*159159*

*G00002*

D E S C R I P T I O N

# VIDEO ENCODING/DECODING METHOD AND APPARATUS

## Technical Field

The present invention relates to a video encoding/decoding method and apparatus which encode/decode a fade video and dissolving video, in particular, at high efficiency.

## Background Art

Motion compensation predictive inter-frame encoding is used as one of encoding modes in a video encoding standard scheme such as ITU-TH.261, H.263, ISO/IEC MPEG-2, or MPEG-4. As a predictive model in motion compensation predictive inter-frame encoding, a model that exhibits the highest predictive efficiency when no change in brightness occurs in the time direction is used. In the case of a fade video which changes in the brightness of pictures, there is no method known up to now which makes a proper prediction against a change in the brightness of pictures when, for example, a normal picture fades in from a black picture. In order to maintain picture quality in a fade video as well, therefore, a large number of bits are required.

In order to solve this problem, for example, in Japanese Patent No. 3166716, "Fade Countermeasure Video

Encoder and Encoding Method", a fade video part is detected to change the allocation of the number of bits. More specifically, in the case of a fadeout video, a large number of bits are allocated to the start part of fadeout that changes in luminance. In general, the last part of fadeout becomes a monochrome picture, and hence can be easily encoded. For this reason, the number of bits allocated to this part is reduced. This makes it possible to improve the overall picture quality without excessively increasing the total number of bits.

In Japanese Patent No. 2938412, "Video Luminance Change Compensation Method, Video Encoding Apparatus, Video Decoding Apparatus, Recording Medium on Which Video Encoding or Decoding Program Is Recorded, and Recording Medium on Which Encoded Data of Video Is Recorded", there is proposed an encoding scheme of properly coping with a fade video by compensating for a reference picture in accordance with two parameters, i.e., a luminance change amount and contrast change amount.

In Thomas Wiegand and Berand Girod, "Multi-frame motion-compensated prediction for video transmission", Kluwer Academic Publishers 2001, an encoding scheme based on a plurality of frame buffers is proposed. In this scheme, an attempt has been made to improve the predictive efficiency by selectively generating a

prediction picture from a plurality of reference frames held in the frame buffers.

According to the conventional techniques, in order to encode a fade video or dissolving video while maintaining high picture quality, a large number of bits are required. Therefore, an improvement in encoding efficiency cannot be expected.

## Disclosure of Invention

It is an object of the present invention to provide a video encoding/decoding method and apparatus which can encode a video which changes in luminance over time, e.g., a fade video or dissolving video, in particular, at high efficiency.

According to a first aspect of the present invention, there is provided a video encoding method of subjecting an input video signal to motion compensation predictive encoding by using a reference picture signal representing at least one reference picture and a motion vector between the input video signal and the reference picture signal, comprising a step of selecting one combination, for each block of the input video signal, from a plurality of combinations each including at least one reference picture number determined in advance for the reference picture and a predictive parameter, a step of generating a prediction picture signal in accordance with the reference picture number and predictive parameter of the selected

combination, a step of generating a predictive error signal representing an error between the input video signal and the prediction picture signal, and a step of encoding the predictive error signal, information of the motion vector, and index information indicating the selected combination.

According to a second aspect of the present invention, there is provided a video decoding method comprising a step of decoding encoded data including a predictive error signal representing an error in a prediction picture signal with respect to a video signal, motion vector information, and index information indicating a combination of at least one reference picture number and a predictive parameter, a step of generating a prediction picture signal in accordance with the reference picture number and predictive parameter of the combination indicated by the decoded index information, and a step of generating a reproduction video signal by using the predictive error signal and the prediction picture signal.

As described above, according to the present invention, there are prepared a plurality of different predictive schemes using combinations of reference picture numbers and predictive parameters or combinations of a plurality of predictive parameters corresponding to designated reference picture numbers. This makes it possible to generate a proper prediction

picture signal, on the basis of a predictive scheme with higher predictive efficiency, with respect to such a video signal that a proper prediction picture signal cannot be generated by a general predictive scheme for video encoding, e.g., a fade video or dissolving video.

In addition, the video signal is a signal including a picture signal obtained for each frame of a progressive signal, a picture signal obtained for each frame obtained by merging two fields of an interlaced signal, and a picture signal obtained for each field of an interlaced signal. When the video signal is a picture signal on a frame basis, the reference picture signal number indicates a reference picture signal on a frame basis. When the video signal is a picture signal on a field basis, the reference picture signal number indicates a reference picture signal on a field basis.

This makes it possible to generate a proper prediction picture signal, on the basis of a predictive scheme with higher predictive efficiency, with respect to such a video signal including both a frame structure and a field structure that a proper prediction picture signal cannot be generated by a general predictive scheme for video encoding, e.g., a fade video or dissolving video.

Furthermore, information of a reference picture number or predictive parameter itself is not send from the encoding side to the decoding side, but index

information indicting a combination of a reference picture number and a predictive parameter is sent, or a reference picture number is separately sent. In this case, the encoding efficiency can be improved by sending index information indicating a combination of predictive parameters.

## Brief Description of Drawings

FIG. 1 is a block diagram showing the arrangement of a video encoding apparatus according to the first embodiment of the present invention;

FIG. 2 is a block diagram showing the detailed arrangement of a frame memory/prediction picture generator in FIG. 1;

FIG. 3 is a view showing an example of a table of combinations of reference frame numbers and predictive parameters, which is used in the first embodiment;

FIG. 4 is a flow chart showing an example of a sequence for selecting a predictive scheme (a combination of a reference frame number and a predictive parameter) for each macroblock and determining an encoding mode in the first embodiment;

FIG. 5 is a block diagram showing the arrangement of a video decoding apparatus according to the first embodiment;

FIG. 6 is a block diagram showing the detailed arrangement of the frame memory/prediction picture generator in FIG. 5;

FIG. 7 is a view showing an example of a table of combinations of predictive parameters in a case wherein the number of reference frames is one and a reference frame number is sent as mode information according to the second embodiment of the present invention;

FIG. 8 is a view showing an example of a table of combinations of predictive parameters in a case wherein the number of reference frames is two and a reference frame number is sent as mode information according to the second embodiment;

FIG. 9 is a view showing an example of a table of combinations of reference picture numbers and predictive parameters in a case wherein the number of reference frame is one according to the third embodiment of the present invention;

FIG. 10 is a view showing an example of a table for only luminance signals according to the third embodiment;

FIG. 11 is a view showing an example of a syntax for each block when index information is to be encoded;

FIG. 12 is a view showing a specific example of an encoded bit stream when a prediction picture is to be generated by using one reference picture;

FIG. 13 is a view showing a specific example of an encoded bit stream when a prediction picture is to be generated by using two reference pictures;

FIG. 14 is a view showing an example of a table of

reference frame numbers, reference field numbers, and predictive parameters when information to be encoded is a top field according to the fourth embodiment of the present invention; and

FIG. 15 is a view showing an example of a table of reference frame numbers, reference field numbers, and predictive parameters when information to be encoded is a bottom field according to the fourth embodiment of the present invention.

## Best Mode for Carrying Out the Invention

The embodiments of the present invention will be described below with reference to the several views of the accompanying drawing.

[First Embodiment]

(About Encoding Side)

FIG. 1 shows the arrangement of a video encoding apparatus according to the first embodiment of the present invention. A video signal 100 is input to the video encoding apparatus, for example, on a frame basis. The video signal 100 is input to a subtracter 101. The subtracter 101 calculates the difference between the video signal 100 and a prediction picture signal 212 to generate a predictive error signal. A mode selection switch 102 selects either the predictive error signal or the video signal 100. An orthogonal transformer 103 subjects the selected signal to an orthogonal transformation, e.g., a discrete cosine

transform (DCT). The orthogonal transformer 103 generates orthogonal transformation coefficient information, e.g., DCT coefficient information. The orthogonal transformation coefficient information is quantized by a quantizer 104 and branched into two paths. One quantization orthogonal transformation coefficient information 210 branched into two paths is guided to a variable-length encoder 111.

The other quantization orthogonal transformation coefficient information 210 branched into the two paths is sequentially subjected to processing reverse to that in the quantizer 104 and orthogonal transformer 103 by a dequantizer or inverse quantizer 105 and inverse orthogonal transformer 106 to be reconstructed into a predictive error signal. Thereafter, an adder 107 adds the reconstructed predictive error signal to the prediction picture signal 212 input through a switch 109 to generate a local decoded video signal 211. The local decoded video signal 211 is input to a frame memory/prediction picture generator 108.

The frame memory/prediction picture generator 108 selects one of a plurality of combinations of prepared reference frame numbers and predictive parameters. The linear sum of the video signal (local decoded video signal 211) of the reference frame indicated by the reference frame number of the selected combination is calculated in accordance with the predictive parameter

of the selected combination, and the resultant signal is added to an offset based on the predictive parameter. With this operation, in this case, a reference picture signal is generated on a frame basis. Subsequently, the frame memory/prediction picture generator 108 motion-compensates for the reference picture signal by using a motion vector to generate the prediction picture signal 212.

In this process the frame memory/prediction picture generator 108 generates motion vector information 214 and index information 215 indicating a selected combination of a reference frame number and a predictive parameter, and sends information necessary for selection of an encoding mode to a mode selector 110. The motion vector information 214 and index information 215 are input to a variable-length encoder 111. The frame memory/prediction picture generator 108 will be described in detail later.

The mode selector 110 selects an encoding mode on a macroblock basis on the basis of predictive information P from the frame memory/prediction picture generator 108, i.e., selects either the intraframe encoding mode or the motion compensated predictive interframe encoding mode, and outputs switch control signals M and S.

In the intraframe encoding mode, the switches 102 and 112 are switched to the A side by the switch

control signals M and S, and the input video signal 100 is input to the orthogonal transformer 103. In the interframe encoding mode, the switches 102 and 112 are switched to the B side by the switch control signals M and S. As a consequence, the predictive error signal from the subtracter 101 is input to the orthogonal transformer 103, and the prediction picture signal 212 from the frame memory/prediction picture generator 108 is input to the adder 107. Mode information 213 is output from the mode selector 110 and input to the variable-length encoder 111.

The variable-length encoder 111 subjects the quantization orthogonal transformation coefficient information 210, mode information 213, motion vector information 214, and index information 215 to variable-length encoding. The variable-length codes generated by this operation are multiplexed by a multiplier 114. The resultant data is then smoothed by an output buffer 115. Encoded data 116 output from the output buffer 115 is sent out to a transmission system or storage system (not shown).

An encoding controller 113 controls an encoding unit 112. More specifically, the encoding controller 113 monitors the buffer amount of the output buffer 115, and controls encoding parameters such as the quantization step size of the quantizer 104 to make the buffer amount constant.

(About Frame Memory/Prediction Picture Generator 108)

FIG. 2 shows the detailed arrangement of the frame memory/prediction picture generator 108 in FIG. 1. Referring to FIG. 2, the local decoded video signal 211 input from the adder 107 in FIG. 1 is stored in a frame memory set 202 under the control of a memory controller 201. The frame memory set 202 has a plurality of (N) frame memories FM1 to FMN for temporarily holding the local decoded video signal 211 as a reference frame.

In a predictive parameter controller 203 is prepared a plurality of combinations of reference frame numbers and predictive parameters in advance as a table. The predictive parameter controller 203 selects, on the basis of the video signal 100, a combination of the reference frame number of a reference frame and a predictive parameter that is used to generate the prediction picture signal 212, and outputs the index information 215 indicating the selected combination.

A multi-frame motion evaluator 204 generates a reference picture signal in accordance with the combination of the reference frame number and the index information selected by the predictive parameter controller 203. The multi-frame motion evaluator 204 evaluates the motion amount and predictive error from this reference picture signal and input video signal 100, and outputs the motion vector information 214 that

minimizes the predictive error. A multi-frame motion compensator 205 carries out motion-compensation for each block using a reference picture signal selected by the multi-frame motion evaluator 204 in accordance with the motion vector to generate the prediction picture signal 212.

The memory controller 201 sets a reference frame number to a local decoded video signal for each frame, and stores each frame in one of the frame memories FM1 to FMN of the frame memory set 202. For example, the respective frames are sequentially numbered from the frame nearest to the input picture. The same reference frame number may be set for different frames. In this case, for example, different predictive parameters are used. A frame near to the input picture is selected from the frame memories FM1 to FMN and sent to the predictive parameter controller 203.

(About Table of Combinations of Reference Frame Numbers and Prediction Parameters)

FIG. 3 shows an example of the table of combinations of reference frame numbers and predictive parameters, which is prepared in the predictive parameter controller 203. "Index" corresponds to prediction pictures that can be selected for each block. In this case, there are eight types of prediction pictures. A reference frame number n is the number of a local decoded video used as a reference

frame, and in this case, indicates the number of a local decoded video corresponding to n past frames.

When the prediction picture signal 212 is generated by using the picture signals of a plurality of reference frames stored in the frame memory set 202, a plurality of reference frame numbers are designated, and (the number of reference frames + 1) coefficients are designated as predictive parameters for each of a luminance signal (Y) and color difference signals (Cb and Cr). In this case, as indicated by equations (1) to (3), n assumes the number of reference frames, n + 1 predictive parameters Di (i =,..., n + 1) are prepared for the luminance signal Y; n + 1 predictive parameters Ei (i =,..., n + 1), for the color difference signal Cb; and n + 1 predictive parameters Fi (i =,..., n + 1), for the color difference signal Cr:

$$Y_t = \sum_{i=1}^{n} D_i Y_{t-i} + D_{n+1} \qquad (1)$$

$$Cb_t = \sum_{i=1}^{n} E_i Cb_{t-i} + E_{n+1} \qquad (2)$$

$$Cr_t = \sum_{i=1}^{n} F_i Cr_{t-i} + F_{n+1} \qquad (3)$$

This operation will be described in more detail with reference to FIG. 3. Referring to FIG. 3, the last numeral of each predictive parameter represents an offset, and the first numeral of each predictive parameter represents a weighting factor (predictive

coefficient). For index 0, the number of reference frames is given by n = 2, the reference frame number is 1, and predictive parameters are 1 and 0 for each of the luminance signal Y and color difference signals Cr and Cb. What the predictive parameters are 1 and 0 as in this case indicates that a local decoded video signal corresponding to the reference frame number "1" is multiplied by 1 and added to offset 0. In other words, the local decoded video signal corresponding to the reference frame number 1 becomes a reference picture signal without any change.

For index 1, two reference frames as local decoded video signals corresponding to the reference frame numbers 1 and 2 are used. In accordance with predictive parameters 2, -1, and 0 for the luminance signal Y, the local decoded video signal corresponding to the reference frame number 1 is doubled, and the local decoded video signal corresponding to the reference frame number 2 is subtracted from the resultant signal. Offset 0 is then added to the resultant signal. That is, extrapolation prediction is performed from the local decoded video signals of two frames to generate a reference picture signal. For the color difference signals Cr and Cb, since predictive parameters are 1, 0, and 0, the local decoded video signal corresponding to the reference frame number 1 is used as a reference picture signal without any change.

This predictive scheme corresponding to index 1 is especially effective for a dissolving video.

For index 2, in accordance with predictive parameters 5/4 and 16, the local decoded video signal corresponding to the reference frame number 1 is multiplied by 5/4 and added with offset 16. For the color difference signals Cr and Cb, since the predictive parameter is 1, the color difference signals Cr and Cb become reference picture signals without any change. This predictive scheme is especially effective for a fade-in video from a black frame.

In this manner, reference picture signals can be selected on the basis of a plurality of predictive schemes with different combinations of the numbers of reference frames to be used and predictive parameters. This makes it possible for this embodiment to properly cope with a fade video and dissolving video that have suffered deterioration in picture quality due to the absence of a proper predictive scheme.

(About Sequence for Selecting Prediction Scheme and Determining Encoding Mode)

An example of a specific sequence for selecting a predictive scheme (a combination of a reference frame numbers and a predictive parameter) for each macroblock and determining an encoding mode in this embodiment will be described next with reference to FIG. 4.

First of all, a maximum assumable value is set to

variable min_D (step S101). LOOP1 (step S102) indicates a repetition for the selection of a predictive scheme in interframe encoding, and variable i represents the value of "index" in FIG. 3. In this case, in order to obtain an optimal motion vector for each predictive scheme, an evaluation value D of each index (each combination of a reference frame number and a predictive parameter) is calculated from the number of bits associated with motion vector information 214 (the number of bits of a variable-length code output from the variable-length encoder 111 in correspondence with the motion vector information 214) and a predictive error absolute value sum, and a motion vector that minimizes the evaluation value D is selected (step S103). The evaluation value D is compared with min_D (step S104). If the evaluation value D is smaller than min_D, the evaluation value D is set to min_D, and index i is assigned to min_i (step S105).

An evaluation value D for intraframe encoding is then calculated (step S106). The evaluation value D is compared with min_D (step S107). If this comparison indicates that min_D is smaller than the evaluation value D, mode MODE is determined as interframe encoding, and min_i is assigned to index information INDEX (step S108). If the evaluation value D is smaller, mode MODE is determined as intraframe encoding

(step S109). In this case, the evaluation value D is set as the estimated value of the number of bits with the same quantization step size.

(About Decoding Side)

A video decoding apparatus corresponding to the video encoding apparatus shown in FIG. 1 will be described next. FIG. 5 shows the arrangement of the video decoding apparatus according to this embodiment. Encoded data 300 sent out from the video encoding apparatus show in FIG. 1 and sent through a transmission system or storage system is temporarily stored in an input buffer 301 and demultiplexed by a demultiplexer 302 for each frame on the basis of a syntax. The resultant data is input to a variable-length decoder 303. The variable-length decoder 303 decodes the variable-length code of each syntax of the encoded data 300 to reproduce a quantization orthogonal transformation coefficient, mode information 413, motion vector information 414, and index information 415.

Of the reproduced information, the quantization orthogonal transformation coefficient is dequantized by a dequantizer 304 and inversely orthogonal-transformed by an inverse orthogonal transformer 305. If the mode information 413 indicates the intraframe encoding mode, a reproduction video signal is output from the inverse orthogonal transformer 305. This signal is then output

as a reproduction video signal 310 thorough an adder 306.

If the mode information 413 indicates the interframe encoding mode, a predictive error signal is output from the inverse orthogonal transformer 305, and a mode selection switch 309 is turned on. The prediction picture signal 412 output from a frame memory/prediction picture generator 308 is added to the predictive error signal by the adder 306. As a consequence, the reproduction video signal 310 is output. The reproduction video signal 310 is stored as a reference picture signal in the frame memory/ prediction picture generator 308.

The mode information 413, motion vector information 414, and index information 415 are input to the frame memory/prediction picture generator 308. The mode information 413 is also input to the mode selection switch 309. In the intraframe encoding mode, the mode selection switch 309 is turned off. In the interframe encoding mode, the switch is turned on.

Like the frame memory/prediction picture generator 108 on the encoding side in FIG. 1, the frame memory/prediction picture generator 308 includes a plurality of prepared combinations of reference frame numbers and predictive parameters as a table, and selects one combination indicated by the index information 415 from the table. The linear sum of the

video signal (reproduction video signal 210) of the reference frame indicated by the reference frame number of the selected combination is calculated in accordance with the predictive parameter of the selected combination, and an offset based on the predictive parameter is added to the resultant signal. With this operation, a reference picture signal is generated. Subsequently, the generated reference picture signal is motion-compensated for by using the motion vector indicated by the motion vector information 414, thereby generating a prediction picture signal 412.

(About Frame Memory/Prediction Picture Generator 308)

FIG. 6 shows the detailed arrangement of the frame memory/prediction picture generator 308 in FIG. 5. Referring to FIG. 6, the reproduction video signal 310 output from the adder 306 in FIG. 5 is stored in the frame memory set 402 under the control of a memory controller 401. The frame memory set 402 has a plurality of (N) frame memories FM1 to FMN for temporarily holding the reproduction video signal 310 as a reference frame.

A predictive parameter controller 403 has in advance combinations of reference frame numbers and predictive parameters as a table like the one shown in FIG. 3. The predictive parameter controller 403 selects a combination of the reference frame number of a reference frame and a predictive parameter, which are

used to generate the prediction picture signal 412, on the basis of the index information 415 from the variable-length decoder 303 in FIG. 5. A plurality of multi-frame motion compensators 404 generate a reference picture signal in accordance with a combination of a reference frame number and index information, which is selected by the predictive parameter controller 403, and performs motion-compensation for each block using this reference picture signal in accordance with the motion vector indicated by the motion vector information 414 from the variable-length decoder 303 in FIG. 5, thereby generating the prediction picture signal 412.

[Second Embodiment]

The second embodiment of the present invention will be described next with reference to FIGS. 7 and 8. Since the overall arrangements of a video encoding apparatus and video decoding apparatus in this embodiment are almost the same as those in the first embodiment, only the differences from the first embodiment will be described.

In this embodiment, there is described an example of the manner of expressing predictive parameters based on a scheme of capable of designating a plurality of reference frame numbers in accordance with mode information of a macroblock basis. A reference frame number is discriminated by the mode information for

each macroblock. This embodiment therefore uses a table of predictive parameters as shown in FIGS. 7 and 8 instead of using a table of combinations of reference frame numbers and predictive parameters as in the first embodiment. That is, index information does not indicate a reference frame number, and only a combination of predictive parameters is designated.

The table in FIG. 7 shows an example of a combination of predictive parameters when the number of reference frames is one. As predictive parameters, (the number of reference frames + 1) parameters, i.e., two parameters (one weighting factor and one offset), are designated for each of a luminance signal (Y) and color difference signals (Cb and Cr).

The table in FIG. 8 shows an example of a combination of predictive parameters when the number of reference frames is two. In this case, as predictive parameters, (the number of reference frames + 1) parameters, i.e., three parameters (two weighting factors and one offset), are designated fro each of a luminance signal (Y) and color difference signals (Cb and Cr). This table is prepared for the encoding side and decoding side each as in the first embodiment.

[Third Embodiment]

The third embodiment of the present invention will be described with reference to FIGS. 9 and 10. Since the overall arrangements of a video encoding apparatus

and video decoding apparatus in this embodiment are almost the same as those in the first embodiment, only the differences from the first and second embodiments will be described below.

In the first and second embodiments, a video is managed on a frame basis. In this embodiment, however, a video is managed on a picture basis. If both a progressive signal and an interlaced signal exist as input picture signals, pictures are not necessarily encoded on a frame basis. In consideration of this, a picture assumes (a) a picture of one frame of a progressive signal, (b) a picture of one frame generated by merging two fields of an interlaced signal, or (c) a picture of one field of an interlaced signal.

If a picture to be encoded is a picture with a frame structure like (a) or (b), a reference picture used in motion compensation prediction is also managed as a frame regardless of whether the encoded picture, which is the reference picture, has a frame structure or field structure. A reference picture number is assigned to this picture. Likewise, if a picture to be encoded is a picture with a field structure like (c), a reference picture used in motion compensation prediction is also managed as a field regardless of whether the encoded picture, which is the reference picture, has a frame structure or field structure. A

reference picture number is assigned to this picture.

Equations (4), (5), and (6) are examples of predictive equations for reference picture numbers and predictive parameters, which are prepared in the predictive parameter controller 203. These examples are predictive equations for generating a prediction picture signal by motion compensation prediction using one reference picture signal.

$$Y = clip\left(\left(D_1(i) \times R_Y(i) + 2^{L_Y - 1}\right) >> L_Y + D_2(i)\right) \quad (4)$$

$$Cb = clip\left(\left(E_1(i) \times \left(R_{Cb}(i) - 128\right) + 2^{L_C - 1}\right) >> L_C + E_2(i) + 128\right) \quad (5)$$

$$Cr = clip\left(\left(F_1(i) \times \left(R_{Cr}(i) - 128\right) + 2^{L_C - 1}\right) >> L_C + F_2(i) + 128\right) \quad (6)$$

where Y is a prediction picture signal of a luminance signal, Cb and Cr are prediction picture signals of two color difference signals, $R_Y(i)$, $R_{Cb}(i)$, and $R_{Cr}(i)$ are the pixel values of the luminance signal and two color difference signals of a reference picture signal with index i, $D_1(i)$ and $D_2(i)$ are the predictive coefficient and offset of the luminance signal with index i, $E_i(i)$ and $E_2(i)$ are the predictive coefficient and offset of the color difference signal Cb with index i, and $F_1(i)$ and $F_2(i)$ are the predictive coefficient and offset of the color difference signal Cr with index i. Index i indicates a value from 0 (the maximum number of reference pictures - 1), and encoded for each block to be encoded (e.g., for each macroblock). The resultant

data is then transmitted to the video decoding apparatus.

The predictive parameters $D_1(i)$, $D_2(i)$, $E_1(i)$, $E_2(i)$, $F_1(i)$, and $F_2(i)$ are represented by values determined in advance between the video encoding apparatus and the video decoding apparatus or a unit of encoding such as a frame, field, or slice, and are encoded together with encoded data to be transmitted from the video encoding apparatus to the video decoding apparatus. With this operation, these parameters are shared by the two apparatuses.

The equations (4), (5), and (6) are predictive equations wherein powers of 2, i.e., 2, 4, 8, 16,... are selected as the denominators of predictive coefficients by which reference picture signals are multiplied. The predictive equations can eliminate the necessity of division and be calculated by arithmetic shifts. This makes it possible to avoid a large increase in calculation cost due to division.

In equations (4), (5), and (6), ">>" of a >> b represents an operator for arithmetically shifting an integer a to the right by b bits. The function "clip" represents a clipping function for setting the value in "()" to 0 when it is smaller than 0, and setting the value to 255 when it is larger than 255.

In this case, assuming that $L_Y$ is the shift amount of a luminance signal, and $L_C$ is the shift amount of a

color difference signal. As these shift amounts $L_Y$ and $L_C$, values determined in advance between the video encoding apparatus and the video decoding apparatus are used. The video encoding apparatus encodes the shift amounts $L_Y$ and $L_C$, together with a table and encoded data, in a predetermined unit of encoding, e.g., a frame, field, or slice, and transmits the resultant data to the video decoding apparatus. This allows the two apparatuses to share the shift amounts $L_Y$ and $L_C$.

In this embodiment, tables of combinations of reference picture numbers and predictive parameters like those shown in FIGS. 9 and 10 are prepared in the predictive parameter controller 203 in FIG. 2. Referring to FIGS. 9 and 10, index i corresponds to prediction pictures that can be selected for each block. In this case, four types of prediction pictures are present in correspondence with 0 to 3 of index i. "Reference picture number" is, in other words, the number of a local decoded video signal used as a reference picture.

"Flag" is a flag indicating whether or not a predictive equation using a predictive parameter is applied to a reference picture number indicated by index i. If Flag is "0", motion compensation prediction is performed by using the local decoded video signal corresponding to the reference picture number indicated by index i without using any

predictive parameter. If Flag is "1", a prediction picture is generated according to equations (4), (5), and (6) by using a local decoded video and predictive parameter corresponding to the reference picture number indicated by index i, thus performing motion compensation prediction. This information of Flag is also encoded, together with a table and encoded data, by using a value determined in advance between the video encoding apparatus and the video decoding apparatus or in a predetermined unit of encoding, e.g., a frame, field, or slice, in the video encoding apparatus. The resultant data is transmitted to the video decoding apparatus. This allows the two apparatuses to share the information of Flag.

In these cases, a prediction picture is generated by using a predictive parameter when index i = 0 with respect to a reference picture number 105, and motion compensation prediction is performed without using any predictive parameter when i = 1. As described above, a plurality of predictive schemes may exist for the same reference picture number.

The table shown in FIG. 9 has predictive parameters $D_1(i)$, $D_2(i)$, $E_1(i)$, $E_2(i)$, $F_1(i)$, and $F_2(i)$ assigned to a luminance signal and two color difference signals in correspondence with equations (4), (5), and (6). FIG. 10 shows an example of a table in which predictive parameters are assigned to only luminance

signals. In general, the number of bits of a color difference signal is not very large compared with the number of bits of a luminance signal. For this reason, in order to reduce the amount of calculation required to generate a prediction picture and the number of bits transmitted in a table, a table is prepared, in which predictive parameters for color difference signals are omitted as shown in FIG. 10 and predictive parameters are assigned to only luminance signals. In this case, only equation (4) is used as a predictive equation.

Equations (7) to (12) are predictive equations in a case wherein a plurality of (two in this case) reference pictures are used.

$$P_Y(i) = \left(D_1(i) \times R_Y(i) + 2^{L_Y - 1}\right) >> L_Y + D_2(i) \quad (7)$$

$$P_{Cb}(i) = \left(E_1(i) \times \left(R_{Cb}(i) - 128\right) + 2^{L_C - 1}\right) >> L_C + E_2(i) + 128 \quad (8)$$

$$P_{Cr}(i) = \left(F_1(i) \times \left(R_{Cr}(i) - 128\right) + 2^{L_C - 1}\right) >> L_C + F_2(i) + 128 \quad (9)$$

$$Y = clip\left(\left(P_Y(i) + P_Y(j) + 1\right) >> 1\right) \quad (10)$$

$$Cb = clip\left(\left(P_{Cb}(i) + P_{Cb}(j) + 1\right) >> 1\right) \quad (11)$$

$$Cr = clip\left(\left(P_{Cr}(i) + P_{Cr}(j) + 1\right) >> 1\right) \quad (12)$$

The pieces of information of the predictive parameters $D_1(i)$, $D_2(i)$, $E_1(i)$, $E_2(i)$, $F_1(i)$, $F_2(i)$, $L_Y$, and $L_C$ and Flag are values determined in advance between the video encoding apparatus and the video decoding apparatus or encoded, together with encoded

data, in a unit of encoding such as a frame, field, or slice, and are transmitted from the video encoding apparatus to the video decoding apparatus. This allows the two apparatuses to share these pieces of information.

If a picture to be decoded is a picture having a frame structure, a reference picture used for motion compensation prediction is also managed as a frame regardless of whether a decoded picture as a reference picture has a frame structure or field structure. A reference picture number is assigned to this picture. Likewise, if a picture to be programmed is a picture having a field structure, a reference picture used for motion compensation prediction is also managed as a field regardless of whether a decoded picture as a reference picture has a frame structure or field structure. A reference picture number is assigned to this picture.

(About Syntax of Index Information)

FIG. 11 shows an example of a syntax in a case wherein index information is encoded in each block. First of all, mode information MODE is present for each block. It is determined in accordance with the mode information MODE whether or not index information IDi indicating the value of index i and index information IDj indicating the value of index j are encoded. Encoded information of motion vector information MVi

for the motion compensation prediction of index i and motion vector information MVj for the motion predictive compensation of index j is added as motion vector information for each block after encoded index information.

(About Data Structure of Encoded Bit Stream)

FIG. 12 shows a specific example of an encoded bit stream for each block when a prediction picture is generated by using one reference picture. The index information IDi is set after mode information MODE, and the motion vector information MVi is set thereafter. The motion vector information MVi is generally two-dimensional vector information. Depending on a motion compensation method in a block which is indicated by mode information, a plurality of two-dimensional vectors may further be sent.

FIG. 13 shows a specific example of an encoded bit stream for each block when a prediction picture is generated by using two reference pictures. Index information IDi and index information IDj are set after mode information MODE, and motion vector information MVi and motion vector information MVj are set thereafter. The motion vector information MVi and motion vector information j are generally two-dimensional vector information. Depending on a motion compensation method in a block indicated by mode information, a plurality of two-dimensional vectors may

be further sent.

Note that the above structures of a syntax and bit stream can be equally applied to all the embodiments.

[Fourth Embodiment]

The fourth embodiment of the present invention will be described next with reference to FIGS. 14 and 15. Since the overall arrangements of a video encoding apparatus and video decoding apparatus in this embodiment are almost the same as those in the first embodiment, only differences from the first, second, and third embodiments will be described. In the third embodiment, encoding on a frame basis and encoding on a field basis are switched for each picture. In the fourth embodiment, encoding on a frame basis and encoding on a field basis are switched for each macroblock.

When encoding on a frame basis and encoding on a field basis are switched for each macroblock, the same reference picture number indicates different pictures, even within the same picture, depending on whether a macroblock is encoded on the frame basis or on the field basis. For this reason, with the tables shown in FIGS. 9 and 10 used in the third embodiment, a proper prediction picture signal may not be generated.

In order to solve this problem, in this embodiment, tables of combinations of reference picture numbers and predictive parameters like those shown in

FIGS. 14 and 15 are prepared in a predictive parameter controller 203 in FIG. 2. Assume that when a macroblock is to be encoded on the field basis, the same predictive parameter as that corresponding to a reference picture number (reference frame index number) used when the macroblock is encoded on the frame basis is used.

FIG. 14 shows a table used when the macroblock is encoded on a field basis and a picture to be encoded is a top field. The upper and lower rows of each field index column correspond to the top field and bottom field, respectively. As shown in FIG. 14, frame index j and field index k are related such that when k = 2j in the top field, k = 2j + 1 in the bottom field. Reference frame number m and reference field number n are related such that when n = 2m in the top field, n = 2m + 1 in the bottom field.

FIG. 15 shows a table used when the macroblock is encoded on a field basis, and a picture to be encoded is a bottom field. As in the table shown in FIG. 14, the upper and lower rows of each field index column correspond to a top field and the bottom field, respectively. In the table in FIG. 15, frame index j and field index k are related such that when k = 2 + 1 in the top field, k = 2j in the bottom field. This makes it possible to assign a small value as field index k to an in-phase bottom field. The relationship

between reference frame number m and reference field number n is the same as that in the table in FIG. 14.

When the macroblock is to be encoded on a field basis, a frame index and field index are encoded as index information by using the tables shown in FIGS. 14 and 15. When the macroblock is to be encoded on a frame basis, only the frame index common to the tables in FIGS. 14 and 15 is index-encoded as index information.

In this embodiment, predictive parameters are assigned to a frame and field by using one table. However, a table for frames and a table for fields may be separately prepared for one picture or slice.

Each embodiment described above has exemplified the video encoding/decoding scheme using orthogonal transformation on a block basis. Even if, however, another transformation technique such as wavelet transformation is used, the technique of the present invention which has been described in the above embodiments can be used.

Video encoding and decoding processing according to the present invention may be implemented as hardware (apparatus) or software using a computer. Some processing may be implemented by hardware, and the other processing may be performed by software. According to the present invention, there can be provided a program for causing a computer to execute

the above video encoding or video decoding or a storage medium storing the program.

## Industrial Applicability

As has been described above, the video encoding/decoding method and apparatus according to the present invention are suited to the image processing field in which a video which changes in luminance over time, such as a fade video or dissolving video, in particular, is encoded and decoded.

C L A I M S

1. A video encoding method of subjecting an input video signal to motion compensation predictive encoding by using a reference picture signal representing at least one reference picture and a motion vector between the input video signal and the reference picture signal, comprising:

a step of selecting one combination, for each block of the input video signal, from a plurality of combinations each including a predictive parameter and at least one reference picture number determined in advance for the reference picture;

a step of generating a prediction picture signal in accordance with the reference picture number and predictive parameter of the selected combination;

a step of generating a predictive error signal representing an error between the input video signal and the prediction picture signal; and

a step of encoding the predictive error signal, information of the motion vector, and index information indicating the selected combination.

2. A video encoding method according to claim 1, wherein the predictive parameter includes information of a weighting factor and offset, and the step of generating the prediction picture signal includes a process of calculating a linear sum of a reference picture signal, indicated by the reference picture

number included in the selected combination, in accordance with the weighting factor, and then adding the offset to the linear sum.

3. A video encoding method of subjecting an input video signal to motion compensation predictive encoding by using a reference picture and a motion vector between the input video signal and the reference picture, comprising:

a step of selecting one combination, for each block of the input video signal, from a plurality of combinations of predictive parameters prepared in advance;

a step of designating at least one reference picture number set to at least one reference picture;

a step of generating a prediction picture signal in accordance with a reference picture corresponding to the designated reference picture number and the predictive parameters of the selected combination;

a step of generating a predictive error signal representing an error between the input video signal and the prediction picture signal; and

a step of encoding the predictive error signal, information of the motion vector, the designated reference picture number, and index information indicating the selected combination.

4. A video encoding method according to claim 3, wherein the predictive parameter includes information

of a weighting factor and offset, and the step of generating the prediction picture signal includes a process of calculating a linear sum of a reference picture signal corresponding to the designated reference picture number in accordance with the weighting factor, and then adding the offset to the linear sum.

5. A video encoding method according to claim 2 or 4, wherein the weighting factor has a power of 2 as a denominator.

6. A video encoding method according to any one of claims 1 to 4, wherein the input video signal has a luminance signal and two color difference signals, and the predictive parameter is prepared for each of the luminance signal and the two color difference signals.

7. A video encoding method according to claim 1 or 3, wherein the input video signal is a picture signal input for each frame of a progressive signal or a picture signal input for each frame obtained by merging two fields of an interlaced signal, and the reference picture signal is a picture signal on a frame basis.

8. A video encoding method according to claim 1 or 3, wherein the input video signal is a picture signal input for each field of an interlaced signal, and the reference picture signal is a picture signal on a field basis.

9. A video encoding method according to claim 1 or 3, wherein the input video signal is a signal including a picture signal input for each frame of a progressive signal, a picture signal input for each frame obtained by merging two fields of an interlaced signal, and a picture signal input for each field of an interlaced signal, the reference picture signal is a picture signal on a frame basis when the input video signal is the picture signal input for each frame, and the reference picture signal is a picture signal on a field basis when the input video signal is the picture signal input for each field.

10. A video decoding method comprising:

a step of decoding encoded data including a predictive error signal representing an error in a prediction picture signal with respect to a video signal, motion vector information, and index information indicating a combination of at least one reference picture number and a predictive parameter;

a step of generating a prediction picture signal in accordance with the reference picture number and predictive parameter of the combination indicated by the decoded index information; and

a step of generating a reproduction video signal by using the predictive error signal and the prediction picture signal.

11. A video decoding method according to claim 10,

wherein the predictive parameter includes information of a weighting factor and offset, and the step of generating the prediction picture signal includes a process of calculating a linear sum of a reference picture signal indicated by the reference picture number included in the decoded index information in accordance with the weighting factor included in the index information, and then adding the offset included in the index information to the linear sum.

12. A video decoding method comprising:

a step of decoding encoded data including a predictive error signal representing an error in a prediction picture signal with respect to a video signal, motion vector information, and index information indicating a combination of a designated reference picture number and a predictive parameter;

a step of generating a prediction picture signal in accordance with the decoded reference picture number and the predictive parameter of the combination indicated by the decoded index information; and

a step of generating a reproduction video signal by using the predictive error signal and the prediction picture signal.

13. A video decoding method according to claim 12, wherein the predictive parameter includes information of a weighting factor and offset, and the step of generating the prediction picture signal includes a

process of calculating a linear sum of a reference picture signal, indicated by the decoded reference picture number, in accordance with the weighting factor included in the index information, and then adding the offset included in the index information to the linear sum.

14. A video decoding method according to claim 10 or 12, wherein the weighting factor has a power of 2 as a denominator.

15. A video decoding method according to claim 10 or 12, wherein the video signal is a picture signal obtained for each frame of a progressive signal or a picture signal obtained for each frame obtained by merging two fields of an interlaced signal, and the reference picture number indicates the number of a reference picture signal on a frame basis.

16. A video decoding method according to claim 10 or 12, wherein the video signal is a picture signal input for each field of an interlaced signal, and the reference picture signal number indicates the number of a reference picture signal on a field basis.

17. A video decoding method according to claim 10 or 12, wherein the video signal is a signal including a picture signal obtained for each frame of a progressive signal, a picture signal obtained for each frame obtained by merging two fields of an interlaced signal, and a picture signal obtained for each field of an

interlaced signal, the reference picture signal number indicates a reference picture signal on a frame basis when the video signal is the picture signal on a frame basis, and the reference picture signal number indicates a reference picture signal on a field basis when the video signal is the picture signal on a field basis.

18. A video encoding apparatus for subjecting an input video signal to motion compensation predictive encoding by using a reference picture and a motion vector between the input video signal and the reference picture, comprising:

means for selecting one combination, for each block of the input video signal, from a plurality of combinations each including a predictive parameter and at least one reference picture number determined in advance for the reference picture;

means for generating a prediction picture signal in accordance with the reference picture number and predictive parameter of the selected combination;

means for generating a predictive error signal representing an error between the input video signal and the prediction picture signal; and

means for encoding the predictive error signal, information of the motion vector, and index information indicating the selected combination.

19. A video encoding apparatus for subjecting an

input video signal to motion compensation predictive encoding by using a reference picture and a motion vector between the input video signal and the reference picture, comprising:

means for selecting one combination, for each block of the input video signal, from a plurality of combinations of predictive parameters prepared in advance;

means for designating at least one reference picture number set to at least one reference picture;

means for generating a prediction picture signal in accordance with a reference picture corresponding to the designated reference picture number and the predictive parameters of the selected combination;

means for generating a predictive error signal representing an error between the input video signal and the prediction picture signal; and

means for encoding the predictive error signal, information of the motion vector, the designated reference picture number, and index information indicating the selected combination.

20. A video decoding apparatus comprising:

means for decoding encoded data including a predictive error signal representing an error in a prediction picture signal with respect to a video signal, motion vector information, and index information indicating a combination of at least one

reference picture number and a predictive parameter;

means for generating a prediction picture signal in accordance with the reference picture number and predictive parameter of the combination indicated by the decoded index information; and

means for generating a reproduction video signal by using the predictive error signal and the prediction picture signal.

21. A video decoding apparatus comprising:

means for decoding encoded data including a predictive error signal representing an error in a prediction picture signal with respect to a video signal, motion vector information, and index information indicating a combination of a designated reference picture number and a predictive parameter;

means for generating a prediction picture signal in accordance with the decoded reference picture number and the predictive parameter of the combination indicated by the decoded index information; and

means for generating a reproduction video signal by using the predictive error signal and the prediction picture signal.