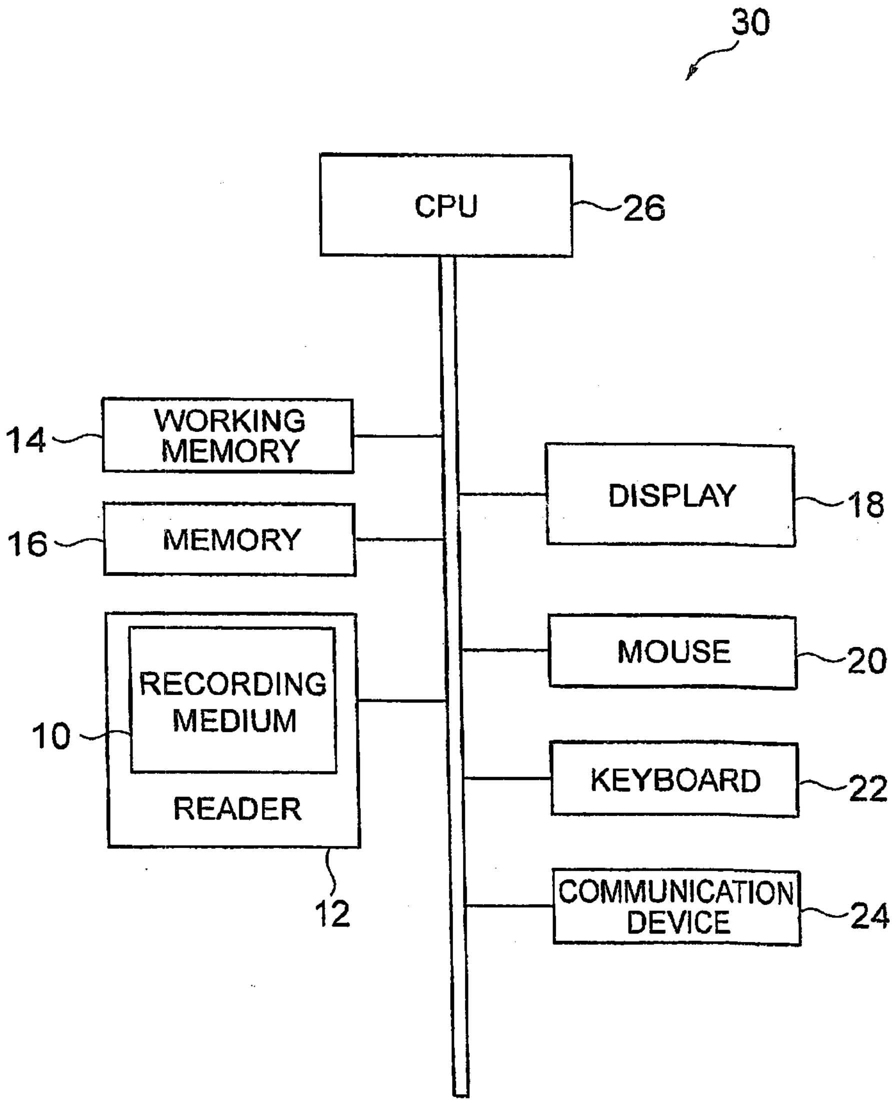
Fig.6



'16 OCT 19 P2:29

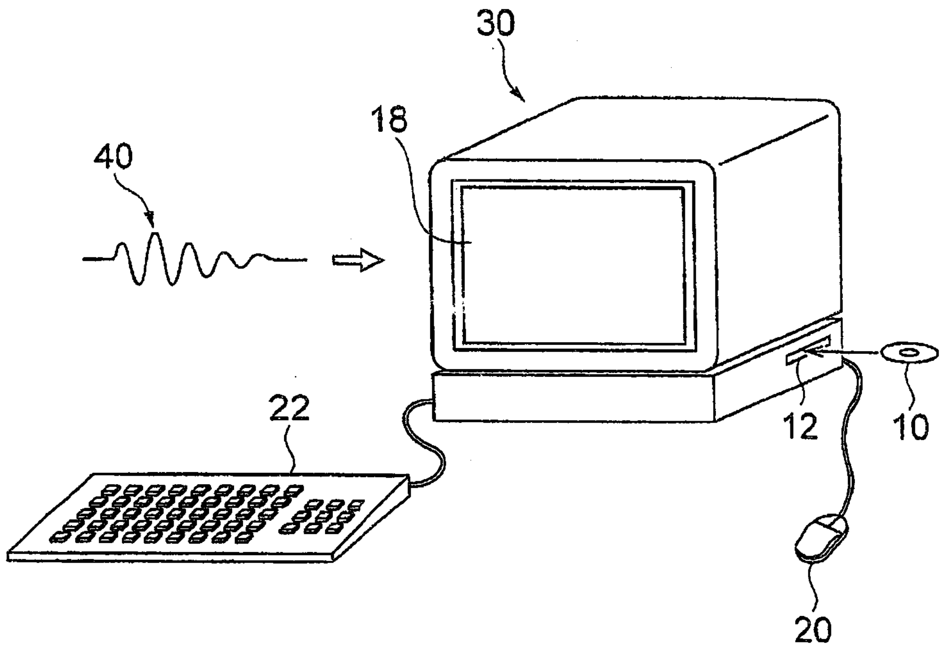
NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO & ONGSIAKO
BY:
CELSO L. CRUZ

Fig.7



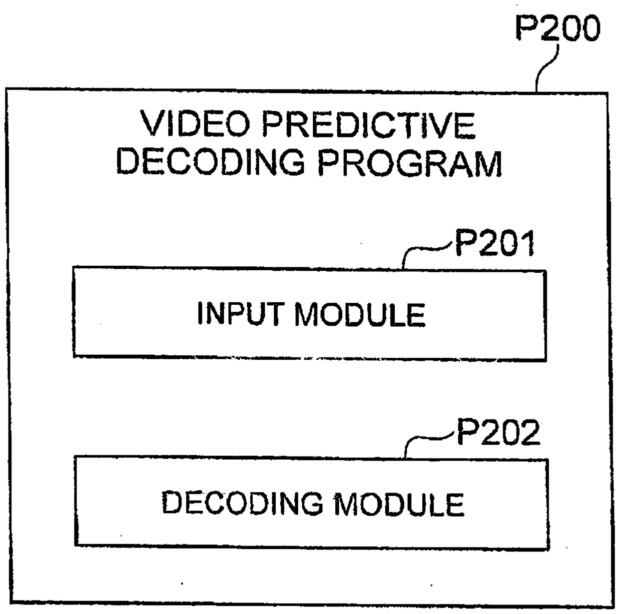
NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

Fig.8



NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

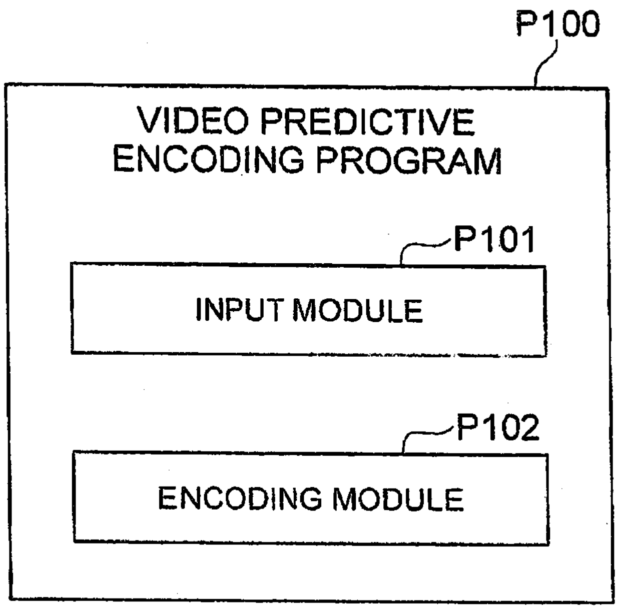
Fig.10



'16 OCT 19 P2:29
RECEIVED BY

NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ

Fig.9



NTT DOCOMO INC.
APPLICANT
SIGUION REYNA MONTECILLO
& ONGSIAKO
BY:
CELSO L. CRUZ