US 20250267292A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2025/0267292 A1

LEE et al. (43) Pub. Date: Aug. 21, 2025

(54) METHOD AND DEVICE FOR IMAGE ENCODING/DECODING, AND RECORDING MEDIUM HAVING BITSTREAM STORED THEREON

(71) Applicants: **Electronics and Telecommunications Research Institute**, Daejeon (KR); **CHIPS&MEDIA, INC**, Seoul (KR); **INDUSTRY-UNIVERSITY COOPERATION FOUNDATION KOREA AEROSPACE UNIVERSITY**, Goyang-si (KR); **INDUSTRY ACADEMY COOPERATION FOUNDATION OF SEJONG UNIVERSITY**, Seoul (KR)

(72) Inventors: **Jin Ho LEE**, Daejeon (KR); **Jung Won KANG**, Daejeon (KR); **Ha Hyun LEE**, Seoul (KR); **Sung Chang LIM**, Daejeon (KR); **Hui Yong KIM**, Daejeon (KR); **Dae Yeon KIM**, Seoul (KR); **Do Hyeon PARK**, Goyang-si (KR); **Jae Gon KIM**, Goyang-si (KR); **Yong Uk YOON**, Jeju-si (KR); **Yung Lyul LEE**, Seoul (KR)

(21) Appl. No.: **19/190,511**

(22) Filed: **Apr. 25, 2025**

Related U.S. Application Data

(63) Continuation of application No. 17/259,842, filed on Jan. 12, 2021, filed as application No. PCT/KR2019/008299 on Jul. 5, 2019, now Pat. No. 12,309,396.

(30) **Foreign Application Priority Data**

Jul. 13, 2018 (KR) 10-2018-0081836
Mar. 11, 2019 (KR) 10-2019-0027678

Publication Classification

(51) **Int. Cl.**

H04N 19/18	(2014.01)
H04N 19/119	(2014.01)
H04N 19/129	(2014.01)
H04N 19/176	(2014.01)
H04N 19/60	(2014.01)

(52) **U.S. Cl.**

CPC ***H04N 19/18*** (2014.11); ***H04N 19/119*** (2014.11); ***H04N 19/129*** (2014.11); ***H04N 19/176*** (2014.11); ***H04N 19/60*** (2014.11)

(57) **ABSTRACT**

Disclosed in a method of decoding an image. The method includes determining a zero out region within a current block; partitioning a region except for the zero out region within the current block, on a per transform coefficient group basis; and decoding a transform coefficient on the per transform coefficient group basis.

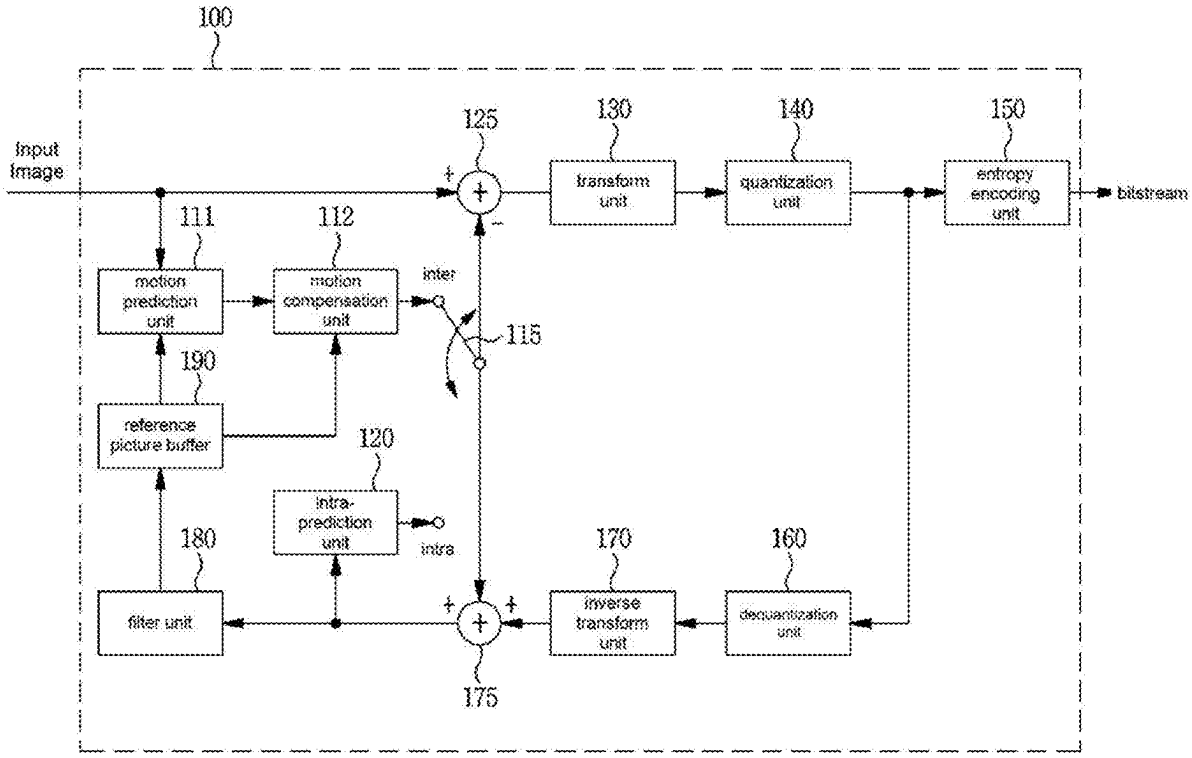


FIG. 1

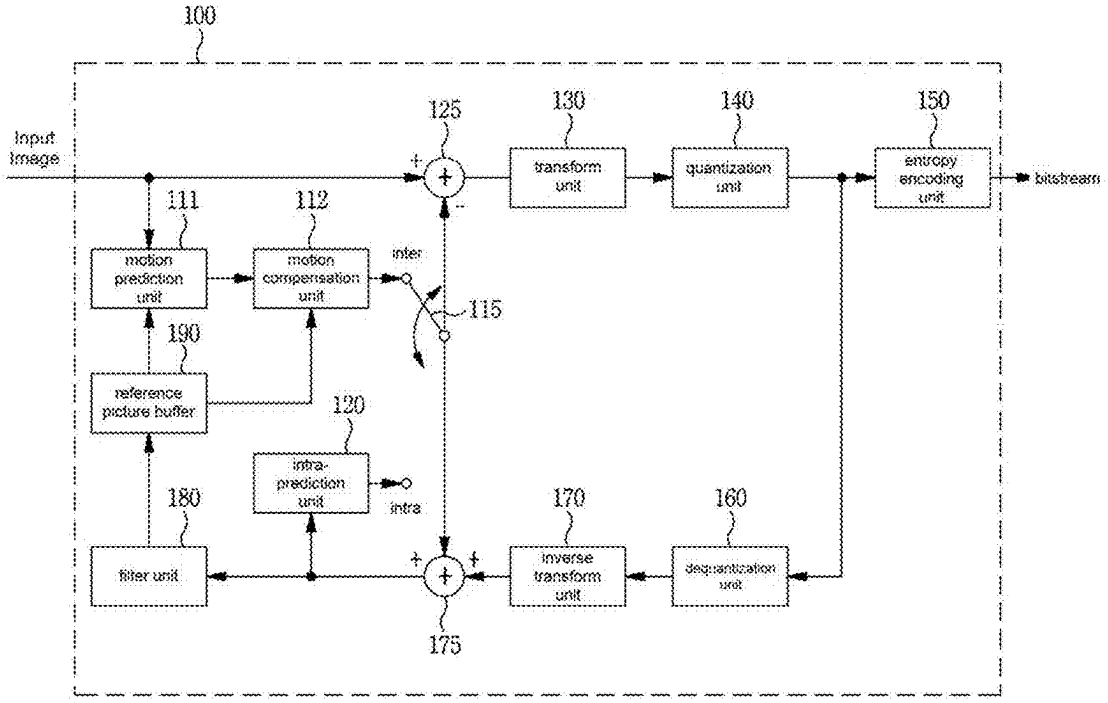

100
Input Image
125
130
transform unit
140
quantization unit
150
entropy encoding unit
bitstream
111
motion prediction unit
112
motion compensation unit
inter
115
190
reference picture buffer
120
intra-prediction unit
intra
180
filter unit
170
inverse transform unit
160
dequantization unit
175

FIG. 2

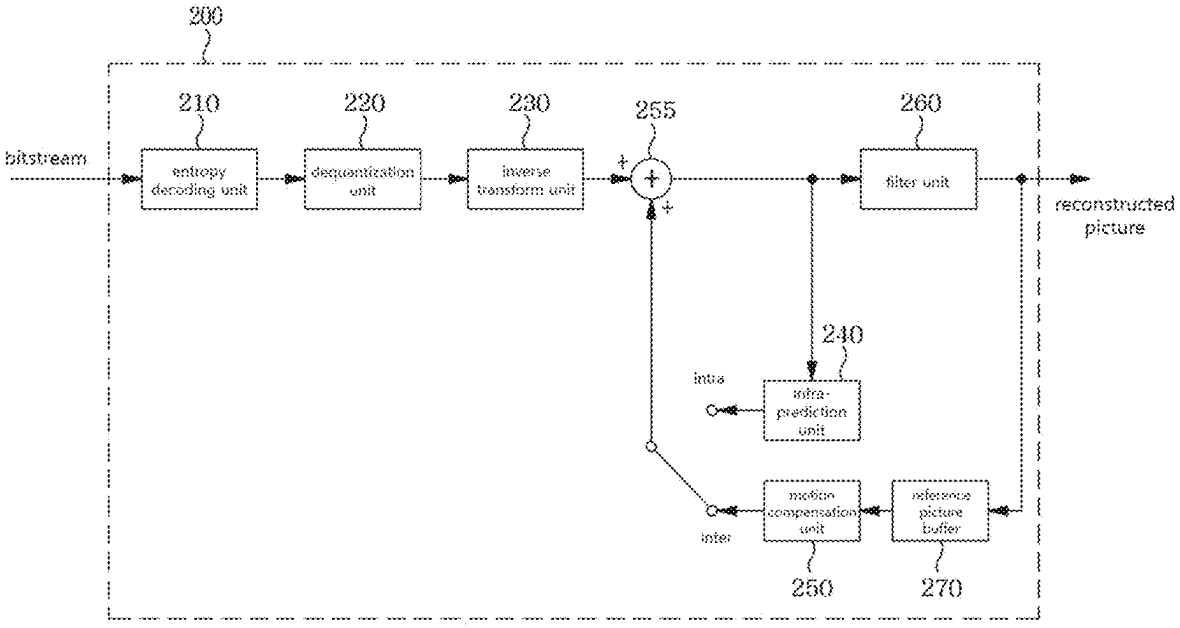

200
210
220
230
255
260
bitstream
entropy decoding unit
dequantization unit
inverse transform unit
filter unit
reconstructed picture
240
intra
intra-prediction unit
motion compensation unit
reference picture buffer
inter
250
270

FIG. 3

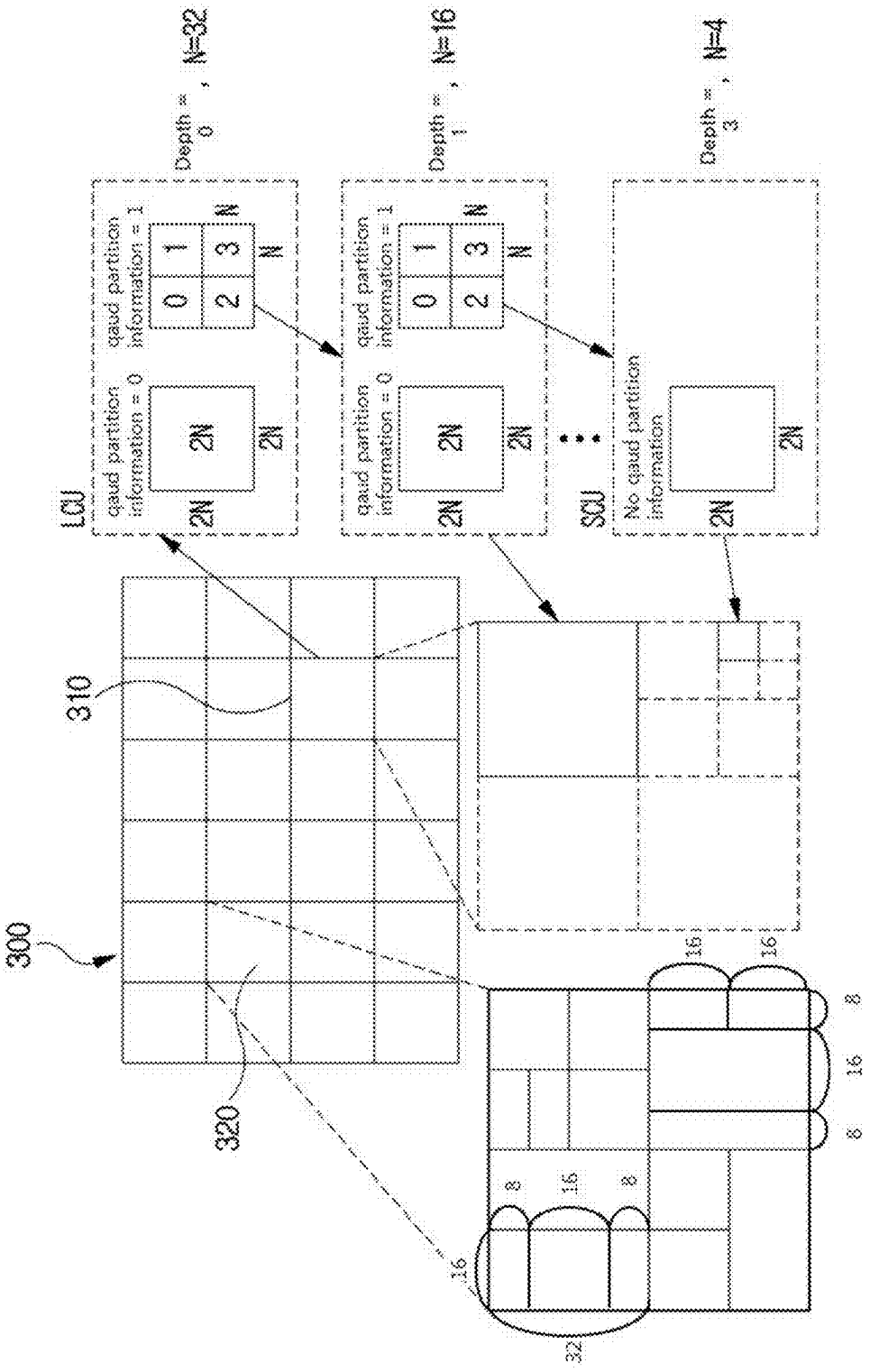

Depth = 0, N=32
Depth = 1, N=16
Depth = 3, N=4
quad partition information = 1
quad partition information = 0
No quad partition information
LCU
SCU
2N
N
0
1
2
3
300
310
320
16
8
32

FIG. 4

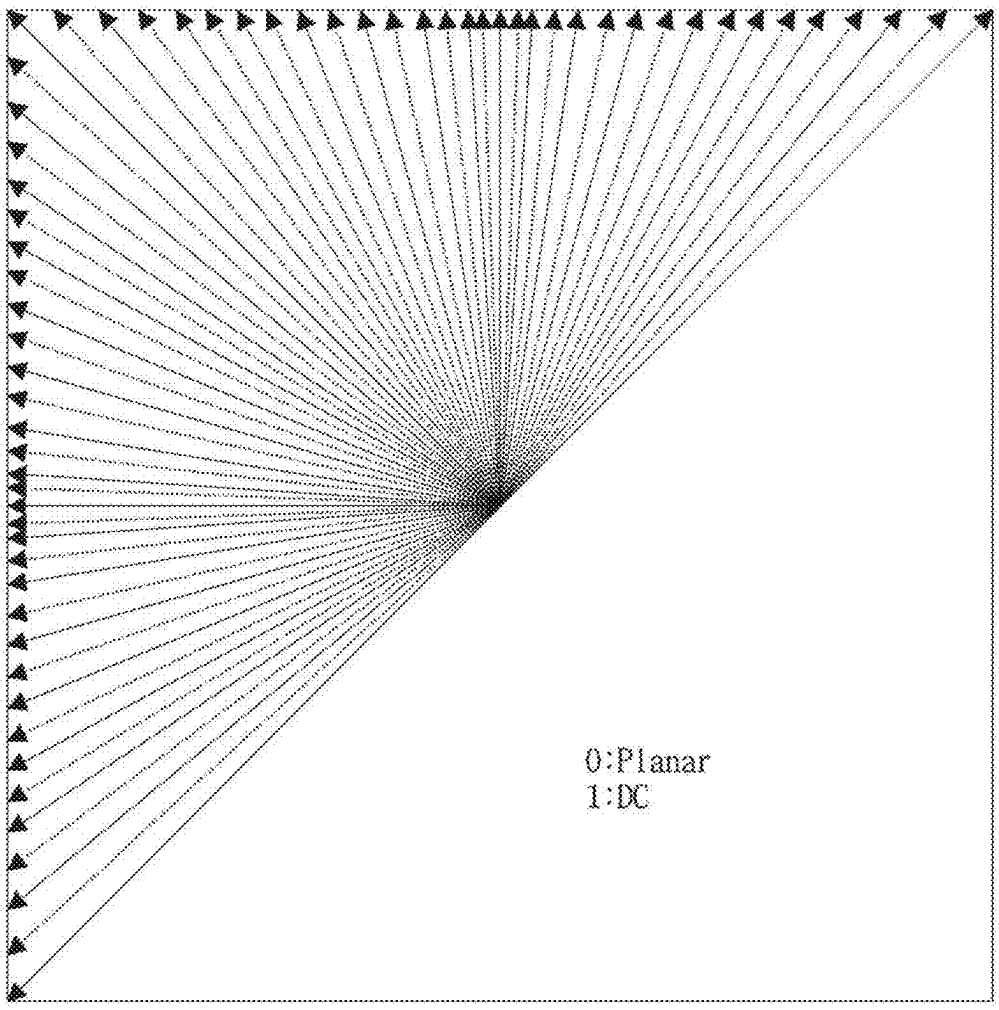


FIG. 5

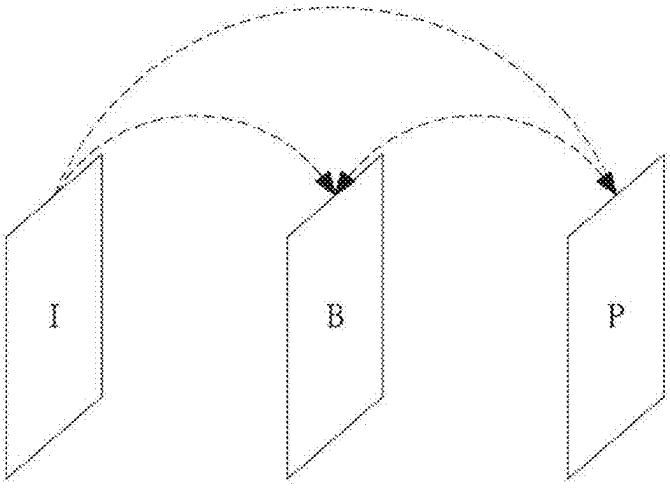

I
B
P

FIG. 6

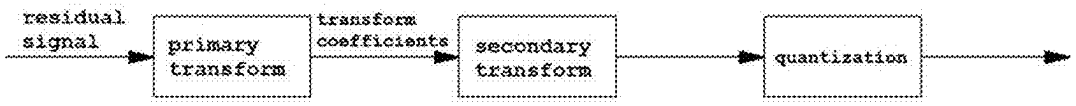

residual signal
primary transform
transform coefficients
secondary transform
quantization

FIG. 7

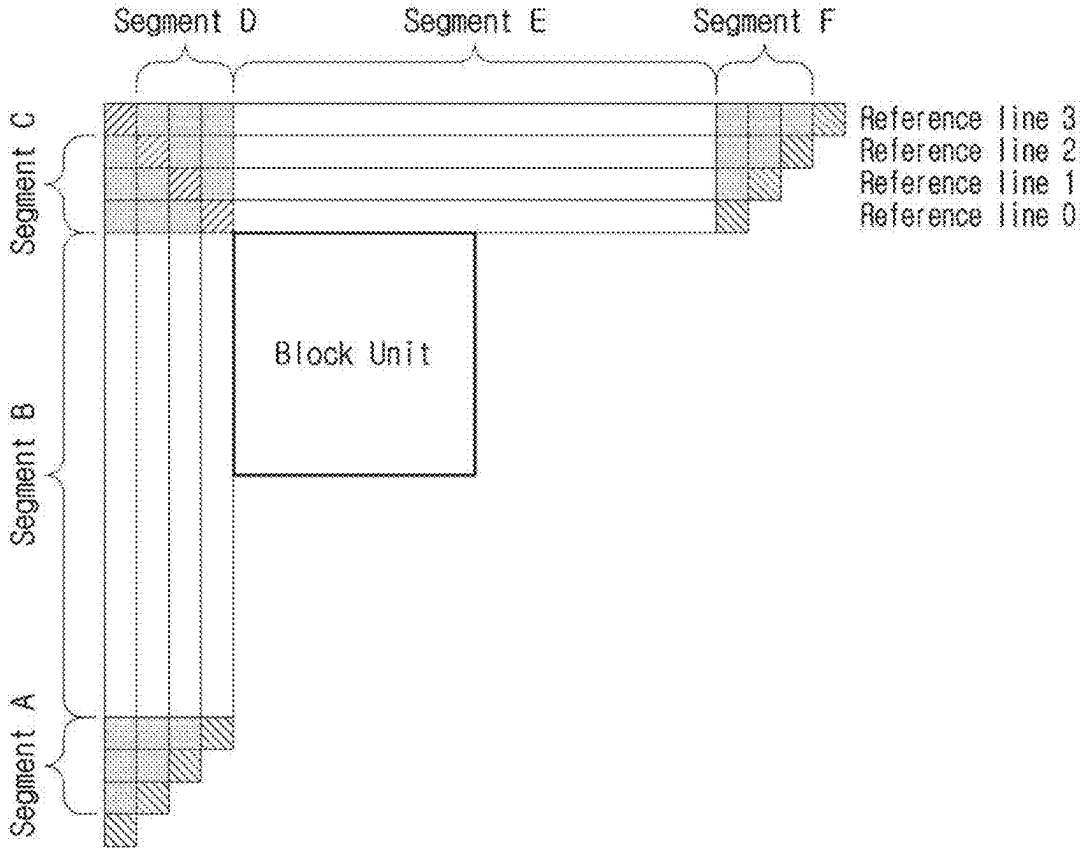

Segment D
Segment E
Segment F
Segment C
Segment B
Segment A
Reference line 3
Reference line 2
Reference line 1
Reference line 0
Block Unit

FIG. 8

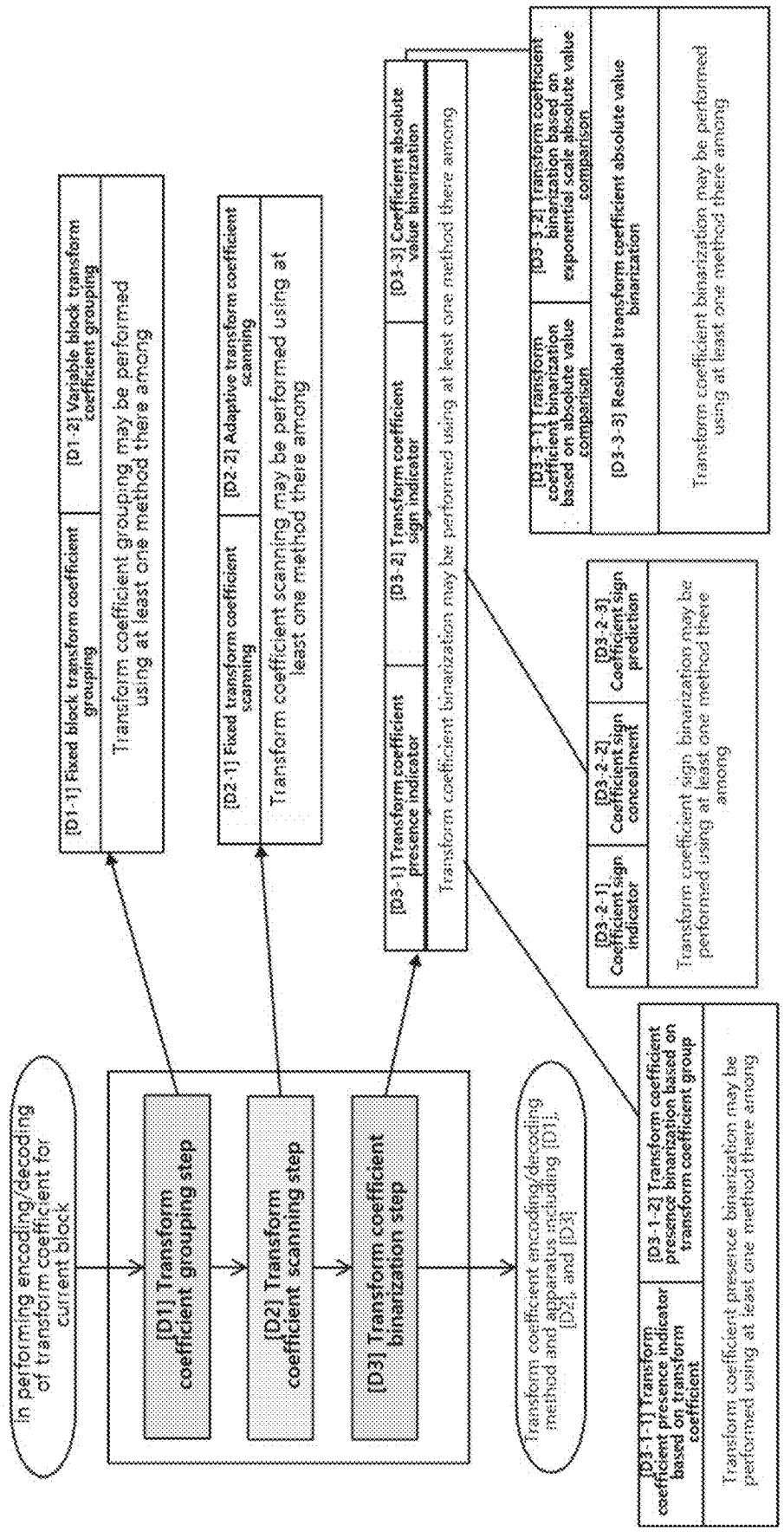

In performing encoding/decoding of transform coefficient for current block
[D1] Transform coefficient grouping step
[D2] Transform coefficient scanning step
[D3] Transform coefficient binarization step
Transform coefficient encoding/decoding method and apparatus including [D1], [D2], and [D3]
[D1-1] Fixed block transform coefficient grouping
[D1-2] Variable block transform coefficient grouping
Transform coefficient grouping may be performed using at least one method there among
[D2-1] Fixed transform coefficient scanning
[D2-2] Adaptive transform coefficient scanning
Transform coefficient scanning may be performed using at least one method there among
[D3-1] Transform coefficient presence indicator
[D3-2] Transform coefficient sign indicator
[D3-3] Coefficient absolute value binarization
Transform coefficient binarization may be performed using at least one method there among
[D3-1-1] Transform coefficient presence indicator based on transform coefficient
[D3-1-2] Transform coefficient presence binarization based on transform coefficient group
Transform coefficient presence binarization may be performed using at least one method there among
[D3-2-1] Coefficient sign indicator
[D3-2-2] Coefficient sign concealment
[D3-2-3] Coefficient sign prediction
Transform coefficient sign binarization may be performed using at least one method there among
[D3-3-1] Transform coefficient binarization based on absolute value comparison
[D3-3-2] Transform coefficient binarization based on exponential scale absolute value comparison
[D3-3-3] Residual transform coefficient absolute value binarization
Transform coefficient binarization may be performed using at least one method there among

FIG. 9

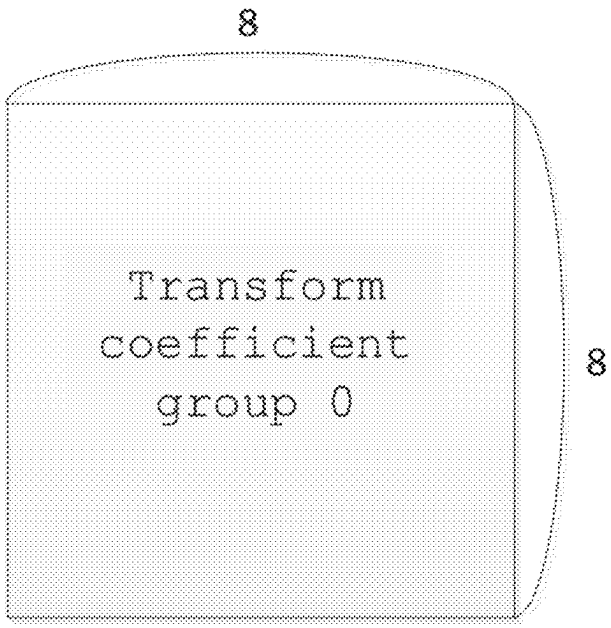

8
Transform coefficient group 0
8

FIG. 10

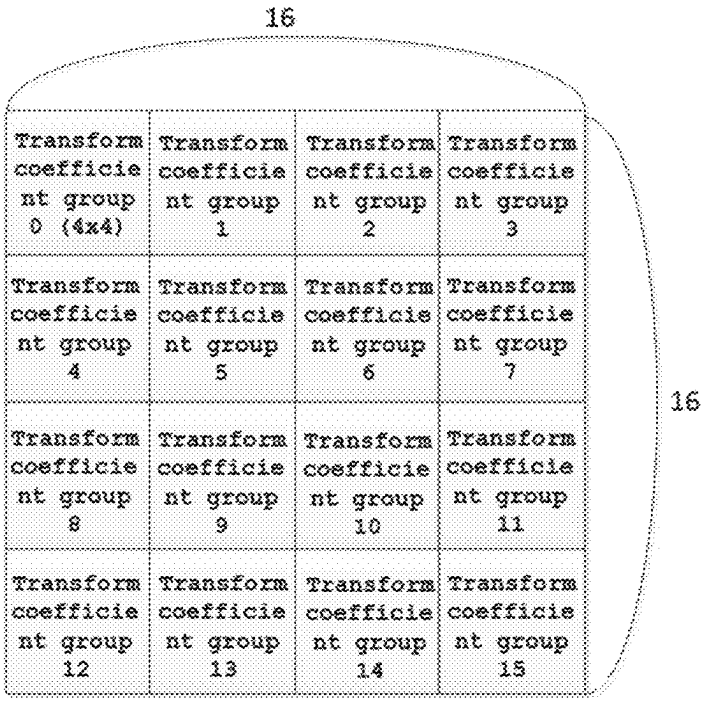

16
Transform coefficient group 0 (4x4)
Transform coefficient group 1
Transform coefficient group 2
Transform coefficient group 3
Transform coefficient group 4
Transform coefficient group 5
Transform coefficient group 6
Transform coefficient group 7
Transform coefficient group 8
Transform coefficient group 9
Transform coefficient group 10
Transform coefficient group 11
Transform coefficient group 12
Transform coefficient group 13
Transform coefficient group 14
Transform coefficient group 15
16

FIG. 11

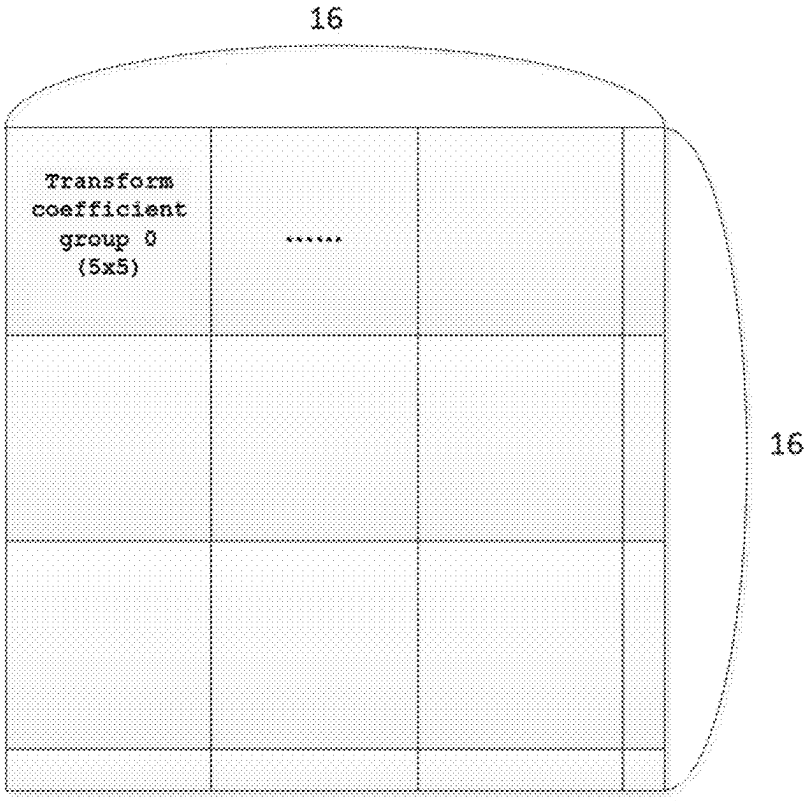


FIG. 12

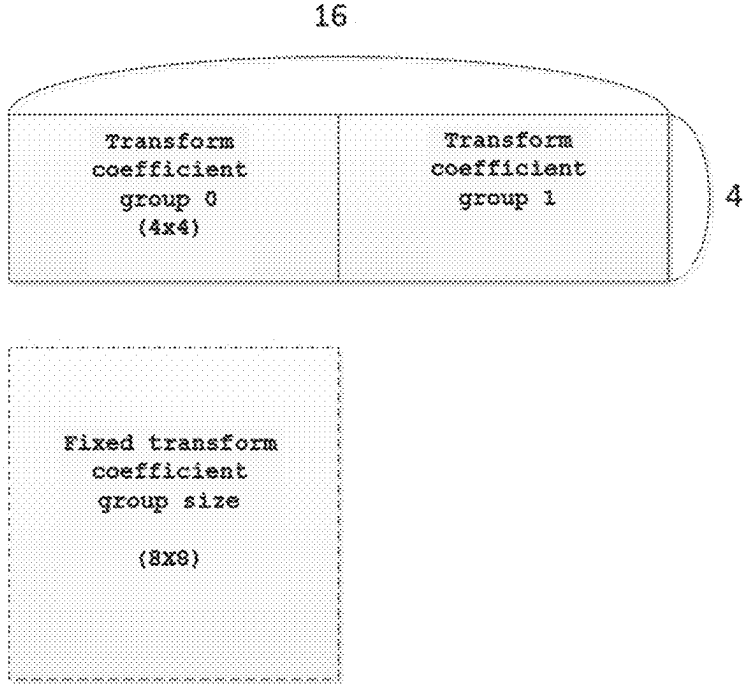

16
Transform coefficient group 0 (4x4)
Transform coefficient group 1
4
Fixed transform coefficient group size
(8X8)

FIG. 13

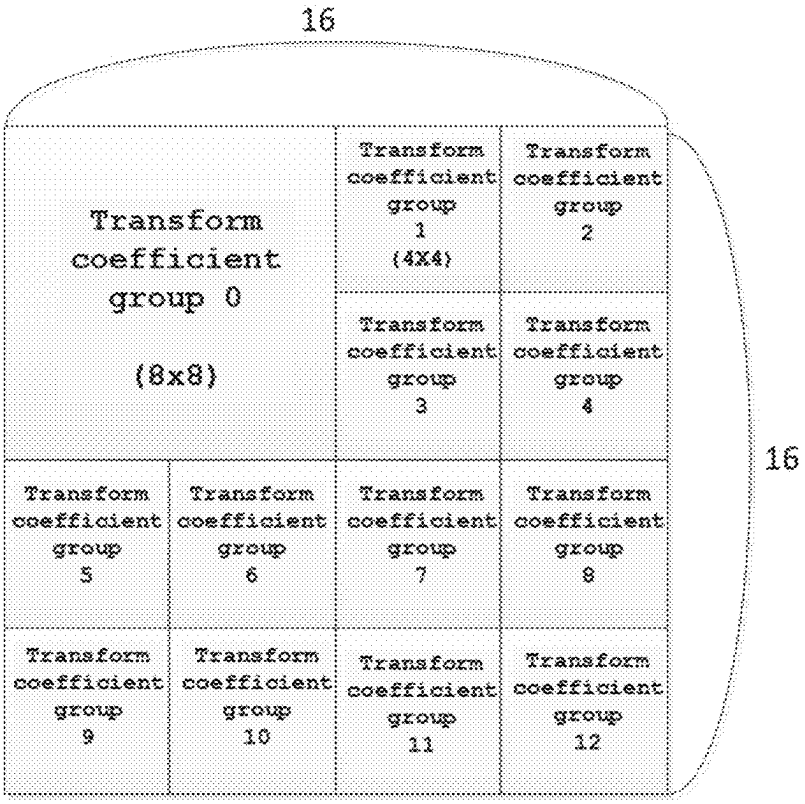

16
Transform coefficient group 0 (8x8)
Transform coefficient group 1 (4X4)
Transform coefficient group 2
Transform coefficient group 3
Transform coefficient group 4
Transform coefficient group 5
Transform coefficient group 6
Transform coefficient group 7
Transform coefficient group 8
Transform coefficient group 9
Transform coefficient group 10
Transform coefficient group 11
Transform coefficient group 12
16

FIG. 14

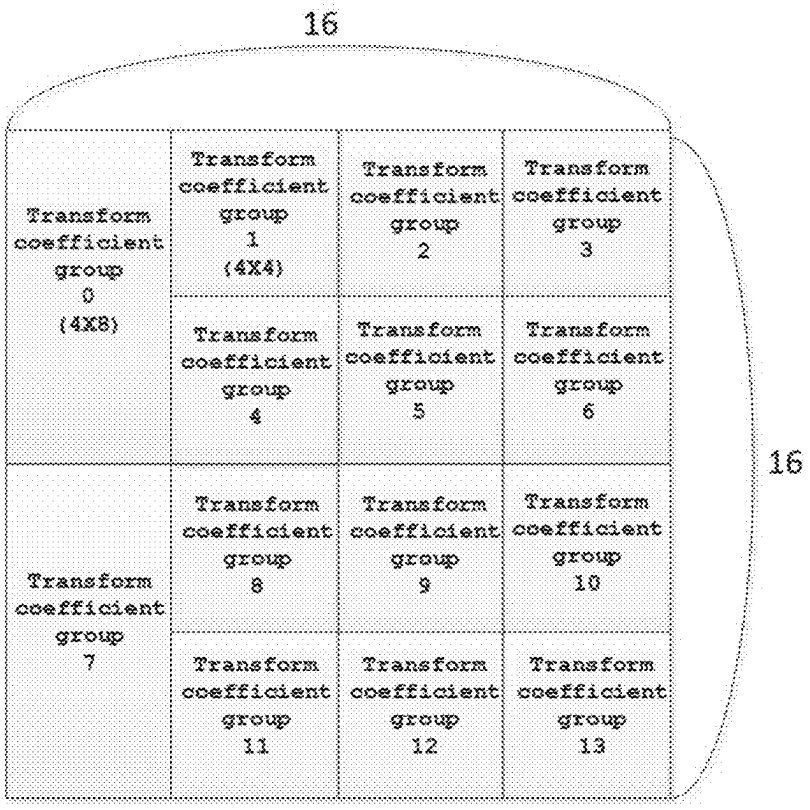

16
Transform coefficient group 0 (4X8)
Transform coefficient group 1 (4X4)
Transform coefficient group 2
Transform coefficient group 3
Transform coefficient group 4
Transform coefficient group 5
Transform coefficient group 6
Transform coefficient group 7
Transform coefficient group 8
Transform coefficient group 9
Transform coefficient group 10
Transform coefficient group 11
Transform coefficient group 12
Transform coefficient group 13
16

FIG. 15

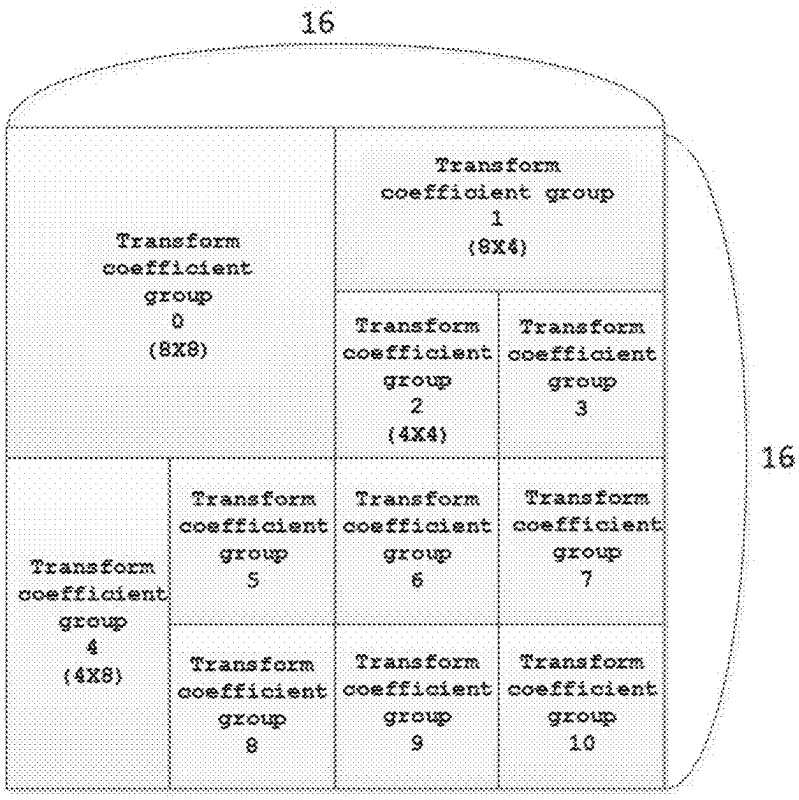

16
Transform coefficient group 0 (8X8)
Transform coefficient group 1 (8X4)
Transform coefficient group 2 (4X4)
Transform coefficient group 3
Transform coefficient group 4 (4X8)
Transform coefficient group 5
Transform coefficient group 6
Transform coefficient group 7
Transform coefficient group 8
Transform coefficient group 9
Transform coefficient group 10
16

FIG. 16

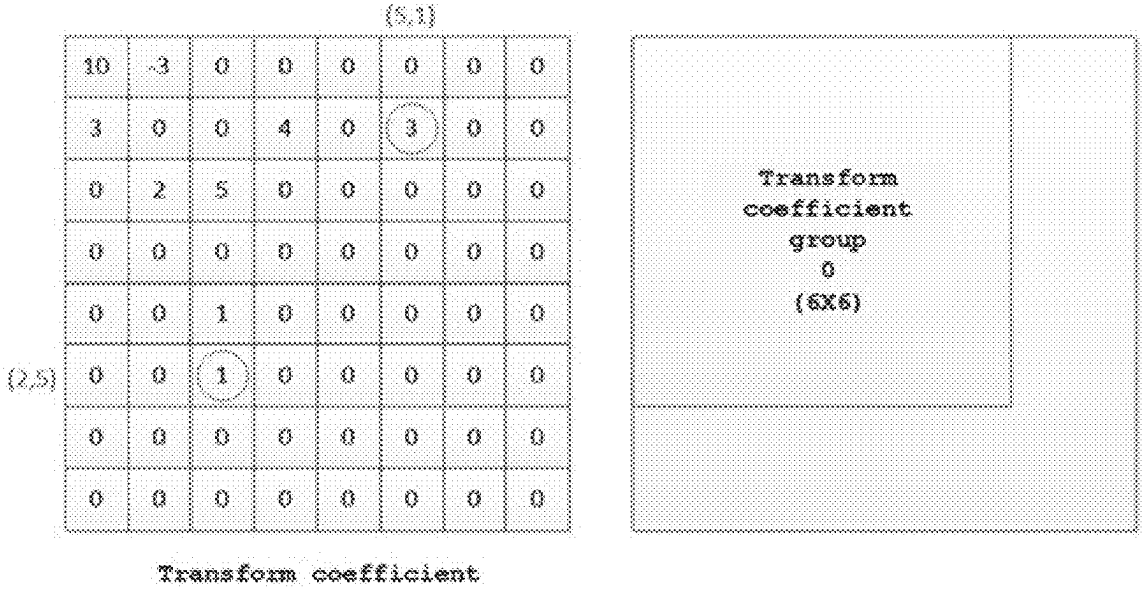

(5,1)
10 -3 0 0 0 0 0 0
3 0 0 4 0 3 0 0
0 2 5 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0
(2,5) 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
Transform coefficient
Transform coefficient group 0 (6X6)

FIG. 17

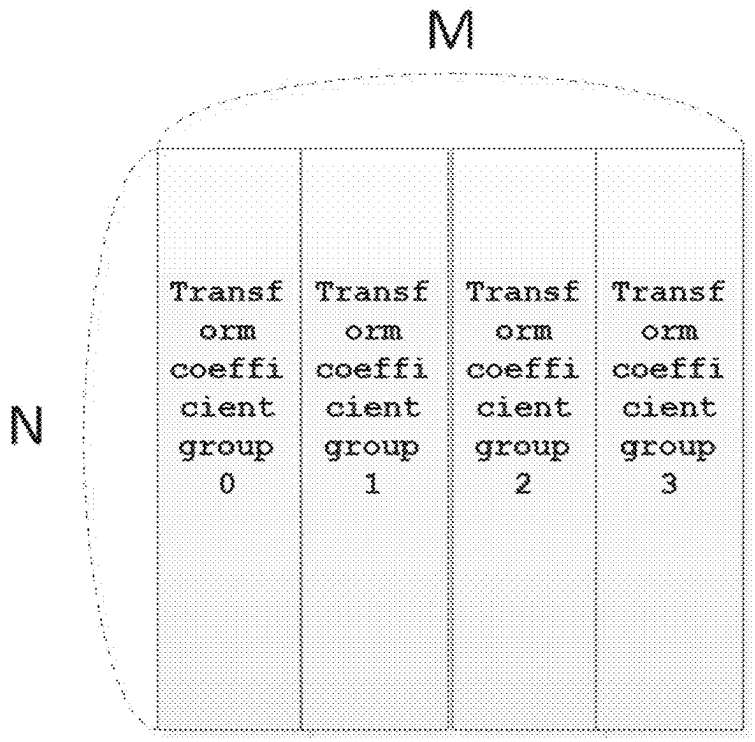

M
N
Transform coefficient group 0
Transform coefficient group 1
Transform coefficient group 2
Transform coefficient group 3

FIG. 18

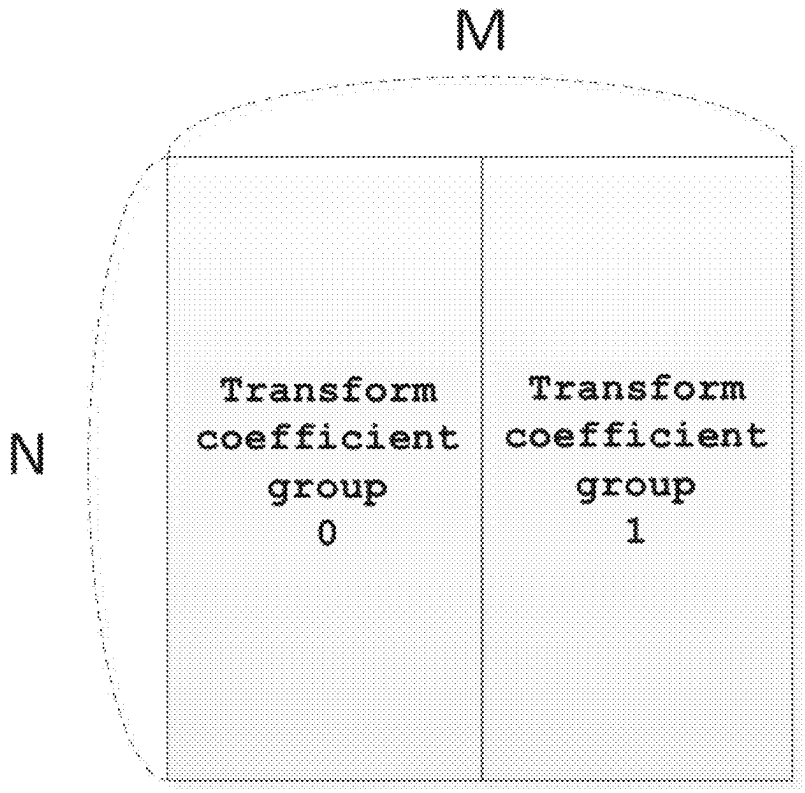

M
N
Transform coefficient group 0
Transform coefficient group 1

FIG. 19

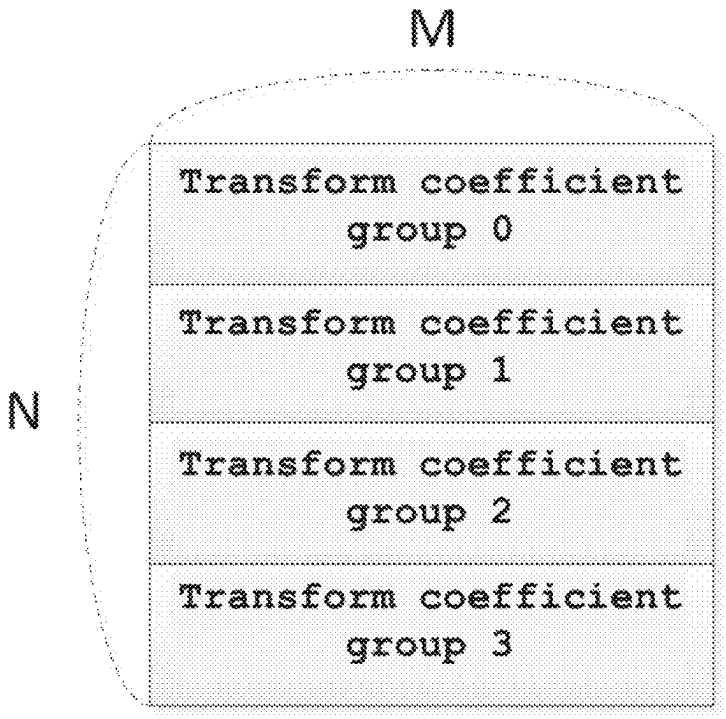

M
N
Transform coefficient group 0
Transform coefficient group 1
Transform coefficient group 2
Transform coefficient group 3

FIG. 20

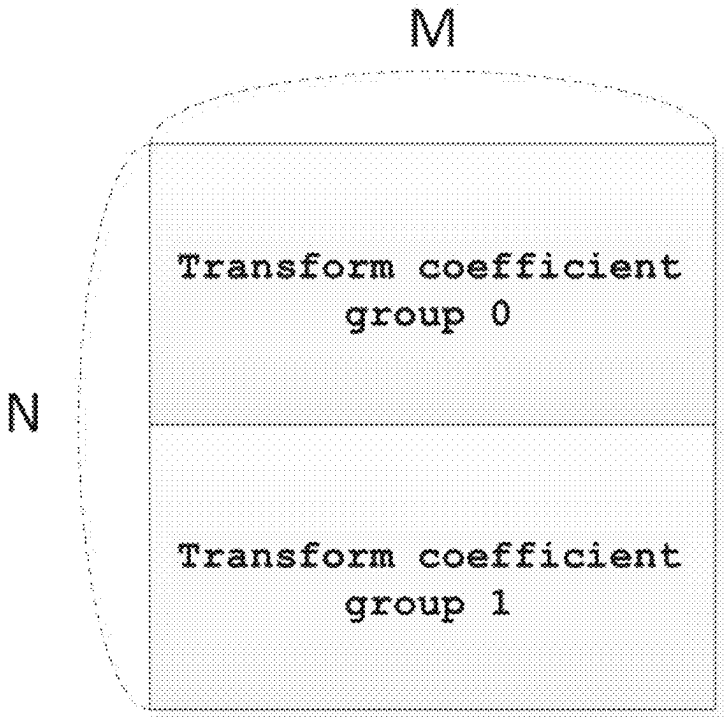

M
N
Transform coefficient group 0
Transform coefficient group 1

FIG. 21

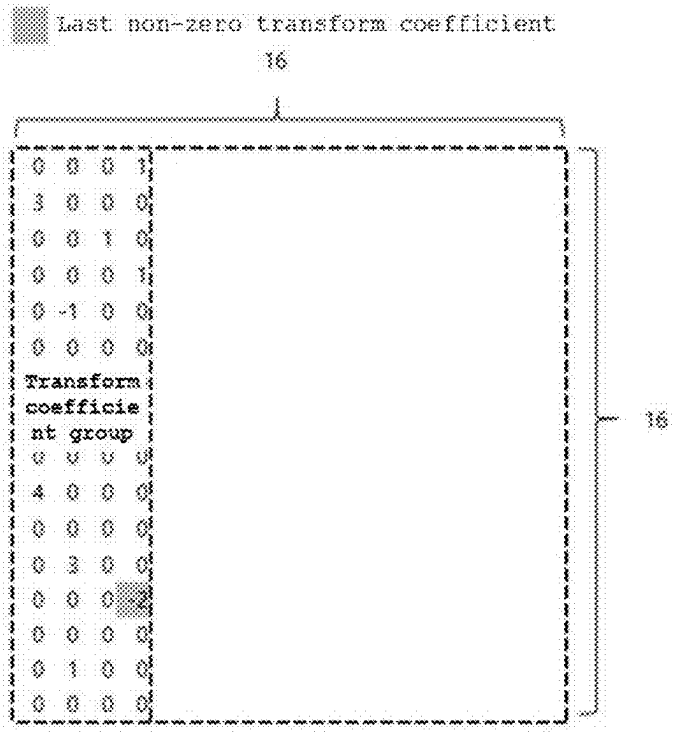

Last non-zero transform coefficient
16
16
Transform coefficient group

FIG. 22

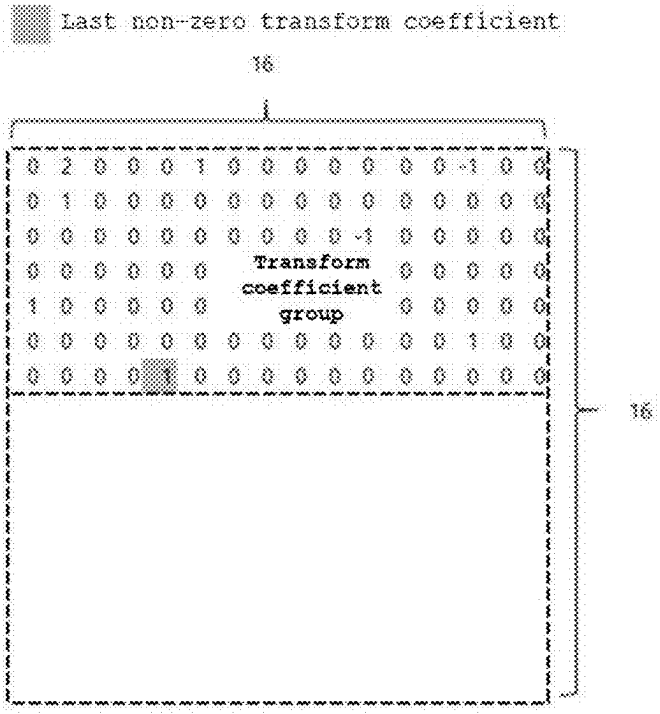

Last non-zero transform coefficient
16
Transform coefficient group
16

FIG. 23

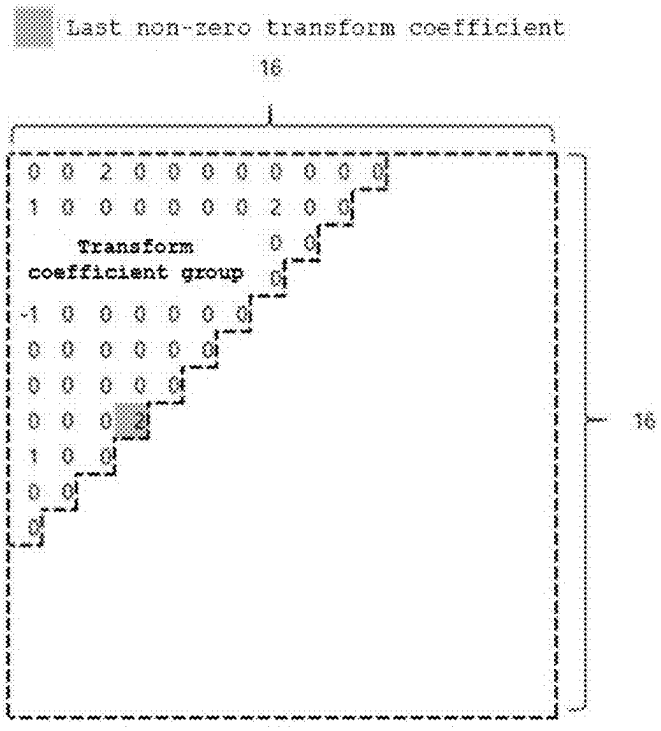

Last non-zero transform coefficient
16
Transform coefficient group
16

FIG. 24

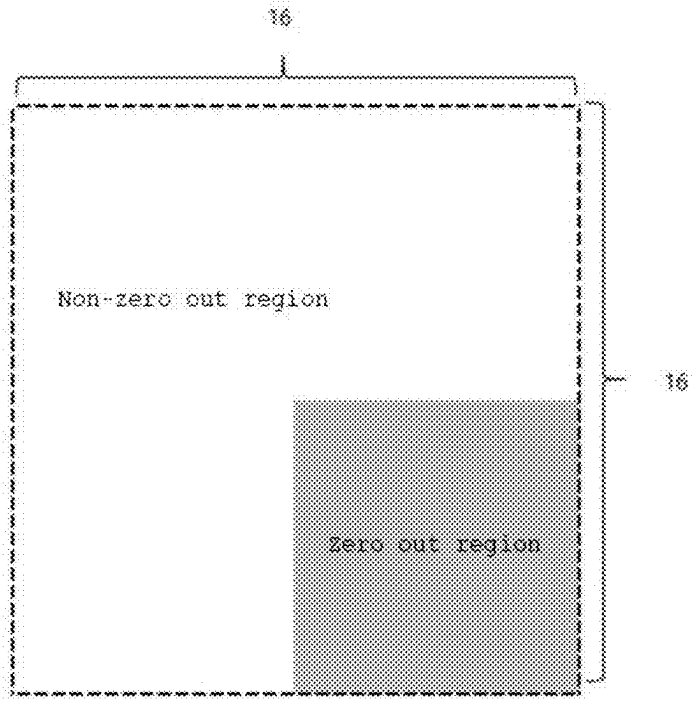

16
Non-zero out region
Zero out region
16

FIG. 25

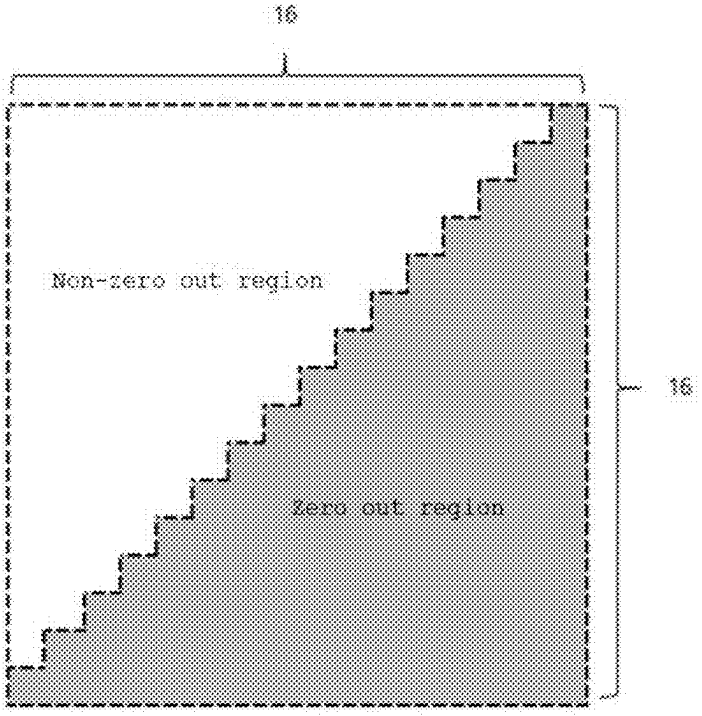


FIG. 26

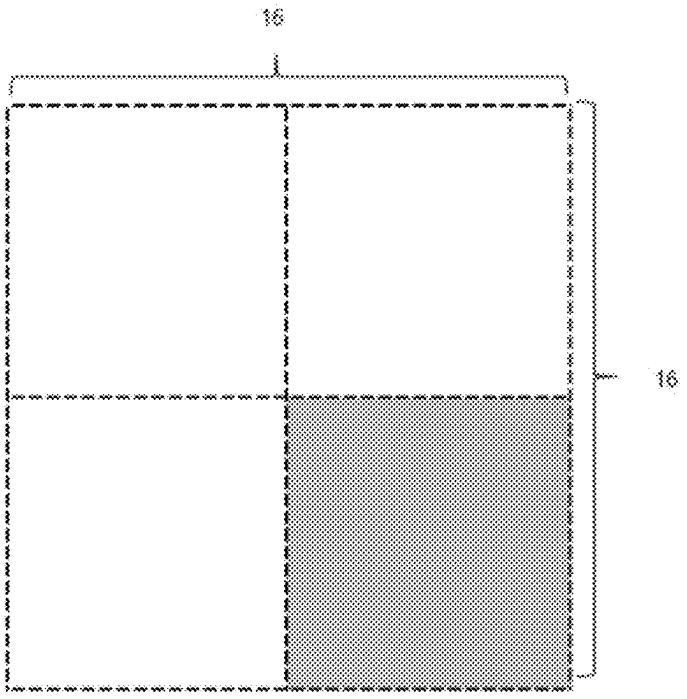


FIG. 27

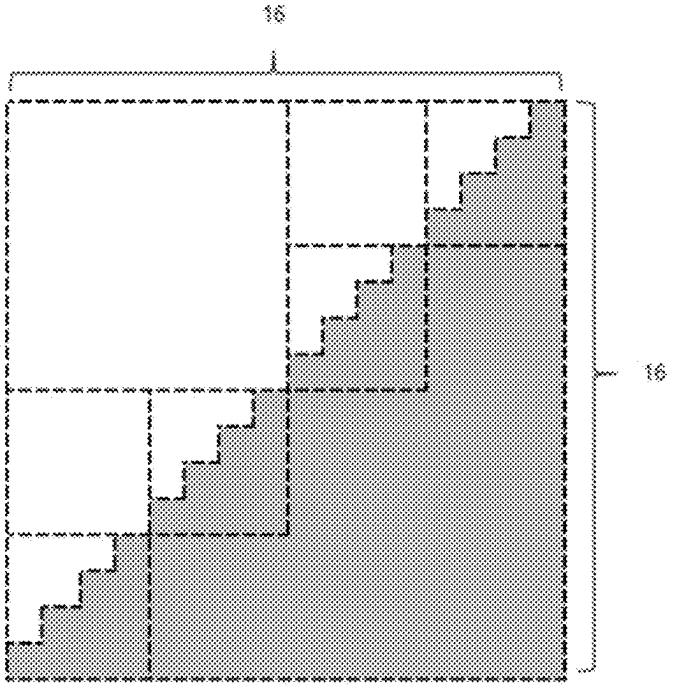
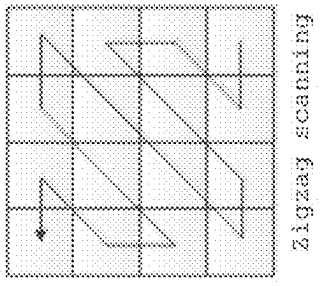
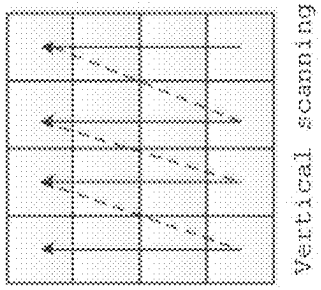
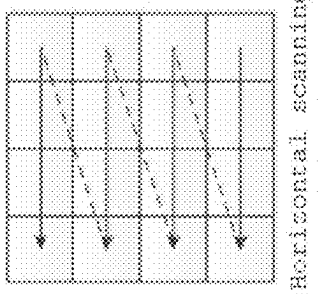


FIG. 28


Zigzag scanning


Vertical scanning


Horizontal scanning

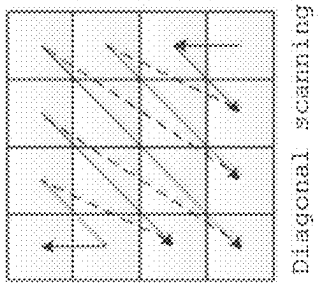

Diagonal scanning

FIG. 29

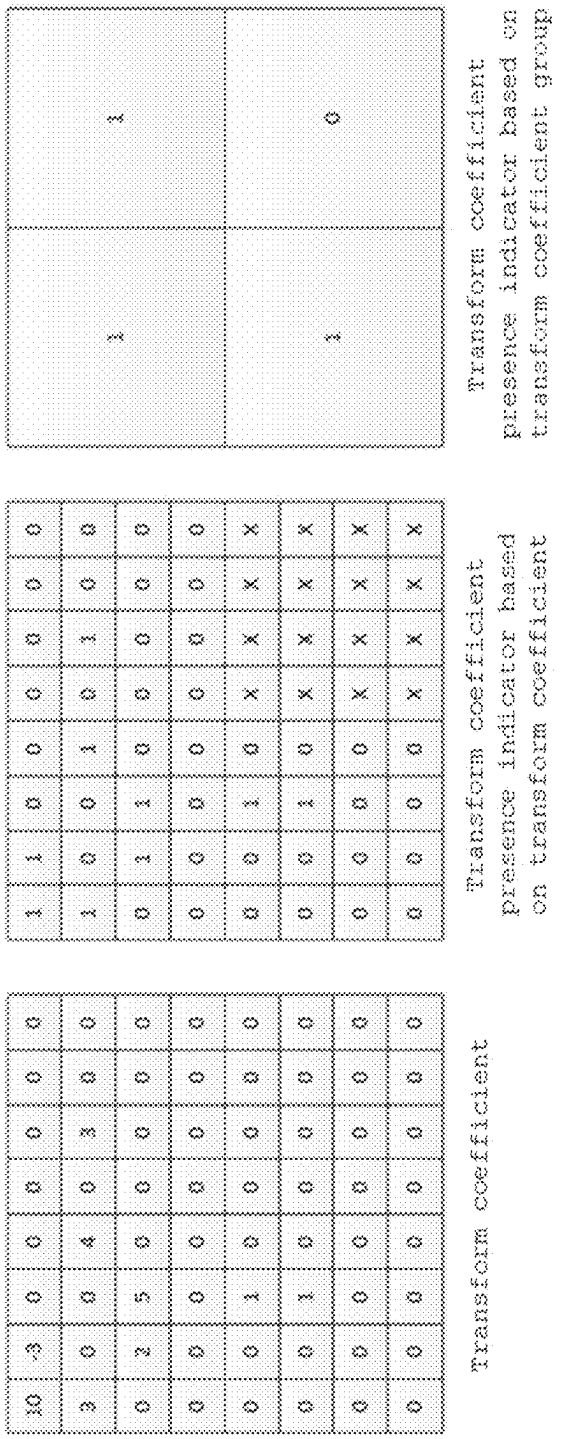

10 -3 0 0 0 0 0 0
3 0 0 4 0 3 0 0
0 2 5 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0
Transform coefficient
1 1 0 0 0 0 0 0
1 0 0 1 0 0 0 0
0 3 1 0 0 0 0 0
0 0 0 0 0 0 0 0
0 0 1 0 x x x x
0 0 1 0 x x x x
0 0 0 0 x x x x
0 0 0 0 x x x x
Transform coefficient presence indicator based on transform coefficient
1 1
1 0
Transform coefficient presence indicator based on transform coefficient group

FIG. 30

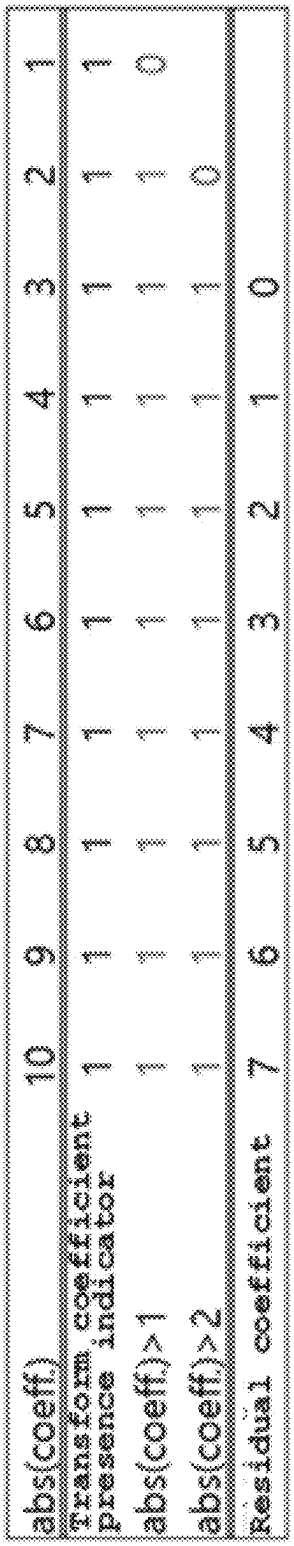

abs(coeff) 10 9 8 7 6 5 4 3 2 1
Transform coefficient presence indicator 1 1 1 1 1 1 1 1 1 1
abs(coeff)>1 1 1 1 1 1 1 1 1 0
abs(coeff)>2 1 1 1 1 1 1 1 1 0
Residual coefficient 7 6 5 4 3 2 1 0

FIG. 31

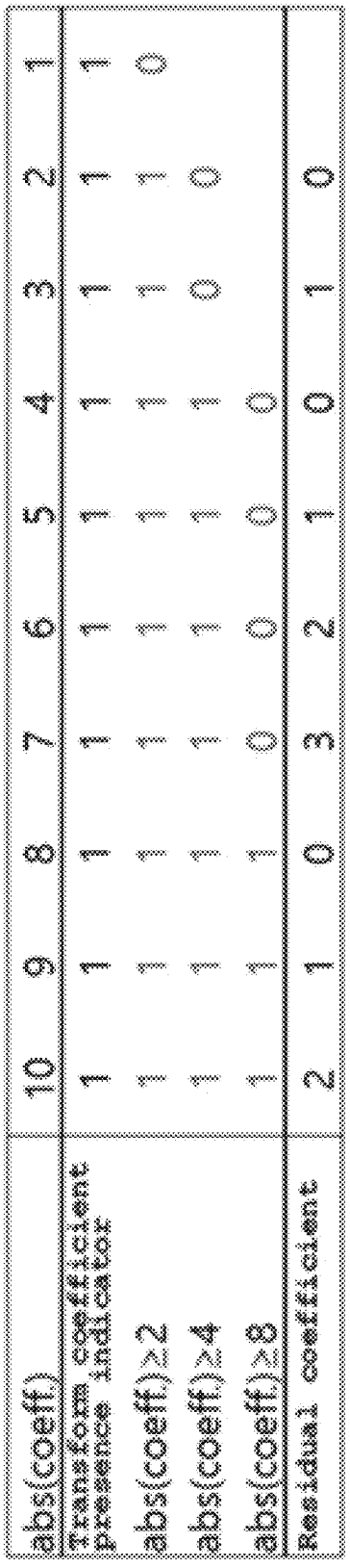
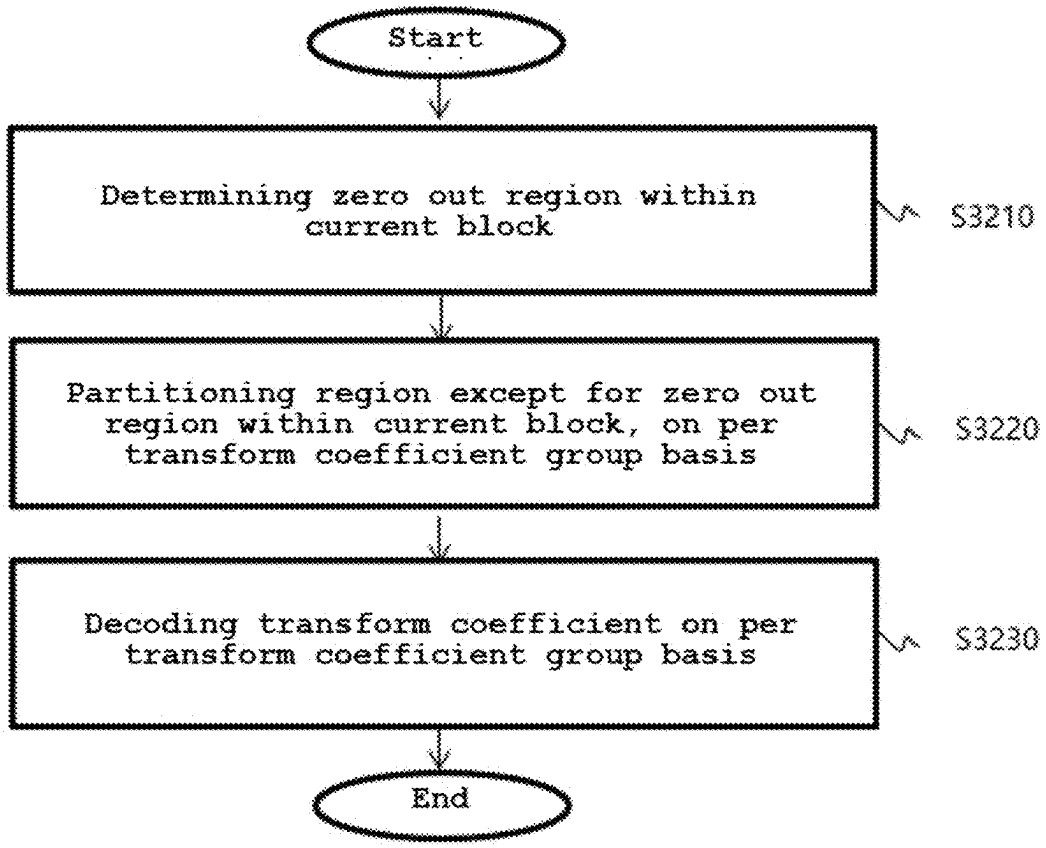

abs(coeff)
10 9 8 7 6 5 4 3 2 1
Transform coefficient presence indicator
1 1 1 1 1 1 1 1 1 1
abs(coeff)≥2
1 1 1 1 1 1 1 1 1 0
abs(coeff)≥4
1 1 1 1 1 1 1 0 0
abs(coeff)≥8
1 1 1 0 0 0 0
Residual coefficient
2 1 0 3 2 1 0 1 0

FIG. 32


Start
Determining zero out region within current block
S3210
Partitioning region except for zero out region within current block, on per transform coefficient group basis
S3220
Decoding transform coefficient on per transform coefficient group basis
S3230
End

METHOD AND DEVICE FOR IMAGE ENCODING/DECODING, AND RECORDING MEDIUM HAVING BITSTREAM STORED THEREON

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application is a continuation of U.S. application Ser. No. 17/259,842, filed on Jan. 12, 2021, which is a National Phase Entry Application of PCT Application No. PCT/KR2019/008299, filed on Jul. 5, 2019, which claims priority to Korean Patent Application No. 10-2018-0081836, filed on Jul. 13, 2018, and Korean Patent Application No. 10-2019-0027678, filed on Mar. 11, 2019, in the Korean Intellectual Property Office, the entire contents of which are hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present invention relates to a method and apparatus for encoding/decoding image and recording medium for storing bitstream. In particular, the present invention relates to an improved transform coefficient encoding/decoding method.

BACKGROUND ART

[0003] Recently, demands for high-resolution and high-quality images such as high definition (H D) images and ultra-high definition (U H D) images have increased in various application fields. However, higher resolution and quality image data has increasing amounts of data in comparison with conventional image data. Therefore, when transmitting image data by using a medium such as conventional wired and wireless broadband networks, or when storing image data by using a conventional storage medium, costs of transmitting and storing increase. In order to solve these problems occurring with an increase in resolution and quality of image data, a high-efficiency image encoding/decoding technique for an image with high-resolution and high image quality is required.

[0004] Image compression techniques include various techniques, such as an inter prediction technique of predicting a pixel value included in a current picture from the preceding or following picture of the current picture, an intra prediction technique of predicting a pixel value included in a current picture by using pixel information within the current picture, transform and quantization techniques for compressing the energy of a residual signal, an entropy encoding technique of assigning a short code to a value with a high appearance frequency and of assigning a long code to a value with a low appearance frequency, and the like. These image compression techniques are used to compress image data effectively for transmission or storage.

[0005] In video encoding/decoding, a residual signal resulting from prediction encoding is transformed into a frequency domain, and a coefficient thereof is quantized to obtain a transform coefficient, and then encoding/decoding is performed on the transform coefficient. In transform coefficient encoding, various techniques are used: a transform coefficient grouping technique for determining a coding unit of a transform coefficient; a transform coefficient scanning technique for arranging coefficients of a transform coefficient group processed in a 2D form into a 1D form; and a transform coefficient binarization and entropy encoding technique for binary encoding coefficient values arranged in one dimension and representing the resulting values into a bit string for final storage and transmission.

[0006] For effective encoding of the transform coefficients, transform coefficient grouping, which can well reflect the values of the transform coefficients generated and the distribution characteristic, and a transform coefficient scanning method are required. However, in the related art, the same scanning method is applied to a transform coefficient group in a fixed size and to a transform coefficient group within a coding block, so that there is a limit of the performance in encoding the transform coefficients.

[0007] A Iso, according to a quantization parameter and the positions of the transform coefficients, the values of the transform coefficients vary in size. H ow ever, in the related art, only one transform coefficient binarization method which is relatively effective in the case where the values of the transform coefficients are small is applied, so that there is a limit of the performance in encoding the transform coefficients.

[0008] In the related art, to encode a transform coefficient of a current coding block, a coding block is divided into transform coefficient groups in a fixed size, and the transform coefficient groups resulting from the division are subjected to transform coefficient encoding using the same scanning method. Since within one coding block, the same scanning method applied to the transform coefficient group in the same size is applied, there is a limit in that the distribution characteristic of the transform coefficients within the coding block is not adaptively reflected.

[0009] According to an applied quantization parameter and the positions of the transform coefficients within a transform block, the values of the transform coefficients generated vary in size. When the value of the quantization parameter is small due to the characteristics of an image or the transform coefficient corresponds to a low frequency component of the image, the value of the transform coefficient is large. However, in the related art, the method of binary encoding the transform coefficient mainly considers the case where the values of the transform coefficients are not large. Thus, a transform coefficient binarization method that can well reflect the case where the values of the transform coefficients are large is required.

DISCLOSURE

Technical Problem

[0010] The present invention is intended to adaptively and flexibly configure a transform coefficient grouping, which is a unit of transform coefficient encoding, and a transform coefficient scanning method.

[0011] A Iso, the present invention is intended to propose a binarization method that is capable of well representing a transform coefficient in a case where a value of a transform coefficient is large, compared to the conventional transform coefficient binarization method.

Technical Solution

[0012] A method of decoding an image according to the present invention, the method may comprise determining a zero out region within a current block, partitioning a region except for the zero out region within the current block, on a

per transform coefficient group basis and decoding a transform coefficient on the per transform coefficient group basis.

[0013] In the method of decoding an image according to the present invention, wherein at the determining of the zero out region, when a width or a height of the current block is larger than a first predefined size, a region of which a size is equal to or larger than the first predefined size within the current block is determined to be the zero out region.

[0014] In the method of decoding an image according to the present invention, wherein at the determining of the zero out region, the determining is based on a type of frequency transform of the current block.

[0015] In the method of decoding an image according to the present invention, wherein at the determining of the zero out region, when a type of frequency transform of the current block is DST-7 or DCT-8, a region of which a size is equal to or larger than a second predefined size within the current block is determined to be the zero out region.

[0016] In the method of decoding an image according to the present invention, wherein at the determining of the zero out region, when a type of frequency transform of the current block is DCT-2, a region of which a size is equal to or larger than a third predefined size within the current block is determined to be the zero out region.

[0017] In the method of decoding an image according to the present invention, wherein a size of the transform coefficient group is determined on the basis of a width and a height of the current block.

[0018] In the method of decoding an image according to the present invention, wherein when a width or a height of the current block is smaller than a fourth predefined size, a size of the transform coefficient group is determined to be a fifth predefined size, and when the width and the height of the current block is larger than the fourth predefined size, the size of the transform coefficient group is determined to be a sixth predefined size, wherein the sixth predefined size is larger than the fifth predefined size.

[0019] In the method of decoding an image according to the present invention, wherein a shape of the transform coefficient group is determined to be a non-square shape when an area of the current block is larger than a predefined area and a shape of the current block is a non-square shape.

[0020] A method of encoding an image according to the present invention, the method may comprise determining a zero out region within a current block, partitioning a region except for the zero out region within the current block, on a per transform coefficient group basis and encoding a transform coefficient on the per transform coefficient group basis.

[0021] In the method of encoding an image according to the present invention, wherein at the determining of the zero out region, when a width or a height of the current block is larger than a first predefined size, a region of which a size is equal to or larger than the first predefined size within the current block is determined to be the zero out region.

[0022] In the method of encoding an image according to the present invention, wherein at the determining of the zero out region, the determining is based on a type of frequency transform of the current block.

[0023] In the method of encoding an image according to the present invention, wherein at the determining of the zero out region, when a type of frequency transform of the current block is DST-7 or DCT-8, a region of which a size is equal to or larger than a second predefined size within the current block is determined to be the zero out region.

[0024] In the method of encoding an image according to the present invention, wherein at the determining of the zero out region, when a type of frequency transform of the current block is DCT-2, a region of which a size is equal to or larger than a third predefined size within the current block is determined to be the zero out region.

[0025] In the method of encoding an image according to the present invention, wherein a size of the transform coefficient group is determined on the basis of a width and a height of the current block.

[0026] In the method of encoding an image according to the present invention, wherein when a width or a height of the current block is smaller than a fourth predefined size, a size of the transform coefficient group is determined to be a fifth predefined size, and when the width and the height of the current block is larger than the fourth predefined size, the size of the transform coefficient group is determined to be a sixth predefined size, wherein the sixth predefined size is larger than the fifth predefined size.

[0027] In the method of encoding an image according to the present invention, wherein a shape of the transform coefficient group is determined to be a non-square shape when an area of the current block is larger than a predefined area and a shape of the current block is a non-square shape.

Advantageous Effects

[0028] The present invention proposes transform coefficient groups in various sizes within a coding block and different scanning methods for the respective transform coefficient groups so that highly flexible transform coefficient grouping and transform coefficient scanning are possible, and proposes an effective binarization method for a transform coefficient having a large value so that much effective transform coefficient encoding is possible.

[0029] The present invention can reduce the number of transform coefficients that need to be encoded through efficient transform coefficient grouping and scanning methods.

[0030] The present invention can perform efficient binarization in which a transform coefficient of which a value is large is assigned relatively few binary bits.

[0031] According to the present invention, image encoding and decoding efficiency can be enhanced.

[0032] According to the present invention, the computational complexity of an image encoder and an image decoder can be reduced.

DESCRIPTION OF DRAWINGS

[0033] FIG. **1** is a block diagram illustrating a configuration according to an example of an encoding apparatus to which the present invention applies.

[0034] FIG. **2** is a block diagram illustrating a configuration according to an example of a decoding apparatus to which the present invention applies.

[0035] FIG. **3** is a diagram schematically illustrating a partitioning structure of an image when encoding and decoding the image.

[0036] FIG. **4** is a diagram illustrating an example of an intra prediction process.

[0037] FIG. **5** is a diagram illustrating an example of an inter prediction process.

[0038] FIG. **6** is a diagram illustrating transform and quantization processes.

[0039] FIG. **7** is a diagram illustrating reference samples available for intra prediction.

[0040] FIG. **8** is a diagram illustrating filtering for block boundaries adjacent to each other according to an embodiment of the present invention.

[0041] FIGS. **9** to **12** are diagrams illustrating an example of fixed block transform coefficient grouping of the present invention.

[0042] FIGS. **13** to **15** are diagrams illustrating an example of variable block transform coefficient grouping of the present invention.

[0043] FIG. **16** is a diagram illustrating a method of determining a transform coefficient group on the basis of a position of a non-zero transform coefficient within a current block.

[0044] FIGS. **17** to **20** are diagrams illustrating how to determine a transform coefficient grouping method on the basis of a width (M) or a height (N) of a current block (M×N).

[0045] FIGS. **21** to **23** are diagrams illustrating a method of setting a transform coefficient group by using an intra prediction mode and position information of the last non-zero coefficient.

[0046] FIGS. **24** and **25** are diagrams illustrating a zero out region and a non-zero out region according to an embodiment of the present invention.

[0047] FIGS. **26** and **27** are diagrams illustrating transform coefficient grouping in a non-zero out region.

[0048] FIG. **28** is a diagram illustrating examples of a fixed transform coefficient scanning method.

[0049] FIG. **29** is a diagram illustrating a transform coefficient presence indicator based on a transform coefficient and a transform coefficient presence indicator based on a transform coefficient group.

[0050] FIG. **30** is a diagram illustrating transform coefficient binarization based on absolute value comparison.

[0051] FIG. **31** is a diagram illustrating transform coefficient binarization based on exponential scale comparison.

[0052] FIG. **32** is a flowchart illustrating a method of decoding an image according to an embodiment of the present invention.

MODE FOR INVENTION

[0053] A variety of modifications may be made to the present invention and there are various embodiments of the present invention, examples of which will now be provided with reference to drawings and described in detail. However, the present invention is not limited thereto, although the exemplary embodiments can be construed as including all modifications, equivalents, or substitutes in a technical concept and a technical scope of the present invention. The similar reference numerals refer to the same or similar functions in various aspects. In the drawings, the shapes and dimensions of elements may be exaggerated for clarity. In the following detailed description of the present invention, references are made to the accompanying drawings that show, by way of illustration, specific embodiments in which the invention may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to implement the present disclosure. It should be understood that various embodiments of the present disclosure, although different, are not necessarily mutually exclusive. For example, specific features, structures, and characteristics described herein, in connection with one embodiment, may be implemented within other embodiments without departing from the spirit and scope of the present disclosure. In addition, it should be understood that the location or arrangement of individual elements within each disclosed embodiment may be modified without departing from the spirit and scope of the present disclosure. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present disclosure is defined only by the appended claims, appropriately interpreted, along with the full range of equivalents to what the claims claim.

[0054] Terms used in the specification, 'first', 'second', etc. can be used to describe various components, but the components are not to be construed as being limited to the terms. The terms are only used to differentiate one component from other components. For example, the 'first' component may be named the 'second' component without departing from the scope of the present invention, and the 'second' component may also be similarly named the 'first' component. The term 'and/or' includes a combination of a plurality of items or any one of a plurality of terms.

[0055] It will be understood that when an element is simply referred to as being 'connected to' or 'coupled to' another element without being 'directly connected to' or 'directly coupled to' another element in the present description, it may be 'directly connected to' or 'directly coupled to' another element or be connected to or coupled to another element, having the other element intervening therebetween. In contrast, it should be understood that when an element is referred to as being "directly coupled" or "directly connected" to another element, there are no intervening elements present.

[0056] Furthermore, constitutional parts shown in the embodiments of the present invention are independently shown so as to represent characteristic functions different from each other. Thus, it does not mean that each constitutional part is constituted in a constitutional unit of separated hardware or software. In other words, each constitutional part includes each of enumerated constitutional parts for convenience. Thus, at least two constitutional parts of each constitutional part may be combined to form one constitutional part or one constitutional part may be divided into a plurality of constitutional parts to perform each function. The embodiment where each constitutional part is combined and the embodiment where one constitutional part is divided are also included in the scope of the present invention, if not departing from the essence of the present invention.

[0057] The terms used in the present specification are merely used to describe particular embodiments, and are not intended to limit the present invention. An expression used in the singular encompasses the expression of the plural, unless it has a clearly different meaning in the context. In the present specification, it is to be understood that terms such as "including", "having", etc. are intended to indicate the existence of the features, numbers, steps, actions, elements, parts, or combinations thereof disclosed in the specification, and are not intended to preclude the possibility that one or more other features, numbers, steps, actions, elements, parts, or combinations thereof may exist or may be added. In other words, when a specific element is referred to as being "included", elements other than the corresponding element are not excluded, but additional elements may be included in embodiments of the present invention or the scope of the present invention.

[0058] In addition, some of constituents may not be indispensable constituents performing essential functions of the present invention but be selective constituents improving only performance thereof. The present invention may be implemented by including only the indispensable constitutional parts for implementing the essence of the present invention except the constituents used in improving performance. The structure including only the indispensable constituents except the selective constituents used in improving only performance is also included in the scope of the present invention.

[0059] Hereinafter, embodiments of the present invention will be described in detail with reference to the accompanying drawings. In describing exemplary embodiments of the present invention, well-known functions or constructions will not be described in detail since they may unnecessarily obscure the understanding of the present invention. The same constituent elements in the drawings are denoted by the same reference numerals, and a repeated description of the same elements will be omitted.

[0060] Hereinafter, an image may mean a picture configuring a video, or may mean the video itself. For example, "encoding or decoding or both of an image" may mean "encoding or decoding or both of a moving picture", and may mean "encoding or decoding or both of one image among images of a moving picture."

[0061] Hereinafter, terms "moving picture" and "video" may be used as the same meaning and be replaced with each other.

[0062] Hereinafter, a target image may be an encoding target image which is a target of encoding and/or a decoding target image which is a target of decoding. A Iso, a target image may be an input image inputted to an encoding apparatus, and an input image inputted to a decoding apparatus. H ere, a target image may have the same meaning with the current image.

[0063] Hereinafter, terms "image", "picture, "frame" and "screen" may be used as the same meaning and be replaced with each other.

[0064] Hereinafter, a target block may be an encoding target block which is a target of encoding and/or a decoding target block which is a target of decoding. A Iso, a target block may be the current block which is a target of current encoding and/or decoding. For example, terms "target block" and "current block" may be used as the same meaning and be replaced with each other.

[0065] Hereinafter, terms "block" and "unit" may be used as the same meaning and be replaced with each other. Or a "block" may represent a specific unit.

[0066] Hereinafter, terms "region" and "segment" may be replaced with each other.

[0067] Hereinafter, a specific signal may be a signal representing a specific block. For example, an original signal may be a signal representing a target block. A prediction signal may be a signal representing a prediction block. A residual signal may be a signal representing a residual block.

[0068] In embodiments, each of specific information, data, flag, index, element and attribute, etc. may have a value. A value of information, data, flag, index, element and attribute equal to "0" may represent a logical false or the first predefined value. In other words, a value "0", a false, a logical false and the first predefined value may be replaced with each other. A value of information, data, flag, index, element and attribute equal to "1" may represent a logical true or the second predefined value. In other words, a value "1", a true, a logical true and the second predefined value may be replaced with each other.

[0069] When a variable i or j is used for representing a column, a row or an index, a value of i may be an integer equal to or greater than 0, or equal to or greater than 1. That is, the column, the row, the index, etc. may be counted from 0 or may be counted from 1.

DESCRIPTION OF TERMS

[0070] Encoder: means an apparatus performing encoding. That is, means an encoding apparatus.

[0071] Decoder: means an apparatus performing decoding. That is, means an decoding apparatus.

[0072] Block: is an M×N array of a sample. Herein, M and N may mean positive integers, and the block may mean a sample array of a two-dimensional form. The block may refer to a unit. A current block my mean an encoding target block that becomes a target when encoding, or a decoding target block that becomes a target when decoding. In addition, the current block may be at least one of an encode block, a prediction block, a residual block, and a transform block.

[0073] Sample: is a basic unit constituting a block. It may be expressed as a value from 0 to 2 Bd−1 according to a bit depth (Bd). In the present invention, the sample may be used as a meaning of a pixel. That is, a sample, a pel, a pixel may have the same meaning with each other.

[0074] Unit: may refer to an encoding and decoding unit. When encoding and decoding an image, the unit may be a region generated by partitioning a single image. In addition, the unit may mean a subdivided unit when a single image is partitioned into subdivided units during encoding or decoding. That is, an image may be partitioned into a plurality of units. When encoding and decoding an image, a predetermined process for each unit may be performed. A single unit may be partitioned into sub-units that have sizes smaller than the size of the unit. Depending on functions, the unit may mean a block, a macroblock, a coding tree unit, a code tree block, a coding unit, a coding block), a prediction unit, a prediction block, a residual unit), a residual block, a transform unit, a transform block, etc. In addition, in order to distinguish a unit from a block, the unit may include a luma component block, a chroma component block associated with the luma component block, and a syntax element of each color component block. The unit may have various sizes and forms, and particularly, the form of the unit may be a two-dimensional geometrical figure such as a square shape, a rectangular shape, a trapezoid shape, a triangular shape, a pentagonal shape, etc. In addition, unit information may include at least one of a unit type indicating the coding unit, the prediction unit, the transform unit, etc., and a unit size, a unit depth, a sequence of encoding and decoding of a unit, etc.

[0075] Coding Tree Unit: is configured with a single coding tree block of a luma component Y, and two coding tree blocks related to chroma components Cb and Cr. In addition, it may mean that including the blocks and a syntax element of each block. Each coding tree unit may be partitioned by using at least one of a quad-tree partitioning method, a binary-tree partitioning method and ternary-tree partitioning method to configure a lower unit such as coding unit, prediction unit, transform unit, etc. It may be used as a term for designating a sample block that becomes a process

unit when encoding/decoding an image as an input image. Here, the quad-tree may mean a quarternary-tree.

[0076] When the size of the coding block is within a predetermined range, the division is possible using only quad-tree partitioning. Here, the predetermined range may be defined as at least one of a maximum size and a minimum size of a coding block in which the division is possible using only quad-tree partitioning. Information indicating a maximum/minimum size of a coding block in which quad-tree partitioning is allowed may be signaled through a bitstream, and the information may be signaled in at least one unit of a sequence, a picture parameter, a tile group, or a slice (segment). Alternatively, the maximum/minimum size of the coding block may be a fixed size predetermined in the coder/decoder. For example, when the size of the coding block corresponds to 256×256 to 64×64, the division is possible only using quad-tree partitioning. Alternatively, when the size of the coding block is larger than the size of the maximum conversion block, the division is possible only using quad-tree partitioning. Herein, the block to be divided may be at least one of a coding block and a transform block. In this case, information indicating the division of the coded block (for example, split flag) may be a flag indicating whether or not to perform the quad-tree partitioning. When the size of the coding block falls within a predetermined range, the division is possible only using binary tree or ternary tree partitioning. In this case, the above description of the quad-tree partitioning may be applied to binary tree partitioning or ternary tree partitioning in the same manner.

[0077] Coding Tree Block: may be used as a term for designating any one of a Y coding tree block, Cb coding tree block, and Cr coding tree block.

[0078] Neighbor Block: may mean a block adjacent to a current block. The block adjacent to the current block may mean a block that comes into contact with a boundary of the current block, or a block positioned within a predetermined distance from the current block. The neighbor block may mean a block adjacent to a vertex of the current block. Herein, the block adjacent to the vertex of the current block may mean a block vertically adjacent to a neighbor block that is horizontally adjacent to the current block, or a block horizontally adjacent to a neighbor block that is vertically adjacent to the current block.

[0079] Reconstructed Neighbor block: may mean a neighbor block adjacent to a current block and which has been already spatially/temporally encoded or decoded. Herein, the reconstructed neighbor block may mean a reconstructed neighbor unit. A reconstructed spatial neighbor block may be a block within a current picture and which has been already reconstructed through encoding or decoding or both. A reconstructed temporal neighbor block is a block at a corresponding position as the current block of the current picture within a reference image, or a neighbor block thereof.

[0080] Unit Depth: may mean a partitioned degree of a unit. In a tree structure, the highest node(Root Node) may correspond to the first unit which is not partitioned. Also, the highest node may have the least depth value. In this case, the highest node may have a depth of level 0. A node having a depth of level 1 may represent a unit generated by partitioning once the first unit. A node having a depth of level 2 may represent a unit generated by partitioning twice the first unit. A node having a depth of level n may represent a unit generated by partitioning n-times the first unit. A Leaf Node may be the lowest node and a node which cannot be partitioned further. A depth of a Leaf Node may be the maximum level. For example, a predefined value of the maximum level may be 3. A depth of a root node may be the lowest and a depth of a leaf node may be the deepest. In addition, when a unit is expressed as a tree structure, a level in which a unit is present may mean a unit depth.

[0081] Bitstream: may mean a bitstream including encoding image information.

[0082] Parameter Set: corresponds to header information among a configuration within a bitstream. At least one of a video parameter set, a sequence parameter set, a picture parameter set, and an adaptation parameter set may be included in a parameter set. In addition, a parameter set may include a slice header, a tile group header, and tile header information. The term "tile group" means a group of tiles and has the same meaning as a slice.

[0083] Parsing: may mean determination of a value of a syntax element by performing entropy decoding, or may mean the entropy decoding itself.

[0084] Symbol: may mean at least one of a syntax element, a coding parameter, and a transform coefficient value of an encoding/decoding target unit. In addition, the symbol may mean an entropy encoding target or an entropy decoding result.

[0085] Prediction Mode: may be information indicating a mode encoded/decoded with intra prediction or a mode encoded/decoded with inter prediction.

[0086] Prediction Unit: may mean a basic unit when performing prediction such as inter-prediction, intra-prediction, inter-compensation, intra-compensation, and motion compensation. A single prediction unit may be partitioned into a plurality of partitions having a smaller size, or may be partitioned into a plurality of lower prediction units. A plurality of partitions may be a basic unit in performing prediction or compensation. A partition which is generated by dividing a prediction unit may also be a prediction unit.

[0087] Prediction Unit Partition: may mean a form obtained by partitioning a prediction unit.

[0088] Reference picture list may refer to a list including one or more reference pictures used for inter prediction or motion compensation. There are several types of usable reference picture lists, including LC (List combined), L0 (List 0), L1 (List 1), L2 (List 2), L3 (List 3).

[0089] Inter prediction indicator may refer to a direction of inter prediction (unidirectional prediction, bidirectional prediction, etc.) of a current block. Alternatively, it may refer to the number of reference pictures used to generate a prediction block of a current block. Alternatively, it may refer to the number of prediction blocks used at the time of performing inter prediction or motion compensation on a current block.

[0090] Prediction list utilization flag indicates whether a prediction block is generated using at least one reference picture in a specific reference picture list. A n inter prediction indicator can be derived using a prediction list utilization flag, and conversely, a prediction list utilization flag can be derived using an inter prediction indicator. For example, when the prediction list utilization flag has a first value of zero (0), it means that a reference picture in a reference picture list is not used to generate a prediction block. On the other hand, when the prediction list utilization flag has a second value of one (1), it means that a reference picture list is used to generate a prediction block.

[0091] Reference picture index may refer to an index indicating a specific reference picture in a reference picture list.

[0092] Reference picture may mean a reference picture which is referred to by a specific block for the purposes of inter prediction or motion compensation of the specific block. Alternatively, the reference picture may be a picture including a reference block referred to by a current block for inter prediction or motion compensation. Hereinafter, the terms "reference picture" and "reference image" have the same meaning and can be interchangeably.

[0093] Motion vector may be a two-dimensional vector used for inter prediction or motion compensation. The motion vector may mean an offset between an encoding/decoding target block and a reference block. For example, (mvX, mvY) may represent a motion vector. H ere, mvX may represent a horizontal component and mvY may represent a vertical component.

[0094] Search range may be a two-dimensional region which is searched to retrieve a motion vector during inter prediction. For example, the size of the search range may be M×N. Here, M and N are both integers.

[0095] Motion vector candidate may refer to a prediction candidate block or a motion vector of the prediction candidate block when predicting a motion vector. In addition, a motion vector candidate may be included in a motion vector candidate list.

[0096] Motion vector candidate list may mean a list composed of one or more motion vector candidates.

[0097] Motion vector candidate index may mean an indicator indicating a motion vector candidate in a motion vector candidate list. Alternatively, it may be an index of a motion vector predictor.

[0098] Motion information may mean information including at least one of the items including a motion vector, a reference picture index, an inter prediction indicator, a prediction list utilization flag, reference picture list information, a reference picture, a motion vector candidate, a motion vector candidate index, a merge candidate, and a merge index.

[0099] Merge candidate list may mean a list composed of one or more merge candidates.

[0100] Merge candidate may mean a spatial merge candidate, a temporal merge candidate, a combined merge candidate, a combined bi-predictive merge candidate, or a zero merge candidate. The merge candidate may include motion information such as an inter prediction indicator, a reference picture index for each list, a motion vector, a prediction list utilization flag, and an inter prediction indicator.

[0101] Merge index may mean an indicator indicating a merge candidate in a merge candidate list. Alternatively, the merge index may indicate a block from which a merge candidate has been derived, among reconstructed blocks spatially/temporally adjacent to a current block. Alternatively, the merge index may indicate at least one piece of motion information of a merge candidate.

[0102] Transform Unit: may mean a basic unit when performing encoding/decoding such as transform, inverse-transform, quantization, dequantization, transform coefficient encoding/decoding of a residual signal. A single transform unit may be partitioned into a plurality of lower-level transform units having a smaller size. Here, transformation/inverse-transformation may comprise at least one among the first transformation/the first inverse-transformation and the second transformation/the second inverse-transformation.

[0103] Scaling: may mean a process of multiplying a quantized level by a factor. A transform coefficient may be generated by scaling a quantized level. The scaling also may be referred to as dequantization.

[0104] Quantization Parameter: may mean a value used when generating a quantized level using a transform coefficient during quantization. The quantization parameter also may mean a value used when generating a transform coefficient by scaling a quantized level during dequantization. The quantization parameter may be a value mapped on a quantization step size.

[0105] Delta Quantization Parameter: may mean a difference value between a predicted quantization parameter and a quantization parameter of an encoding/decoding target unit.

[0106] Scan: may mean a method of sequencing coefficients within a unit, a block or a matrix. For example, changing a two-dimensional matrix of coefficients into a one-dimensional matrix may be referred to as scanning, and changing a one-dimensional matrix of coefficients into a two-dimensional matrix may be referred to as scanning or inverse scanning.

[0107] Transform Coefficient: may mean a coefficient value generated after transform is performed in an encoder. It may mean a coefficient value generated after at least one of entropy decoding and dequantization is performed in a decoder. A quantized level obtained by quantizing a transform coefficient or a residual signal, or a quantized transform coefficient level also may fall within the meaning of the transform coefficient.

[0108] Quantized Level: may mean a value generated by quantizing a transform coefficient or a residual signal in an encoder. Alternatively, the quantized level may mean a value that is a dequantization target to undergo dequantization in a decoder. Similarly, a quantized transform coefficient level that is a result of transform and quantization also may fall within the meaning of the quantized level.

[0109] Non-zero Transform Coefficient: may mean a transform coefficient having a value other than zero, or a transform coefficient level or a quantized level having a value other than zero.

[0110] Quantization Matrix: may mean a matrix used in a quantization process or a dequantization process performed to improve subjective or objective image quality. The quantization matrix also may be referred to as a scaling list.

[0111] Quantization Matrix Coefficient: may mean each element within a quantization matrix. The quantization matrix coefficient also may be referred to as a matrix coefficient.

[0112] Default Matrix: may mean a predetermined quantization matrix preliminarily defined in an encoder or a decoder.

[0113] Non-default Matrix: may mean a quantization matrix that is not preliminarily defined in an encoder or a decoder but is signaled by a user.

[0114] Statistic Value: a statistic value for at least one among a variable, an encoding parameter, a constant value, etc. which have a computable specific value may be one or more among an average value, a sum value, a weighted average value, a weighted sum value, the minimum value,

the maximum value, the most frequent value, a median value, an interpolated value of the corresponding specific values.

[0115] FIG. **1** is a block diagram showing a configuration of an encoding apparatus according to an embodiment to which the present invention is applied.

[0116] An encoding apparatus **100** may be an encoder, a video encoding apparatus, or an image encoding apparatus. A video may include at least one image. The encoding apparatus **100** may sequentially encode at least one image.

[0117] Referring to FIG. **1**, the encoding apparatus **100** may include a motion prediction unit **111**, a motion compensation unit **112**, an intra-prediction unit **120**, a switch **115**, a subtractor **125**, a transform unit **130**, a quantization unit **140**, an entropy encoding unit **150**, a dequantization unit **160**, a inverse-transform unit **170**, an adder **175**, a filter unit **180**, and a reference picture buffer **190**.

[0118] The encoding apparatus **100** may perform encoding of an input image by using an intra mode or an inter mode or both. In addition, encoding apparatus **100** may generate a bitstream including encoded information through encoding the input image, and output the generated bitstream. The generated bitstream may be stored in a computer readable recording medium, or may be streamed through a wired/wireless transmission medium. When an intra mode is used as a prediction mode, the switch **115** may be switched to an intra. Alternatively, when an inter mode is used as a prediction mode, the switch **115** may be switched to an inter mode. Herein, the intra mode may mean an intra-prediction mode, and the inter mode may mean an inter-prediction mode. The encoding apparatus **100** may generate a prediction block for an input block of the input image. In addition, the encoding apparatus **100** may encode a residual block using a residual of the input block and the prediction block after the prediction block being generated. The input image may be called as a current image that is a current encoding target. The input block may be called as a current block that is current encoding target, or as an encoding target block.

[0119] When a prediction mode is an intra mode, the intra-prediction unit **120** may use a sample of a block that has been already encoded/decoded and is adjacent to a current block as a reference sample. The intra-prediction unit **120** may perform spatial prediction for the current block by using a reference sample, or generate prediction samples of an input block by performing spatial prediction. Herein, the intra prediction may mean intra-prediction,

[0120] When a prediction mode is an inter mode, the motion prediction unit **111** may retrieve a region that best matches with an input block from a reference image when performing motion prediction, and deduce a motion vector by using the retrieved region. In this case, a search region may be used as the region. The reference image may be stored in the reference picture buffer **190**. Here, when encoding/decoding for the reference image is performed, it may be stored in the reference picture buffer **190**.

[0121] The motion compensation unit **112** may generate a prediction block by performing motion compensation for the current block using a motion vector. Herein, inter-prediction may mean inter-prediction or motion compensation.

[0122] When the value of the motion vector is not an integer, the motion prediction unit **111** and the motion compensation unit **112** may generate the prediction block by applying an interpolation filter to a partial region of the reference picture. In order to perform inter-picture prediction or motion compensation on a coding unit, it may be determined that which mode among a skip mode, a merge mode, an advanced motion vector prediction (AMVP) mode, and a current picture referring mode is used for motion prediction and motion compensation of a prediction unit included in the corresponding coding unit. Then, inter-picture prediction or motion compensation may be differently performed depending on the determined mode.

[0123] The subtractor **125** may generate a residual block by using a difference of an input block and a prediction block. The residual block may be called as a residual signal. The residual signal may mean a difference between an original signal and a prediction signal. In addition, the residual signal may be a signal generated by transforming or quantizing, or transforming and quantizing a difference between the original signal and the prediction signal. The residual block may be a residual signal of a block unit.

[0124] The transform unit **130** may generate a transform coefficient by performing transform of a residual block, and output the generated transform coefficient. Herein, the transform coefficient may be a coefficient value generated by performing transform of the residual block. When a transform skip mode is applied, the transform unit **130** may skip transform of the residual block.

[0125] A quantized level may be generated by applying quantization to the transform coefficient or to the residual signal. Hereinafter, the quantized level may be also called as a transform coefficient in embodiments.

[0126] The quantization unit **140** may generate a quantized level by quantizing the transform coefficient or the residual signal according to a parameter, and output the generated quantized level. Herein, the quantization unit **140** may quantize the transform coefficient by using a quantization matrix.

[0127] The entropy encoding unit **150** may generate a bitstream by performing entropy encoding according to a probability distribution on values calculated by the quantization unit **140** or on coding parameter values calculated when performing encoding, and output the generated bitstream. The entropy encoding unit **150** may perform entropy encoding of sample information of an image and information for decoding an image. For example, the information for decoding the image may include a syntax element.

[0128] When entropy encoding is applied, symbols are represented so that a smaller number of bits are assigned to a symbol having a high chance of being generated and a larger number of bits are assigned to a symbol having a low chance of being generated, and thus, the size of bit stream for symbols to be encoded may be decreased. The entropy encoding unit **150** may use an encoding method for entropy encoding such as exponential Golomb, context-adaptive variable length coding (CAVLC), context-adaptive binary arithmetic coding (CABAC), etc. For example, the entropy encoding unit **150** may perform entropy encoding by using a variable length coding/code (VLC) table. In addition, the entropy encoding unit **150** may deduce a binarization method of a target symbol and a probability model of a target symbol/bin, and perform arithmetic coding by using the deduced binarization method, and a context model.

[0129] In order to encode a transform coefficient level (quantized level), the entropy encoding unit **150** may change a two-dimensional block form coefficient into a one-dimensional vector form by using a transform coefficient scanning method.

[0130] A coding parameter may include information (flag, index, etc.) such as syntax element that is encoded in an encoder and signaled to a decoder, and information derived when performing encoding or decoding. The coding parameter may mean information required when encoding or decoding an image. For example, at least one value or a combination form of a unit/block size, a unit/block depth, unit/block partition information, unit/block shape, unit/block partition structure, whether to partition of a quad-tree form, whether to partition of a binary-tree form, a partition direction of a binary-tree form (horizontal direction or vertical direction), a partition form of a binary-tree form (symmetric partition or asymmetric partition), whether or not a current coding unit is partitioned by ternary tree partitioning, direction (horizontal or vertical direction) of the ternary tree partitioning, type (symmetric or asymmetric type) of the ternary tree partitioning, whether a current coding unit is partitioned by multi-type tree partitioning, direction (horizontal or vertical direction) of the multi-type three partitioning, type (symmetric or asymmetric type) of the multi-type tree partitioning, and a tree (binary tree or ternary tree) structure of the multi-type tree partitioning, a prediction mode(intra prediction or inter prediction), a luma intra-prediction mode/direction, a chroma intra-prediction mode/direction, intra partition information, inter partition information, a coding block partition flag, a prediction block partition flag, a transform block partition flag, a reference sample filtering method, a reference sample filter tab, a reference sample filter coefficient, a prediction block filtering method, a prediction block filter tap, a prediction block filter coefficient, a prediction block boundary filtering method, a prediction block boundary filter tab, a prediction block boundary filter coefficient, an intra-prediction mode, an inter-prediction mode, motion information, a motion vector, a motion vector difference, a reference picture index, a inter-prediction angle, an inter-prediction indicator, a prediction list utilization flag, a reference picture list, a reference picture, a motion vector predictor index, a motion vector predictor candidate, a motion vector candidate list, whether to use a merge mode, a merge index, a merge candidate, a merge candidate list, whether to use a skip mode, an interpolation filter type, an interpolation filter tab, an interpolation filter coefficient, a motion vector size, a presentation accuracy of a motion vector, a transform type, a transform size, information of whether or not a primary (first) transform is used, information of whether or not a secondary transform is used, a primary transform index, a secondary transform index, information of whether or not a residual signal is present, a coded block pattern, a coded block flag(CBF), a quantization parameter, a quantization parameter residue, a quantization matrix, whether to apply an intra loop filter, an intra loop filter coefficient, an intra loop filter tab, an intra loop filter shape/form, whether to apply a deblocking filter, a deblocking filter coefficient, a deblocking filter tab, a deblocking filter strength, a deblocking filter shape/form, whether to apply an adaptive sample offset, an adaptive sample offset value, an adaptive sample offset category, an adaptive sample offset type, whether to apply an adaptive loop filter, an adaptive loop filter coefficient, an adaptive loop filter tab, an adaptive loop filter shape/form, a binarization/inverse-binarization method, a context model determining method, a context model updating method, whether to perform a regular mode, whether to perform a bypass mode, a context bin, a bypass bin, a significant coefficient flag, a last significant coefficient flag, a coded flag for a unit of a coefficient group, a position of the last significant coefficient, a flag for whether a value of a coefficient is larger than 1, a flag for whether a value of a coefficient is larger than 2, a flag for whether a value of a coefficient is larger than 3, information on a remaining coefficient value, a sign information, a reconstructed luma sample, a reconstructed chroma sample, a residual luma sample, a residual chroma sample, a luma transform coefficient, a chroma transform coefficient, a quantized luma level, a quantized chroma level, a transform coefficient level scanning method, a size of a motion vector search area at a decoder side, a shape of a motion vector search area at a decoder side, a number of time of a motion vector search at a decoder side, information on a CTU size, information on a minimum block size, information on a maximum block size, information on a maximum block depth, information on a minimum block depth, an image displaying/outputting sequence, slice identification information, a slice type, slice partition information, tile identification information, a tile type, tile partition information, tile group identification information, a tile group type, tile group partition information, a picture type, a bit depth of an input sample, a bit depth of a reconstruction sample, a bit depth of a residual sample, a bit depth of a transform coefficient, a bit depth of a quantized level, and information on a luma signal or information on a chroma signal may be included in the coding parameter.

[0131] Herein, signaling the flag or index may mean that a corresponding flag or index is entropy encoded and included in a bitstream by an encoder, and may mean that the corresponding flag or index is entropy decoded from a bitstream by a decoder.

[0132] When the encoding apparatus **100** performs encoding through inter-prediction, an encoded current image may be used as a reference image for another image that is processed afterwards. Accordingly, the encoding apparatus **100** may reconstruct or decode the encoded current image, or store the reconstructed or decoded image as a reference image in reference picture buffer **190**.

[0133] A quantized level may be dequantized in the dequantization unit **160**, or may be inverse-transformed in the inverse-transform unit **170**. A dequantized or inverse-transformed coefficient or both may be added with a prediction block by the adder **175**. By adding the dequantized or inverse-transformed coefficient or both with the prediction block, a reconstructed block may be generated. Herein, the dequantized or inverse-transformed coefficient or both may mean a coefficient on which at least one of dequantization and inverse-transform is performed, and may mean a reconstructed residual block.

[0134] A reconstructed block may pass through the filter unit **180**. The filter unit **180** may apply at least one of a deblocking filter, a sample adaptive offset (SAG), and an adaptive loop filter (A L F) to a reconstructed sample, a reconstructed block or a reconstructed image. The filter unit may be called as an in-loop filter.

[0135] The deblocking filter may remove block distortion generated in boundaries between blocks. In order to determine whether or not to apply a deblocking filter, whether or not to apply a deblocking filter to a current block may be determined based samples included in several rows or columns which are included in the block. When a deblock-

ing filter is applied to a block, another filter may be applied according to a required deblocking filtering strength.

[0136] In order to compensate an encoding error, a proper offset value may be added to a sample value by using a sample adaptive offset. The sample adaptive offset may correct an offset of a deblocked image from an original image by a sample unit. A method of partitioning samples of an image into a predetermined number of regions, determining a region to which an offset is applied, and applying the offset to the determined region, or a method of applying an offset in consideration of edge information on each sample may be used.

[0137] The adaptive loop filter may perform filtering based on a comparison result of the filtered reconstructed image and the original image. Samples included in an image may be partitioned into predetermined groups, a filter to be applied to each group may be determined, and differential filtering may be performed for each group. Information of whether or not to apply the A L F may be signaled by coding units (CUs), and a form and coefficient of the A L F to be applied to each block may vary.

[0138] The reconstructed block or the reconstructed image having passed through the filter unit may be stored in the reference picture buffer **190**. A reconstructed block processed by the filter unit **180** may be a part of a reference image. That is, a reference image is a reconstructed image composed of reconstructed blocks processed by the filter unit **180**. The stored reference image may be used later in inter prediction or motion compensation.

[0139] FIG. **2** is a block diagram showing a configuration of a decoding apparatus according to an embodiment and to which the present invention is applied.

[0140] A decoding apparatus **200** may a decoder, a video decoding apparatus, or an image decoding apparatus.

[0141] Referring to FIG. **2**, the decoding apparatus **200** may include an entropy decoding unit **210**, a dequantization unit **220**, a inverse-transform unit **230**, an intra-prediction unit **240**, a motion compensation unit **250**, an adder **225**, a filter unit **260**, and a reference picture buffer **270**.

[0142] The decoding apparatus **200** may receive a bitstream output from the encoding apparatus **100**. The decoding apparatus **200** may receive a bitstream stored in a computer readable recording medium, or may receive a bitstream that is streamed through a wired/wireless transmission medium. The decoding apparatus **200** may decode the bitstream by using an intra mode or an inter mode. In addition, the decoding apparatus **200** may generate a reconstructed image generated through decoding or a decoded image, and output the reconstructed image or decoded image.

[0143] When a prediction mode used when decoding is an intra mode, a switch may be switched to an intra. Alternatively, when a prediction mode used when decoding is an inter mode, a switch may be switched to an inter mode.

[0144] The decoding apparatus **200** may obtain a reconstructed residual block by decoding the input bitstream, and generate a prediction block. When the reconstructed residual block and the prediction block are obtained, the decoding apparatus **200** may generate a reconstructed block that becomes a decoding target by adding the reconstructed residual block with the prediction block. The decoding target block may be called a current block.

[0145] The entropy decoding unit **210** may generate symbols by entropy decoding the bitstream according to a probability distribution. The generated symbols may include a symbol of a quantized level form. Herein, an entropy decoding method may be a inverse-process of the entropy encoding method described above.

[0146] In order to decode a transform coefficient level (quantized level), the entropy decoding unit **210** may change a one-directional vector form coefficient into a two-dimensional block form by using a transform coefficient scanning method.

[0147] A quantized level may be dequantized in the dequantization unit **220**, or inverse-transformed in the inverse-transform unit **230**. The quantized level may be a result of dequantizing or inverse-transforming or both, and may be generated as a reconstructed residual block. Herein, the dequantization unit **220** may apply a quantization matrix to the quantized level.

[0148] When an intra mode is used, the intra-prediction unit **240** may generate a prediction block by performing, for the current block, spatial prediction that uses a sample value of a block adjacent to a decoding target block and which has been already decoded.

[0149] When an inter mode is used, the motion compensation unit **250** may generate a prediction block by performing, for the current block, motion compensation that uses a motion vector and a reference image stored in the reference picture buffer **270**.

[0150] The adder **225** may generate a reconstructed block by adding the reconstructed residual block with the prediction block. The filter unit **260** may apply at least one of a deblocking filter, a sample adaptive offset, and an adaptive loop filter to the reconstructed block or reconstructed image. The filter unit **260** may output the reconstructed image. The reconstructed block or reconstructed image may be stored in the reference picture buffer **270** and used when performing inter-prediction. A reconstructed block processed by the filter unit **260** may be a part of a reference image. That is, a reference image is a reconstructed image composed of reconstructed blocks processed by the filter unit **260**. The stored reference image may be used later in inter prediction or motion compensation.

[0151] FIG. **3** is a view schematically showing a partition structure of an image when encoding and decoding the image. FIG. **3** schematically shows an example of partitioning a single unit into a plurality of lower units.

[0152] In order to efficiently partition an image, when encoding and decoding, a coding unit (CU) may be used. The coding unit may be used as a basic unit when encoding/decoding the image. In addition, the coding unit may be used as a unit for distinguishing an intra prediction mode and an inter prediction mode when encoding/decoding the image. The coding unit may be a basic unit used for prediction, transform, quantization, inverse-transform, dequantization, or an encoding/decoding process of a transform coefficient.

[0153] Referring to FIG. **3**, an image **300** is sequentially partitioned in a largest coding unit (LCU), and a LCU unit is determined as a partition structure. Herein, the LCU may be used in the same meaning as a coding tree unit (CTU). A unit partitioning may mean partitioning a block associated with to the unit. In block partition information, information of a unit depth may be included. Depth information may represent a number of times or a degree or both in which a unit is partitioned. A single unit may be partitioned into a plurality of lower level units hierarchically associated with depth information based on a tree structure. In other words,

a unit and a lower level unit generated by partitioning the unit may correspond to a node and a child node of the node, respectively. Each of partitioned lower unit may have depth information. Depth information may be information representing a size of a CU, and may be stored in each CU. Unit depth represents times and/or degrees related to partitioning a unit. Therefore, partitioning information of a lower-level unit may comprise information on a size of the lower-level unit.

[0154] A partition structure may mean a distribution of a coding unit (CU) within an LCU **310**. Such a distribution may be determined according to whether or not to partition a single CU into a plurality (positive integer equal to or greater than 2 including 2, 4, 8, 16, etc.) of CU s. A horizontal size and a vertical size of the CU generated by partitioning may respectively be half of a horizontal size and a vertical size of the CU before partitioning, or may respectively have sizes smaller than a horizontal size and a vertical size before partitioning according to a number of times of partitioning. The CU may be recursively partitioned into a plurality of CUs. By the recursive partitioning, at least one among a height and a width of a CU after partitioning may decrease comparing with at least one among a height and a width of a CU before partitioning. Partitioning of the CU may be recursively performed until to a predefined depth or predefined size. For example, a depth of an LCU may be 0, and a depth of a smallest coding unit (SCU) may be a predefined maximum depth. Herein, the LCU may be a coding unit having a maximum coding unit size, and the SCU may be a coding unit having a minimum coding unit size as described above. Partitioning is started from the LCU **310**, a CU depth increases by 1 as a horizontal size or a vertical size or both of the CU decreases by partitioning. For example, for each depth, a CU which is not partitioned may have a size of 2N×2N. Also, in case of a CU which is partitioned, a CU with a size of 2N×2N may be partitioned into four CUs with a size of N×N. A size of N may decrease to half as a depth increase by 1.

[0155] In addition, information whether or not the CU is partitioned may be represented by using partition information of the CU. The partition information may be 1-bit information. All CUs, except for a SCU, may include partition information. For example, when a value of partition information is a first value, the CU may not be partitioned, when a value of partition information is a second value, the CU may be partitioned

[0156] Referring to FIG. **3**, an LCU having a depth 0 may be a 64×64 block. 0 may be a minimum depth. A SCU having a depth 3 may be an 8×8 block. 3 may be a maximum depth. A CU of a 32×32 block and a 16×16 block may be respectively represented as a depth 1 and a depth 2.

[0157] For example, when a single coding unit is partitioned into four coding units, a horizontal size and a vertical size of the four partitioned coding units may be a half size of a horizontal and vertical size of the CU before being partitioned. In one embodiment, when a coding unit having a 32×32 size is partitioned into four coding units, each of the four partitioned coding units may have a 16×16 size. When a single coding unit is partitioned into four coding units, it may be called that the coding unit may be partitioned into a quad-tree form.

[0158] For example, when one coding unit is partitioned into two sub-coding units, the horizontal or vertical size (width or height) of each of the two sub-coding units may be half the horizontal or vertical size of the original coding unit. For example, when a coding unit having a size of 32×32 is vertically partitioned into two sub-coding units, each of the two sub-coding units may have a size of 16×32. For example, when a coding unit having a size of 8×32 is horizontally partitioned into two sub-coding units, each of the two sub-coding units may have a size of 8×16. When one coding unit is partitioned into two sub-coding units, it can be said that the coding unit is binary-partitioned or is partitioned by a binary tree partition structure.

[0159] For example, when one coding unit is partitioned into three sub-coding units, the horizontal or vertical size of the coding unit can be partitioned with a ratio of 1:2:1, thereby producing three sub-coding units whose horizontal or vertical sizes are in a ratio of 1:2:1. For example, when a coding unit having a size of 16×32 is horizontally partitioned into three sub-coding units, the three sub-coding units may have sizes of 16×8, 16×16, and 16×8 respectively, in the order from the uppermost to the lowermost sub-coding unit. For example, when a coding unit having a size of 32×32 is vertically split into three sub-coding units, the three sub-coding units may have sizes of 8×32, 16×32, and 8×32, respectively in the order from the left to the right sub-coding unit. When one coding unit is partitioned into three sub-coding units, it can be said that the coding unit is ternary-partitioned or partitioned by a ternary tree partition structure.

[0160] In FIG. **3**, a coding tree unit (CTU) **320** is an example of a CTU to which a quad tree partition structure, a binary tree partition structure, and a ternary tree partition structure are all applied.

[0161] As described above, in order to partition the CTU, at least one of a quad tree partition structure, a binary tree partition structure, and a ternary tree partition structure may be applied. Various tree partition structures may be sequentially applied to the CTU, according to a predetermined priority order. For example, the quad tree partition structure may be preferentially applied to the CTU. A coding unit that cannot be partitioned any longer using a quad tree partition structure may correspond to a leaf node of a quad tree. A coding unit corresponding to a leaf node of a quad tree may serve as a root node of a binary and/or ternary tree partition structure. That is, a coding unit corresponding to a leaf node of a quad tree may be further partitioned by a binary tree partition structure or a ternary tree partition structure, or may not be further partitioned. Therefore, by preventing a coding block that results from binary tree partitioning or ternary tree partitioning of a coding unit corresponding to a leaf node of a quad tree from undergoing further quad tree partitioning, block partitioning and/or signaling of partition information can be effectively performed.

[0162] The fact that a coding unit corresponding to a node of a quad tree is partitioned may be signaled using quad partition information. The quad partition information having a first value (e.g., "1") may indicate that a current coding unit is partitioned by the quad tree partition structure. The quad partition information having a second value (e.g., "0") may indicate that a current coding unit is not partitioned by the quad tree partition structure. The quad partition information may be a flag having a predetermined length (e.g., one bit).

[0163] There may not be a priority between the binary tree partitioning and the ternary tree partitioning. That is, a coding unit corresponding to a leaf node of a quad tree may further undergo arbitrary partitioning among the binary tree

partitioning and the ternary tree partitioning. In addition, a coding unit generated through the binary tree partitioning or the ternary tree partitioning may undergo a further binary tree partitioning or a further ternary tree partitioning, or may not be further partitioned.

[0164] A tree structure in which there is no priority among the binary tree partitioning and the ternary tree partitioning is referred to as a multi-type tree structure. A coding unit corresponding to a leaf node of a quad tree may serve as a root node of a multi-type tree. Whether to partition a coding unit which corresponds to a node of a multi-type tree may be signaled using at least one of multi-type tree partition indication information, partition direction information, and partition tree information. For partitioning of a coding unit corresponding to a node of a multi-type tree, the multi-type tree partition indication information, the partition direction, and the partition tree information may be sequentially signaled.

[0165] The multi-type tree partition indication information having a first value (e.g., "1") may indicate that a current coding unit is to undergo a multi-type tree partitioning. The multi-type tree partition indication information having a second value (e.g., "0") may indicate that a current coding unit is not to undergo a multi-type tree partitioning.

[0166] When a coding unit corresponding to a node of a multi-type tree is further partitioned by a multi-type tree partition structure, the coding unit may include partition direction information. The partition direction information may indicate in which direction a current coding unit is to be partitioned for the multi-type tree partitioning. The partition direction information having a first value (e.g., "1") may indicate that a current coding unit is to be vertically partitioned. The partition direction information having a second value (e.g., "0") may indicate that a current coding unit is to be horizontally partitioned.

[0167] When a coding unit corresponding to a node of a multi-type tree is further partitioned by a multi-type tree partition structure, the current coding unit may include partition tree information. The partition tree information may indicate a tree partition structure which is to be used for partitioning of a node of a multi-type tree. The partition tree information having a first value (e.g., "1") may indicate that a current coding unit is to be partitioned by a binary tree partition structure. The partition tree information having a second value (e.g., "0") may indicate that a current coding unit is to be partitioned by a ternary tree partition structure.

[0168] The partition indication information, the partition tree information, and the partition direction information may each be a flag having a predetermined length (e.g., one bit).

[0169] At least any one of the quadtree partition indication information, the multi-type tree partition indication information, the partition direction information, and the partition tree information may be entropy encoded/decoded. For the entropy-encoding/decoding of those types of information, information on a neighboring coding unit adjacent to the current coding unit may be used. For example, there is a high probability that the partition type (the partitioned or non-partitioned, the partition tree, and/or the partition direction) of a left neighboring coding unit and/or an upper neighboring coding unit of a current coding unit is similar to that of the current coding unit. Therefore, context information for entropy encoding/decoding of the information on the current coding unit may be derived from the information on the neighboring coding units. The information on the neighboring coding units may include at least any one of quad partition information, multi-type tree partition indication information, partition direction information, and partition tree information.

[0170] As another example, among binary tree partitioning and ternary tree partitioning, binary tree partitioning may be preferentially performed. That is, a current coding unit may primarily undergo binary tree partitioning, and then a coding unit corresponding to a leaf node of a binary tree may be set as a root node for ternary tree partitioning. In this case, neither quad tree partitioning nor binary tree partitioning may not be performed on the coding unit corresponding to a node of a ternary tree.

[0171] A coding unit that cannot be partitioned by a quad tree partition structure, a binary tree partition structure, and/or a ternary tree partition structure becomes a basic unit for coding, prediction and/or transformation. That is, the coding unit cannot be further partitioned for prediction and/or transformation. Therefore, the partition structure information and the partition information used for partitioning a coding unit into prediction units and/or transformation units may not be present in a bit stream.

[0172] However, when the size of a coding unit (i.e., a basic unit for partitioning) is larger than the size of a maximum transformation block, the coding unit may be recursively partitioned until the size of the coding unit is reduced to be equal to or smaller than the size of the maximum transformation block. For example, when the size of a coding unit is 64×64 and when the size of a maximum transformation block is 32×32, the coding unit may be partitioned into four 32×32 blocks for transformation. For example, when the size of a coding unit is 32×64 and the size of a maximum transformation block is 32×32, the coding unit may be partitioned into two 32×32 blocks for the transformation. In this case, the partitioning of the coding unit for transformation is not signaled separately, and may be determined through comparison between the horizontal or vertical size of the coding unit and the horizontal or vertical size of the maximum transformation block. For example, when the horizontal size (width) of the coding unit is larger than the horizontal size (width) of the maximum transformation block, the coding unit may be vertically bisected. For example, when the vertical size (length) of the coding unit is larger than the vertical size (length) of the maximum transformation block, the coding unit may be horizontally bisected.

[0173] Information of the maximum and/or minimum size of the coding unit and information of the maximum and/or minimum size of the transformation block may be signaled or determined at an upper level of the coding unit. The upper level may be, for example, a sequence level, a picture level, a slice level, a tile group level, a tile level, or the like. For example, the minimum size of the coding unit may be determined to be 4×4. For example, the maximum size of the transformation block may be determined to be 64×64. For example, the minimum size of the transformation block may be determined to be 4×4.

[0174] Information of the minimum size (quad tree minimum size) of a coding unit corresponding to a leaf node of a quad tree and/or information of the maximum depth (the maximum tree depth of a multi-type tree) from a root node to a leaf node of the multi-type tree may be signaled or determined at an upper level of the coding unit. For example, the upper level may be a sequence level, a picture level, a

slice level, a tile group level, a tile level, or the like. Information of the minimum size of a quad tree and/or information of the maximum depth of a multi-type tree may be signaled or determined for each of an intra-picture slice and an inter-picture slice.

[0175] Difference information between the size of a CTU and the maximum size of a transformation block may be signaled or determined at an upper level of the coding unit. For example, the upper level may be a sequence level, a picture level, a slice level, a tile group level, a tile level, or the like. Information of the maximum size of the coding units corresponding to the respective nodes of a binary tree (hereinafter, referred to as a maximum size of a binary tree) may be determined based on the size of the coding tree unit and the difference information. The maximum size of the coding units corresponding to the respective nodes of a ternary tree (hereinafter, referred to as a maximum size of a ternary tree) may vary depending on the type of slice. For example, for an intra-picture slice, the maximum size of a ternary tree may be 32×32. For example, for an inter-picture slice, the maximum size of a ternary tree may be 128×128. For example, the minimum size of the coding units corresponding to the respective nodes of a binary tree (hereinafter, referred to as a minimum size of a binary tree) and/or the minimum size of the coding units corresponding to the respective nodes of a ternary tree (hereinafter, referred to as a minimum size of a ternary tree) may be set as the minimum size of a coding block.

[0176] As another example, the maximum size of a binary tree and/or the maximum size of a ternary tree may be signaled or determined at the slice level. Alternatively, the minimum size of the binary tree and/or the minimum size of the ternary tree may be signaled or determined at the slice level.

[0177] Depending on size and depth information of the above-described various blocks, quad partition information, multi-type tree partition indication information, partition tree information and/or partition direction information may be included or may not be included in a bit stream.

[0178] For example, when the size of the coding unit is not larger than the minimum size of a quad tree, the coding unit does not contain quad partition information. Thus, the quad partition information may be deduced from a second value.

[0179] For example, when the sizes (horizontal and vertical sizes) of a coding unit corresponding to a node of a multi-type tree are larger than the maximum sizes (horizontal and vertical sizes) of a binary tree and/or the maximum sizes (horizontal and vertical sizes) of a ternary tree, the coding unit may not be binary-partitioned or ternary-partitioned. Accordingly, the multi-type tree partition indication information may not be signaled but may be deduced from a second value.

[0180] Alternatively, when the sizes (horizontal and vertical sizes) of a coding unit corresponding to a node of a multi-type tree are the same as the maximum sizes (horizontal and vertical sizes) of a binary tree and/or are two times as large as the maximum sizes (horizontal and vertical sizes) of a ternary tree, the coding unit may not be further binary-partitioned or ternary-partitioned. Accordingly, the multi-type tree partition indication information may not be signaled but be derived from a second value. This is because when a coding unit is partitioned by a binary tree partition structure and/or a ternary tree partition structure, a coding unit smaller than the minimum size of a binary tree and/or the minimum size of a ternary tree is generated.

[0181] Alternatively, the binary tree partitioning or the ternary tree partitioning may be limited on the basis of the size of a virtual pipeline data unit (hereinafter, a pipeline buffer size). For example, when the coding unit is divided into sub-coding units which do not fit the pipeline buffer size by the binary tree partitioning or the ternary tree partitioning, the corresponding binary tree partitioning or ternary tree partitioning may be limited. The pipeline buffer size may be the size of the maximum transform block (e.g., 64×64). For example, when the pipeline buffer size is 64×64, the division below may be limited.

[0182] N×M (N and/or M is 128) Ternary tree partitioning for coding units

[0183] 128×N (N<=64) Binary tree partitioning in horizontal direction for coding units

[0184] N×128 (N<=64) Binary tree partitioning in vertical direction for coding units

[0185] Alternatively, when the depth of a coding unit corresponding to a node of a multi-type tree is equal to the maximum depth of the multi-type tree, the coding unit may not be further binary-partitioned and/or ternary-partitioned. Accordingly, the multi-type tree partition indication information may not be signaled but may be deduced from a second value.

[0186] Alternatively, only when at least one of vertical direction binary tree partitioning, horizontal direction binary tree partitioning, vertical direction ternary tree partitioning, and horizontal direction ternary tree partitioning is possible for a coding unit corresponding to a node of a multi-type tree, the multi-type tree partition indication information may be signaled. Otherwise, the coding unit may not be binary-partitioned and/or ternary-partitioned. Accordingly, the multi-type tree partition indication information may not be signaled but may be deduced from a second value.

[0187] Alternatively, only when both of the vertical direction binary tree partitioning and the horizontal direction binary tree partitioning or both of the vertical direction ternary tree partitioning and the horizontal direction ternary tree partitioning are possible for a coding unit corresponding to a node of a multi-type tree, the partition direction information may be signaled. Otherwise, the partition direction information may not be signaled but may be derived from a value indicating possible partitioning directions.

[0188] Alternatively, only when both of the vertical direction binary tree partitioning and the vertical direction ternary tree partitioning or both of the horizontal direction binary tree partitioning and the horizontal direction ternary tree partitioning are possible for a coding tree corresponding to a node of a multi-type tree, the partition tree information may be signaled. Otherwise, the partition tree information may not be signaled but be deduced from a value indicating a possible partitioning tree structure.

[0189] FIG. **4** is a view showing an intra-prediction process.

[0190] Arrows from center to outside in FIG. **4** may represent prediction directions of intra prediction modes.

[0191] Intra encoding and/or decoding may be performed by using a reference sample of a neighbor block of the current block. A neighbor block may be a reconstructed neighbor block. For example, intra encoding and/or decod-

ing may be performed by using an encoding parameter or a value of a reference sample included in a reconstructed neighbor block.

[0192] A prediction block may mean a block generated by performing intra prediction. A prediction block may correspond to at least one among CU, PU and TU. A unit of a prediction block may have a size of one among CU, PU and TU. A prediction block may be a square block having a size of 2×2, 4×4, 16×16, 32×32 or 64×64 etc. or may be a rectangular block having a size of 2×8, 4×8, 2×16, 4×16 and 8×16 etc.

[0193] Intra prediction may be performed according to intra prediction mode for the current block. The number of intra prediction modes which the current block may have may be a fixed value and may be a value determined differently according to an attribute of a prediction block. For example, an attribute of a prediction block may comprise a size of a prediction block and a shape of a prediction block, etc.

[0194] The number of intra-prediction modes may be fixed to N regardless of a block size. Or, the number of intra prediction modes may be 3, 5, 9, 17, 34, 35, 36, 65, or 67 etc. Alternatively, the number of intra-prediction modes may vary according to a block size or a color component type or both. For example, the number of intra prediction modes may vary according to whether the color component is a luma signal or a chroma signal. For example, as a block size becomes large, a number of intra-prediction modes may increase. Alternatively, a number of intra-prediction modes of a luma component block may be larger than a number of intra-prediction modes of a chroma component block.

[0195] An intra-prediction mode may be a non-angular mode or an angular mode. The non-angular mode may be a DC mode or a planar mode, and the angular mode may be a prediction mode having a specific direction or angle. The intra-prediction mode may be expressed by at least one of a mode number, a mode value, a mode numeral, a mode angle, and mode direction. A number of intra-prediction modes may be M, which is larger than 1, including the non-angular and the angular mode. In order to intra-predict a current block, a step of determining whether or not samples included in a reconstructed neighbor block may be used as reference samples of the current block may be performed. When a sample that is not usable as a reference sample of the current block is present, a value obtained by duplicating or performing interpolation on at least one sample value among samples included in the reconstructed neighbor block or both may be used to replace with a non-usable sample value of a sample, thus the replaced sample value is used as a reference sample of the current block.

[0196] FIG. **7** is a diagram illustrating reference samples capable of being used for intra prediction.

[0197] As shown in FIG. **7**, at least one of the reference sample line 0 to the reference sample line may be used for intra prediction of the current block. In FIG. **7**, the samples of a segment A and a segment F may be padded with the samples closest to a segment B and a segment E, respectively, instead of retrieving from the reconstructed neighboring block. Index information indicating the reference sample line to be used for intra prediction of the current block may be signaled. When the upper boundary of the current block is the boundary of the CTU, only the reference sample line 0 may be available. Therefore, in this case, the index information may not be signaled. When a reference sample line other than the reference sample line 0 is used, filtering for a prediction block, which will be described later, may not be performed.

[0198] When intra-predicting, a filter may be applied to at least one of a reference sample and a prediction sample based on an intra-prediction mode and a current block size.

[0199] In case of a planar mode, when generating a prediction block of a current block, according to a position of a prediction target sample within a prediction block, a sample value of the prediction target sample may be generated by using a weighted sum of an upper and left side reference sample of a current sample, and a right upper side and left lower side reference sample of the current block. In addition, in case of a DC mode, when generating a prediction block of a current block, an average value of upper side and left side reference samples of the current block may be used. In addition, in case of an angular mode, a prediction block may be generated by using an upper side, a left side, a right upper side, and/or a left lower side reference sample of the current block. In order to generate a prediction sample value, interpolation of a real number unit may be performed.

[0200] In the case of intra prediction between color components, a prediction block for the current block of the second color component may be generated on the basis of the corresponding reconstructed block of the first color component. For example, the first color component may be a luma component, and the second color component may be a chroma component. For intra prediction between color components, the parameters of the linear model between the first color component and the second color component may be derived on the basis of the template. The template may include upper and/or left neighboring samples of the current block and upper and/or left neighboring samples of the reconstructed block of the first color component corresponding thereto. For example, the parameters of the linear model may be derived using a sample value of a first color component having a maximum value among samples in a template and a sample value of a second color component corresponding thereto, and a sample value of a first color component having a minimum value among samples in the template and a sample value of a second color component corresponding thereto. When the parameters of the linear model are derived, a corresponding reconstructed block may be applied to the linear model to generate a prediction block for the current block. According to a video format, subsampling may be performed on the neighboring samples of the reconstructed block of the first color component and the corresponding reconstructed block. For example, when one sample of the second color component corresponds to four samples of the first color component, four samples of the first color component may be sub-sampled to compute one corresponding sample. In this case, the parameter derivation of the linear model and intra prediction between color components may be performed on the basis of the corresponding sub-sampled samples. Whether or not to perform intra prediction between color components and/or the range of the template may be signaled as the intra prediction mode.

[0201] The current block may be partitioned into two or four sub-blocks in the horizontal or vertical direction. The partitioned sub-blocks may be sequentially reconstructed. That is, the intra prediction may be performed on the sub-block to generate the sub-prediction block. In addition, dequantization and/or inverse transform may be performed on the sub-blocks to generate sub-residual blocks. A recon-

structed sub-block may be generated by adding the sub-prediction block to the sub-residual block. The reconstructed sub-block may be used as a reference sample for intra prediction of the sub-sub-blocks. The sub-block may be a block including a predetermined number (for example, 16) or more samples. Accordingly, for example, when the current block is an 8×4 block or a 4×8 block, the current block may be partitioned into two sub-blocks. Also, when the current block is a 4×4 block, the current block may not be partitioned into sub-blocks. When the current block has other sizes, the current block may be partitioned into four sub-blocks. Information on whether or not to perform the intra prediction based on the sub-blocks and/or the partitioning direction (horizontal or vertical) may be signaled. The intra prediction based on the sub-blocks may be limited to be performed only when reference sample line 0 is used. When the intra prediction based on the sub-block is performed, filtering for the prediction block, which will be described later, may not be performed.

[0202] The final prediction block may be generated by performing filtering on the prediction block that is intra-predicted. The filtering may be performed by applying predetermined weights to the filtering target sample, the left reference sample, the upper reference sample, and/or the upper left reference sample. The weight and/or the reference sample (range, position, etc.) used for the filtering may be determined on the basis of at least one of a block size, an intra prediction mode, and a position of the filtering target sample in the prediction block. The filtering may be performed only in the case of a predetermined intra prediction mode (e.g., DC, planar, vertical, horizontal, diagonal, and/or adjacent diagonal modes). The adjacent diagonal mode may be a mode in which k is added to or subtracted from the diagonal mode. For example, k may be a positive integer of 8 or less.

[0203] An intra-prediction mode of a current block may be entropy encoded/decoded by predicting an intra-prediction mode of a block present adjacent to the current block. When intra-prediction modes of the current block and the neighbor block are identical, information that the intra-prediction modes of the current block and the neighbor block are identical may be signaled by using predetermined flag information. In addition, indicator information of an intra-prediction mode that is identical to the intra-prediction mode of the current block among intra-prediction modes of a plurality of neighbor blocks may be signaled. When intra-prediction modes of the current block and the neighbor block are different, intra-prediction mode information of the current block may be entropy encoded/decoded by performing entropy encoding/decoding based on the intra-prediction mode of the neighbor block.

[0204] FIG. **5** is a diagram illustrating an embodiment of an inter-picture prediction process.

[0205] In FIG. **5**, a rectangle may represent a picture. In FIG. **5**, an arrow represents a prediction direction. Pictures may be categorized into intra pictures (I pictures), predictive pictures (P pictures), and Bi-predictive pictures (B pictures) according to the encoding type thereof.

[0206] The I picture may be encoded through intra-prediction without requiring inter-picture prediction. The P picture may be encoded through inter-picture prediction by using a reference picture that is present in one direction (i.e., forward direction or backward direction) with respect to a current block. The B picture may be encoded through inter-picture prediction by using reference pictures that are preset in two directions (i.e., forward direction and backward direction) with respect to a current block. When the inter-picture prediction is used, the encoder may perform inter-picture prediction or motion compensation and the decoder may perform the corresponding motion compensation.

[0207] Hereinbelow, an embodiment of the inter-picture prediction will be described in detail.

[0208] The inter-picture prediction or motion compensation may be performed using a reference picture and motion information.

[0209] Motion information of a current block may be derived during inter-picture prediction by each of the encoding apparatus **100** and the decoding apparatus **200**. The motion information of the current block may be derived by using motion information of a reconstructed neighboring block, motion information of a collocated block (also referred to as a col block or a co-located block), and/or a block adjacent to the co-located block. The co-located block may mean a block that is located spatially at the same position as the current block, within a previously reconstructed collocated picture (also referred to as a col picture or a co-located picture). The co-located picture may be one picture among one or more reference pictures included in a reference picture list.

[0210] The derivation method of the motion information may be different depending on the prediction mode of the current block. For example, a prediction mode applied for inter prediction includes an AMVP mode, a merge mode, a skip mode, a merge mode with a motion vector difference, a subblock merge mode, a triangle partition mode, an inter-intra combination prediction mode, affine mode, and the like. Herein, the merge mode may be referred to as a motion merge mode.

[0211] For example, when the AMVP is used as the prediction mode, at least one of motion vectors of the reconstructed neighboring blocks, motion vectors of the co-located blocks, motion vectors of blocks adjacent to the co-located blocks, and a (0, 0) motion vector may be determined as motion vector candidates for the current block, and a motion vector candidate list is generated by using the emotion vector candidates. The motion vector candidate of the current block can be derived by using the generated motion vector candidate list. The motion information of the current block may be determined based on the derived motion vector candidate. The motion vectors of the collocated blocks or the motion vectors of the blocks adjacent to the collocated blocks may be referred to as temporal motion vector candidates, and the motion vectors of the reconstructed neighboring blocks may be referred to as spatial motion vector candidates.

[0212] The encoding apparatus **100** may calculate a motion vector difference (MVD) between the motion vector of the current block and the motion vector candidate and may perform entropy encoding on the motion vector difference (MVD). In addition, the encoding apparatus **100** may perform entropy encoding on a motion vector candidate index and generate a bitstream. The motion vector candidate index may indicate an optimum motion vector candidate among the motion vector candidates included in the motion vector candidate list. The decoding apparatus may perform entropy decoding on the motion vector candidate index included in the bitstream and may select a motion vector

candidate of a decoding target block from among the motion vector candidates included in the motion vector candidate list by using the entropy-decoded motion vector candidate index. In addition, the decoding apparatus **200** may add the entropy-decoded MVD and the motion vector candidate extracted through the entropy decoding, thereby deriving the motion vector of the decoding target block.

[0213] Meanwhile, the coding apparatus **100** may perform entropy-coding on resolution information of the calculated MVD. The decoding apparatus **200** may adjust the resolution of the entropy-decoded MVD using the MVD resolution information.

[0214] Meanwhile, the coding apparatus **100** calculates a motion vector difference (M VD) between a motion vector and a motion vector candidate in the current block on the basis of an affine model, and performs entropy-coding on the MVD. The decoding apparatus **200** derives a motion vector on a per sub-block basis by deriving an affine control motion vector of a decoding target block through the sum of the entropy-decoded MVD and an affine control motion vector candidate.

[0215] The bitstream may include a reference picture index indicating a reference picture. The reference picture index may be entropy-encoded by the encoding apparatus **100** and then signaled as a bitstream to the decoding apparatus **200**. The decoding apparatus **200** may generate a prediction block of the decoding target block based on the derived motion vector and the reference picture index information.

[0216] Another example of the method of deriving the motion information of the current may be the merge mode. The merge mode may mean a method of merging motion of a plurality of blocks. The merge mode may mean a mode of deriving the motion information of the current block from the motion information of the neighboring blocks. When the merge mode is applied, the merge candidate list may be generated using the motion information of the reconstructed neighboring blocks and/or the motion information of the collocated blocks. The motion information may include at least one of a motion vector, a reference picture index, and an inter-picture prediction indicator. The prediction indicator may indicate one-direction prediction (L0 prediction or L1 prediction) or two-direction predictions (L0 prediction and L1 prediction).

[0217] The merge candidate list may be a list of motion information stored. The motion information included in the merge candidate list may be at least one of motion information (spatial merge candidate) of a neighboring block adjacent to the current block, motion information (temporal merge candidate) of the collocated block of the current block in the reference picture, new motion information generated by a combination of the motion information exiting in the merge candidate list, motion information (history-based merge candidate) of the block that is encoded/decoded before the current block, and zero merge candidate.

[0218] The encoding apparatus **100** may generate a bitstream by performing entropy encoding on at least one of a merge flag and a merge index and may signal the bitstream to the decoding apparatus **200**. The merge flag may be information indicating whether or not to perform the merge mode for each block, and the merge index may be information indicating that which neighboring block, among the neighboring blocks of the current block, is a merge target block. For example, the neighboring blocks of the current block may include a left neighboring block on the left side of the current block, an upper neighboring block disposed above the current block, and a tempor al neighboring block temporally adjacent to the current block.

[0219] Meanwhile, the coding apparatus **100** performs entropy-coding on the correction information for correcting the motion vector among the motion information of the merge candidate and signals the same to the decoding apparatus **200**. The decoding apparatus **200** can correct the motion vector of the merge candidate selected by the merge index on the basis of the correction information. H ere, the correction information may include at least one of information on whether or not to perform the correction, correction direction information, and correction size information. As described above, the prediction mode that corrects the motion vector of the merge candidate on the basis of the signaled correction information may be referred to as a merge mode having the motion vector difference.

[0220] The skip mode may be a mode in which the motion information of the neighboring block is applied to the current block as it is. When the skip mode is applied, the encoding apparatus **100** may perform entropy encoding on information of the fact that the motion information of which block is to be used as the motion information of the current block to generate a bitstream, and may signal the bitstream to the decoding apparatus **200**. The encoding apparatus **100** may not signal a syntax element regarding at least any one of the motion vector difference information, the encoding block flag, and the transform coefficient level to the decoding apparatus **200**.

[0221] The subblock merge mode may mean a mode that derives the motion information in units of sub-blocks of a coding block (CU). When the subblock merge mode is applied, a subblock merge candidate list may be generated using motion information (sub-block based temporal merge candidate) of the sub-block collocated to the current sub-block in the reference image and/or an affine control point motion vector merge candidate.

[0222] The triangle partition mode may mean a mode that derives motion information by partitioning the current block into diagonal directions, derives each prediction sample using each of the derived motion information, and derives the prediction sample of the current block by weighting each of the derived prediction samples.

[0223] The inter-intra combined prediction mode may mean a mode that derives a prediction sample of the current block by weighting a prediction sample generated by inter prediction and a prediction sample generated by intra prediction.

[0224] The decoding apparatus **200** may correct the derived motion information by itself. The decoding apparatus **200** may search the predetermined region on the basis of the reference block indicated by the derived motion information and derive the motion information having the minimum SAD as the corrected motion information.

[0225] The decoding apparatus **200** may compensate a prediction sample derived via inter prediction using an optical flow.

[0226] FIG. **6** is a diagram illustrating a transform and quantization process.

[0227] As illustrated in FIG. **6**, a transform and/or quantization process is performed on a residual signal to generate a quantized level signal. The residual signal is a difference between an original block and a prediction block (i.e., an

intra prediction block or an inter prediction block). The prediction block is a block generated through intra prediction or inter prediction. The transform may be a primary transform, a secondary transform, or both. The primary transform of the residual signal results in transform coefficients, and the secondary transform of the transform coefficients results in secondary transform coefficients.

[0228] At least one scheme selected from among various transform schemes which are preliminarily defined is used to perform the primary transform. For example, examples of the predefined transform schemes include discrete cosine transform (DCT), discrete sine transform (DST), and Karhunen-Loeve transform (K LT). The transform coefficients generated through the primary transform may undergo the secondary transform. The transform schemes used for the primary transform and/or the secondary transform may be determined according to coding parameters of the current block and/or neighboring blocks of the current block. Alternatively, transform information indicating the transform scheme may be signaled. The DCT-based transform may include, for example, DCT-2, DCT-8, and the like. The DST-based transform may include, for example, DST-7.

[0229] A quantized-level signal (quantization coefficients) may be generated by performing quantization on the residual signal or a result of performing the primary transform and/or the secondary transform. The quantized level signal may be scanned according to at least one of a diagonal up-right scan, a vertical scan, and a horizontal scan, depending on an intra prediction mode of a block or a block size/shape. For example, as the coefficients are scanned in a diagonal up-right scan, the coefficients in a block form change into a one-dimensional vector form. Aside from the diagonal up-right scan, the horizontal scan of horizontally scanning a two-dimensional block form of coefficients or the vertical scan of vertically scanning a two-dimensional block form of coefficients may be used depending on the intra prediction mode and/or the size of a transform block. The scanned quantized-level coefficients may be entropy-encoded to be inserted into a bitstream.

[0230] A decoder entropy-decodes the bitstream to obtain the quantized-level coefficients. The quantized-level coefficients may be arranged in a two-dimensional block form through inverse scanning. For the inverse scanning, at least one of a diagonal up-right scan, a vertical scan, and a horizontal scan may be used.

[0231] The quantized-level coefficients may then be dequantized, then be secondary-inverse-transformed as necessary, and finally be primary-inverse-transformed as necessary to generate a reconstructed residual signal.

[0232] Inverse mapping in a dynamic range may be performed for a luma component reconstructed through intra prediction or inter prediction before in-loop filtering. The dynamic range may be divided into 16 equal pieces and the mapping function for each piece may be signaled. The mapping function may be signaled at a slice level or a tile group level. An inverse mapping function for performing the inverse mapping may be derived on the basis of the mapping function. In-loop filtering, reference picture storage, and motion compensation are performed in an inverse mapped region, and a prediction block generated through inter prediction is converted into a mapped region via mapping using the mapping function, and then used for generating the reconstructed block. However, since the intra prediction is performed in the mapped region, the prediction block generated via the intra prediction may be used for generating the reconstructed block without mapping/inverse mapping.

[0233] When the current block is a residual block of a chroma component, the residual block may be converted into an inverse mapped region by performing scaling on the chroma component of the mapped region. The availability of the scaling may be signaled at the slice level or the tile group level. The scaling may be applied only when the mapping for the luma component is available and the division of the luma component and the division of the chroma component follow the same tree structure. The scaling may be performed on the basis of an average of sample values of a luma prediction block corresponding to the color difference block. In this case, when the current block uses inter prediction, the luma prediction block may mean a mapped luma prediction block. A value necessary for the scaling may be derived by referring to a lookup table using an index of a piece to which an average of sample values of a luma prediction block belongs. Finally, by scaling the residual block using the derived value, the residual block may be switched to the inverse mapped region. Then, chroma component block restoration, intra prediction, inter prediction, in-loop filtering, and reference picture storage may be performed in the inverse mapped area.

[0234] Information indicating whether the mapping/inverse mapping of the luma component and chroma component is available may be signaled through a set of sequence parameters.

[0235] The prediction block of the current block may be generated on the basis of a block vector indicating a displacement between the current block and the reference block in the current picture. In this way, a prediction mode for generating a prediction block with reference to the current picture is referred to as an intra block copy (IBC) mode. The IBC mode may be applied to M×N (M<=64, N<=64) coding units. The IBC mode may include a skip mode, a merge mode, an A M V P mode, and the like. In the case of a skip mode or a merge mode, a merge candidate list is constructed, and the merge index is signaled so that one merge candidate may be specified. The block vector of the specified merge candidate may be used as a block vector of the current block. The merge candidate list may include at least one of a spatial candidate, a history-based candidate, a candidate based on an average of two candidates, and a zero-merge candidate. In the case of an AMVP mode, the difference block vector may be signaled. In addition, the prediction block vector may be derived from the left neighboring block and the upper neighboring block of the current block. The index on which neighboring block to use may be signaled. The prediction block in the IBC mode is included in the current CTU or the left CTU and limited to a block in the already reconstructed area. For example, a value of the block vector may be limited such that the prediction block of the current block is positioned in an area of three 64×64 blocks preceding the 64×64 block to which the current block belongs in the coding/decoding order. By limiting the value of the block vector in this way, memory consumption and device complexity according to the BC mode implementation may be reduced.

[0236] Hereinafter, a transform coefficient encoding/decoding method according to an embodiment of the present invention will be described. In the embodiment below, a current block may be a transform unit or a transform block.

[0237] In the present invention, the transform coefficient for the current block may mean a coefficient derived as a result of transforming a residual signal for the current block, or a quantized level value of the transform coefficient.

[0238] FIG. **8** is a diagram illustrating a transform coefficient encoding/decoding method according to an embodiment of the present invention.

[0239] Referring to FIG. **8**, the transform coefficient encoding/decoding method may include: [D1] a step of grouping transform coefficients; [D2] a step of scanning the transform coefficients; and [D3] a step of binarizing the transform coefficients.

[D1] A Transform Coefficient Grouping Step

[0240] In encoding/decoding of the transform coefficient for the current block, the transform coefficient grouping may be performed using at least one method among fixed block transform coefficient grouping and variable block transform coefficient grouping.

[0241] The encoding/decoding of the transform coefficient for the current block may be performed on a per-transform coefficient group basis.

[0242] K transform coefficient groups may be independently subjected to transform coefficient encoding/decoding. Further, the K transform coefficient groups may be independently subjected to prediction encoding/decoding, and may be independently subjected to transform and quantization. Herein, K may be a positive integer.

[0243] In the meantime, the transform coefficient group may be defined as a transform coefficient block, a child transform block, or a sub transform block.

[0244] Transform coefficient grouping for the current block according to the embodiment of the present invention may be performed as the fixed block transform coefficient grouping.

[0245] Specifically, the transform coefficients for the current block may be divided into transform coefficient groups, each in a fixed M (width)×N (height) size.

[0246] FIGS. **9** to **12** are diagrams illustrating an example of fixed block transform coefficient grouping of the present invention.

[0247] FIG. **9** shows an example in which transform coefficient grouping is performed with a size that is the same as the size of the current block. Referring to FIG. **9**, when the current block is in an 8×8 size, the size of the transform coefficient group is set to 8×8 and the current block may be a transform coefficient group.

[0248] FIG. **10** shows an example in which the current block is partitioned with a predetermined size and is subjected to transform coefficient grouping. Referring to FIG. **10**, when the current block is in a 16×16 size and the transform coefficient group is in a predetermined size, which is a 4×4 size, the current block is divided into 16 transform coefficient groups for grouping.

[0249] FIG. **11** shows an example of transform coefficient grouping when the current block is not exactly divided by the transform coefficient group in size. Referring to FIG. **11**, when the current block is in a 16×16 size and the transform coefficient group is in a predetermined size, which is a 5×5 size, transform coefficient groups in subblock sizes, such as 5×1, 1×5, and 1×1, are allowed with respect to the right and bottom boundaries of the current block.

[0250] FIG. **12** shows an example of transform coefficient grouping when the horizontal or vertical size of a predetermined transform coefficient group is larger than the horizontal or vertical size of the current block. Referring to FIG. **12**, when the current block is in a 16×4 size and the predetermined transform coefficient group is in an 8×8 size, an 8×4-sized transform coefficient group is allowed. That is, the size of the predetermined transform coefficient group may vary in accordance with the size of the current block.

[0251] In the meantime, when the current block is present at the right or lower boundary of the picture or when the current block is not exactly divided by the transform coefficient group in size, a transform coefficient group which is in a size different from a fixed transform coefficient group size is allowed, as described above with reference to FIGS. **11** and **12**.

[0252] Transform coefficient grouping for the current block according to the embodiment of the present invention may be performed as variable block transform coefficient grouping.

[0253] Specifically, the transform coefficients for the current block may be divided into k transform coefficient groups in different sizes. In variable block transform coefficient grouping, the current block may be partitioned into k transform coefficient groups that have the position and the size of the transform coefficient group having the optimum cost. In the meantime, the k transform coefficient groups resulting from the variable block transform coefficient grouping may not overlap each other.

[0254] FIGS. **13** to **15** are diagrams illustrating an example of variable block transform coefficient grouping of the present invention.

[0255] As shown in FIG. **13**, when the current block is in a 16×16 size, the current block is divided in such a manner that the transform coefficient group at the top left position is in an 8×8 size and the remaining transform coefficient groups are in a 4×4 size.

[0256] As shown in FIG. **14**, when the current block is in a 16×16 size, the current block is divided in such a manner that the transform coefficient group at the left position is in a 4×8 size and the remaining transform coefficient groups are in a 4×4 size. Alternatively, when the current block is in a 16×16 size, the current block is divided in such a manner that the transform coefficient group at the top position is in an 8×4 size and the remaining transform coefficient groups are in a 4×4 size.

[0257] As shown in FIG. **15**, when the current block is in a 16×16 size, the current block is divided in such a manner that the transform coefficient group at the top left position is in an 8×8 size, the transform coefficient group at the left position is in a 4×8 size, the transform coefficient group at the top position is in an 8×4 size, and the remaining transform coefficient groups are in a 4×4 size.

[0258] In the meantime, the transform coefficient grouping method for the current block may be indexed into a grouping method having the optimum cost among N grouping methods, wherein N is any positive integer.

[0259] For example, an index list of five grouping methods may be constructed as follows, and index information indicating a method having the minimum R D-cost may be signaled.

[0260] index 0: a fixed block transform coefficient grouping method;

[0261] index 1: a first variable block transform coefficient grouping method;

[0262] index 2: a second variable block transform coefficient grouping method;

[0263] index 3: a third variable block transform coefficient grouping method;

[0264] index 4: a fourth variable block transform coefficient grouping method;

[0265] In performing transform coefficient grouping, the transform coefficient grouping method may be determined on the basis of an encoding parameter, for example, at least one among the slice type, the tile group type, the encoding mode, the intra prediction mode, the inter prediction mode, the value of the transform coefficient, and the size/shape of the current block. The transform coefficient grouping method may be at least one among the fixed transform coefficient grouping method and the variable transform coefficient grouping method described above.

[0266] On the basis of the slice type, the transform coefficient grouping method may be determined.

[0267] For example, when the current slice is slice **1**, the variable transform coefficient grouping method is determined. When the current slice is slice P or slice B, the fixed transform coefficient grouping method is determined.

[0268] On the basis of the position of the non-zero transform coefficient within the current block, the transform coefficient grouping method may be determined.

[0269] For example, one transform coefficient group in the minimum size which includes all the non-zero transform coefficients within the current block may be determined. Herein, the transform coefficient group may be in a size of (X_max+1)×(Y_max+1), wherein X_max is the maximum value for the X-coordinate positions of the non-zero transform coefficients within the current block and Y_max is the maximum value for the Y-coordinate positions.

[0270] FIG. **16** is a diagram illustrating a method of determining a transform coefficient group on the basis of a position of a non-zero transform coefficient within a current block.

[0271] Referring to FIG. **16**, in a 16×16-sized coding block, one transform coefficient group 0 in the minimum size (6×6) which includes all the non-zero transform coefficients may be determined.

[0272] The example of determining the transform coefficient grouping method on the basis of the position of the non-zero coefficient within the current block may be performed only, in a case where the current block is in an N×M size, when min(N, M) or max(N, M) is larger than a threshold value K, which is any positive integer.

[0273] On the basis of the prediction mode, the transform coefficient grouping method may be determined.

[0274] For example, when the current block is in the intra prediction mode, the variable transform coefficient grouping method is determined. When the current block is in the inter prediction mode, the fixed transform coefficient grouping method is determined.

[0275] On the basis of the intra prediction mode, the transform coefficient grouping method may be determined.

[0276] For example, when the intra prediction mode of the current block is the DC/Planar mode, it is determined that the variable coefficient grouping method described with reference to FIG. **13** is used. When the intra prediction mode of the current block is a horizontal angular mode, it is determined that the variable coefficient grouping method described with reference to FIG. **14** is used. When the intra prediction mode of the current block is a vertical angular mode, it is determined that the transform coefficient grouping method described with reference to FIG. **15** is used.

[0277] On the basis of the size of the current block, the transform coefficient grouping method may be determined.

[0278] For example, when the minimum value (min(N, M)) or the maximum value (max(N, M)) of the width (M) and the height (N) of the current block (M×N) is larger than the threshold value K, which is any positive integer, the variable transform coefficient grouping method is determined. Otherwise, the fixed transform coefficient grouping method is determined.

[0279] For example, when the width (M) and the height (N) of the current block (M×N) are the same, the fixed transform coefficient grouping method is determined. When the width (M) and the height (N) of the current block (M×N) differ, the variable transform coefficient grouping method is determined.

[0280] For example, when the width (M) of the current block (M×N) is smaller than the height (N) and is also smaller than a predefined value (P), the width of the transform coefficient group is fixed to the width (M) of the current block and then transform coefficient grouping is performed. Herein, P may be a positive integer, for example, 4.

[0281] For example, when the height (N) of the current block (M×N) is smaller than the width (M) and is also smaller than a predefined value (Q), the height of the transform coefficient group is fixed to the height (N) of the current block and then transform coefficient grouping is performed. Herein, Q may be a positive integer, for example, 4.

[0282] FIGS. **17** to **20** are diagrams illustrating how to determine a transform coefficient grouping method on the basis of a width (M) or a height (N) of a current block (M×N).

[0283] For example, when the width (M) of the current block is larger than any threshold value, transform coefficient grouping is performed with division of M by 4 as shown in FIG. **17**.

[0284] For example, when the width (M) of the current block is larger than any threshold value, transform coefficient grouping is performed with division of M by 2 as shown in FIG. **18**.

[0285] For example, when the height (N) of the current block is larger than any threshold value, transform coefficient grouping is performed with division of N by 4 as shown in FIG. **19**.

[0286] For example, when the height (N) of the current block is larger than any threshold value, transform coefficient grouping is performed with division of N by 2 as shown in FIG. **20**.

[0287] For example, when the width (M) or the height (N) of the current block is smaller than any threshold value (T), transform coefficient grouping is performed into a transform coefficient group in a B×B size. Conversely, when the width (M) and the height (N) of the current block are equal to or larger than the any threshold value, transform coefficient grouping is performed into a transform coefficient group in a 2B×2B size. Herein, T may be a positive integer, for example, 4, and B may be a positive integer, for example, 2.

[0288] For example, when the area of the current block (M×N) is larger than any threshold value (R) and is in a non-square shape, transform coefficient grouping is per-

formed into a transform coefficient group in a non-square shape. Herein, R may be a positive integer, for example, 8.

[0289] The area of the transform coefficient group may be fixed into at least one unit among a sub-CU, a CU, a CTU, a tile, a tile group, a brick, a slice, a picture, and a sequence. Herein, the area of the transform coefficient group may be the product of the width and the height of the transform coefficient group, and may be a positive integer.

[0290] For example, the area of the transform coefficient group may have a value of {4, 8, 16, 32, 64, 128, 256, 512}. Herein, the area of the transform coefficient group may be transmitted on the basis of at least one among a sub-CU, a CU, a CTU, a tile, a tile group, a brick, a slice, a picture, and a sequence.

[0291] For example, a value obtained by applying log **2** to the area of the transform coefficient group may be transmitted.

[0292] For example, a value obtained by applying log **2** to the area of the transform coefficient group is added to −2, and the resulting value may be transmitted.

[0293] For example, a value obtained by applying log **2** to the area of the transform coefficient group is added to −3, and the resulting value may be transmitted.

[0294] The transform coefficient group may be partitioned into sub transform coefficient groups.

[0295] For example, transform and quantization may be performed on a per transform coefficient group basis, and scanning and binarization of transform coefficients may be performed on a per sub transform coefficient group basis.

[0296] Using the intra prediction mode and the position information of the last non-zero coefficient, the transform coefficient grouping method may be determined.

[0297] A residual signal after intra prediction may have a correlation according to the intra prediction direction. Therefore, there is a characteristic that when horizontal direction prediction is used, the most non-zero coefficients after frequency transform occur in the first column; and conversely, when vertical direction prediction is used, the most non-zero coefficients after frequency transform occur in the first row.

[0298] Using such a characteristic and the position information of the last non-zero coefficient after frequency transform, transform coefficient grouping may be effectively performed. Herein, the wording effective may mean that a much larger transform coefficient group is generated to reduce the number of coefficient presence indicators based on a transform coefficient group (block) to be signaled.

[0299] FIGS. **21** to **23** are diagrams illustrating a method of setting a transform coefficient group by using an intra prediction mode and position information of the last non-zero coefficient.

[0300] FIG. **21** shows an example where when the intra prediction mode is the horizontal direction, transform coefficient grouping is performed using the position information of the last non-zero coefficient.

[0301] Referring to FIG. **21**, in intra horizontal direction prediction, the first transform coefficient group has the vertical size, starting from the position (0, 0), which may be the same as the vertical size of the current block; and the horizontal size of the first transform coefficient group may be determined to be the x position +1 of the last non-zero transform coefficient. Herein, intra prediction horizontal direction may include an intra prediction mode having a direction similar to the horizontal direction, namely, a direction at an angle of less than +−90 degrees, with the horizontal direction in the center.

[0302] FIG. **22** shows an example where when the intra prediction mode is the vertical direction, transform coefficient grouping is performed using position information of the last non-zero coefficient.

[0303] Referring to FIG. **22**, in intra vertical direction prediction, the first transform coefficient group has the horizontal size, starting from the position (0, 0), which is the same as the horizontal size of the current block; and the vertical size of the first transform coefficient group may be determined to be the y position +1 of the last non-zero transform coefficient. Herein, intra prediction vertical direction may include an intra prediction mode having a direction similar to the vertical direction, namely, a direction at an angle of less than +−90 degrees, with the vertical direction in the center.

[0304] FIG. **23** shows an example where the intra prediction mode is a non-angular prediction mode, transform coefficient grouping is performed using the position information of the last non-zero coefficient.

[0305] Referring to FIG. **23**, in non-angular intra prediction, the first transform coefficient group may be determined to be in a minimum-sized triangular shape including the position of the last non-zero coefficient. Herein, the non-angular prediction modes may include the Planar mode and the D C mode.

[0306] In the meantime, using the transform coefficient scanning method, instead of the intra prediction mode, and the position information of the last non-zero coefficient, the transform coefficient grouping method may be determined.

[0307] In the examples described with reference to FIGS. **21** and **22**, the transform coefficient group may be set according to the scanning method, without using the intra prediction mode.

[0308] For example, in the case of using horizontal scanning, the transform coefficient group may be set in the same manner as in FIG. **21**. In the case of using vertical scanning, the transform coefficient group may be set in the same manner as in FIG. **22**. Besides, in the case of using zigzag or diagonal scanning, the transform coefficient group may be set using the method as in FIG. **23**.

[0309] With respect to a remaining non-zero out region except for a zero out region, transform coefficient grouping may be performed.

[0310] Specifically, the remaining region except for the zero out region may be subjected to transform coefficient grouping. That is, transform coefficient grouping may be performed in the non-zero out region.

[0311] For example, in the non-zero out region, transform coefficient groups including all the non-zero transform coefficients may be generated.

[0312] In the meantime, the zero out region/non-zero out region may be predefined in the encoder and the decoder, or may be determined on the basis of signaled information in a particular header (a VPS, an SPS, a PPS, a slice, a brick, a tile group, or the like).

[0313] In the meantime, in the current block, a region having a predefined value (T) or larger may be set as a zero out region. Herein, T may be a positive integer, for example, 32. That is, when the width or height of the current block is equal to or larger than a predefined size, a region in a

predefined size or larger within the current block is determined as the zero out region.

[0314] For example, when the current block is in a 64×64 size and the predefined value is 32, the region other than a 32×32-sized region at the top left of the current block is set as the zero out region.

[0315] For example, when the current block is in a 16×64 size and the predefined value is 32, the region other than a 16×32-sized region at the top left of the current block is set as the zero out region.

[0316] Alternatively, the zero out region/non-zero out region may be determined on the basis of at least one among the size of the current block and the type of frequency transform.

[0317] For example, when as the type of frequency transform, the type of transform (for example, DST-7 or DCT-8) other than DCT-2 is used and the current block (or the transform block) is in a 64×64 size, the region except for the top left transform region in a 32×32 size is determined as the zero out region.

[0318] For example, when the type of transform (for example, DST-7 or DCT-8) other than DCT-2 is used as the type of frequency transform, the non-zero out region is set to be the M×M-sized region at the top left of the current block. In this case, the region other than the M×M-sized region at the top left may be set as the zero out region. Herein, M may be a positive integer, for example, 16.

[0319] In the meantime, the determination of the zero out region/non-zero out region may be performed by a combination of one or more examples described above.

[0320] FIGS. **24** and **25** are diagrams illustrating a zero out region and a non-zero out region according to an embodiment of the present invention.

[0321] FIG. **24** shows an example of a zero out region in a quadrangular shape.

[0322] FIG. **25** shows an example of a zero out region in a triangular shape.

[0323] Regarding the transform coefficient grouping forth non-zero out region, according to the zero out region, one or more transform coefficient groups in a fixed size may be configured, or transform coefficient groups in different sizes that may minimally include a zero out region may be configured.

[0324] FIGS. **26** and **27** are diagrams illustrating transform coefficient grouping in a non-zero out region.

[0325] Referring to FIG. **26**, when an 8×8-sized region at the bottom right within a 16×16-sized block is defined as the zero out region, grouping into three 8×8-sized transform coefficient groups takes place.

[0326] Referring to FIG. **27**, when the triangular region at the bottom right within the 16×16-sized block is defined as the zero out region and it is assumed that the minimum size of the transform coefficient group is 4×4, six 4×4-sized transform coefficient groups and one 8×8-sized transform coefficient group are configured to minimally include the zero out region.

[D2] A Transform Coefficient Scanning Step

[0327] In performing encoding/decoding of the transform coefficient for the current block, the transform coefficient scanning may be performed using at least one among fixed transform coefficient scanning and adaptive transform coefficient scanning. Further, the above-described scanning method may be performed on the transform coefficient group.

[0328] FIG. **28** is a diagram illustrating examples of a fixed transform coefficient scanning method.

[0329] Referring to FIG. **28**, the fixed transform coefficient scanning method may be performed by any one among diagonal scanning, horizontal scanning, vertical scanning, and zigzag scanning.

[0330] In the meantime, the fixed transform scanning method may be determined on the basis of at least one among a current block, a tile, a tile group, a brick, a slice, a picture, and a sequence.

[0331] Regarding the adaptive transform coefficient scanning, the transform coefficient scanning method may be adaptively determined on the basis of the encoding parameter.

[0332] For example, the transform coefficient scanning method may be determined on the basis of at least one among the quantization parameter, the transform method, the encoding mode, the intra prediction mode, the inter prediction mode, the value of the transform coefficient, and the size/shape of the current block.

[0333] As another example, when the current block is in the intra prediction mode and is processed in the DC/PLANAR prediction mode, the transform coefficient scanning method is determined to be diagonal scanning.

[0334] As still another example, when the current block is in the intra prediction mode and is processed in the horizontal angular prediction mode, the transform coefficient scanning method is determined to be vertical scanning.

[0335] As still another example, when the current block is in the intra prediction mode and is processed in the vertical angular prediction mode, the transform coefficient scanning method is determined to be horizontal scanning.

[0336] In the meantime, a transform coefficient scanning candidate list of N candidates may be configured on the basis of at least one among a picture, a slice, a brick, a tile group, a tile, a coding block, and a transform coefficient group, wherein N is a positive integer. In this case, the transform coefficient scanning candidate list may be configured on the basis of at least one among the encoding mode, the transform method, the inter prediction mode, and the intra prediction mode. The encoder may encode index information indicating the scanning method having the optimum cost, for example, generated bits, and RD-cost, in the transform coefficient candidate list. The decoder may select the transform coefficient scanning method from the transform coefficient candidate list by using the index information.

[0337] In the meantime, the transform coefficient scanning candidate list may be constructed on the basis of the transform coefficient scanning method for the neighboring block of the current block.

[0338] In the meantime, the transform coefficient scanning may be performed on the transform coefficient and the transform coefficient group that are included in the non-zero out region except for the zero out region.

[0339] The syntax elements (last_significant_coeff_x_prefix/suffix, last_significant_coeff_y_prefix/suffix), which mean the start of the scanning or the position of the last non-zero transform coefficient within the current transform block, are intended to represent only the position of the transform coefficient included in the non-zero out region, so

that the maximum range of the syntax elements may be limited, causing the syntax elements to be represented by much shorter bits.

[0340] For example, in the case where only a 32×32-sized region at the top left within a 64×64-sized transform block is defined as the non-zero out region, each of the positions x and y of the last transform coefficient may have a range of 0 to 31. Therefore, when represented by fixed bits, the position is represented by 5 bits.

[0341] The encoder may scan the transform coefficient groups included in the non-zero out region according to a predefined transform coefficient scanning method (or scanning order), and may scan the transform coefficients within the corresponding transform coefficient groups according to the predefined transform coefficient scanning method (or scanning order).

[0342] Herein, the transform coefficients may be encoded by starting the scanning from the transform coefficient that is present at the position of the last transform coefficient included in the non-zero out region.

[0343] The decoder may decode the position of the last transform coefficient from the bitstream and may perform scanning on the non-zero out region starting from the position of the last transform coefficient according to the predefined transform coefficient scanning method (or scanning order) so that the positions of the decoded transform coefficients within the transform coefficient block may be derived.

[D3] A Transform Coefficient Binarization Step

[0344] Binarization of the transform coefficient may be performed using at least one method among transform coefficient presence indicator binarization, transform coefficient sign binarization, and transform coefficient absolute value binarization. Herein, binarization may mean entropy encoding/decoding.

[0345] Hereinafter, the binarization of the transform coefficient presence indicator, of the transform coefficient sign, and of the transform coefficient absolute value will be described.

[D3-1] Binarization of a Transform Coefficient Presence Indicator

[0346] A transform coefficient presence indicator is an indicator indicating whether or not the non-zero transform coefficient is present.

[0347] The transform coefficient presence indicator may be encoded/decoded on the basis of the transform coefficient or the transform coefficient group.

[0348] The transform coefficient presence indicator based on the transform coefficient is set to "1" on the basis of the transform coefficient when the value of the transform coefficient is not 0, and is set to "0" when the value of the transform coefficient is 0.

[0349] The transform coefficient presence indicator based on the transform coefficient group is set to "1" when the non-zero transform coefficient is present in the corresponding transform coefficient group, and is set to "0" when the non-zero transform coefficient is not present.

[0350] When the transform coefficient presence indicator based on the transform coefficient group is "0", the coefficient presence indicator based on the transform coefficient within the transform coefficient group is not encoded/decoded. In this case, the transform coefficient presence indicator based on the transform coefficient within the transform coefficient group may be regarded to be "0".

[0351] FIG. **29** is a diagram illustrating a transform coefficient presence indicator based on a transform coefficient and a transform coefficient presence indicator based on a transform coefficient group.

[0352] In FIG. **29**, the current block is in an 8×8 size, and the transform coefficient group is in a 4×4 size.

[0353] Referring to FIG. **29**, when the transform coefficient presence indicator based on the transform coefficient group is "0", the transform coefficient indicator based on the transform coefficient is not encoded/decoded and thus does not present.

[0354] In the meantime, transform may not be performed in the zero out region, so that the transform coefficient presence indicator based on the transform coefficient group and the transform coefficient presence indicator based on the transform coefficient may be concealed with respect to the transform coefficient groups and the transform coefficients included in the zero out region. Herein, the concealment may mean that encoding/decoding is not performed or transmission from the encoder to the decoder does not take place.

[0355] Accordingly, the decoder may not entropy decode pieces of syntax of the transform coefficient presence indicator based on the transform coefficient group and the transform coefficient presence indicator based on the transform coefficient with respect to the transform coefficient groups and the transform coefficients that are included in the zero out region, and the value of the transform coefficient presence indicator based on the transform coefficient group and the value of the transform coefficient presence indicator based on the transform coefficient may be regarded to be "0".

[0356] A s described above, the zero out region may be predefined in the encoder and the decoder, or may be determined on the basis of signaled information in a particular header (a VPS, an SPS, a PPS, a slice, a brick, a tile group, or the like). Therefore, the transform coefficient presence indicator based on the transform coefficient group and the transform coefficient presence indicator based on the transform coefficient may be concealed on the basis of information for determining the zero out region.

[0357] Further, the zero out region may be determined on the basis of at least one among the size of the current block and the type of frequency transform, so that the transform coefficient presence indicator based on the transform coefficient group and the transform coefficient presence indicator based on the transform coefficient may be concealed on the basis of at least one among the size of the current block and the type of frequency transform.

[0358] For example, as shown in FIG. **26**, when the current block is in a 16×16 size and the grey 8×8-sized region is determined as the zero out region, the encoder does not transmit at least one transform coefficient presence indicator based on the transform group and at least one transform coefficient presence indicator based on the transform coefficient, with respect to the transform coefficient groups included within the grey region. Further, the decoder may not entropy decode at least one transform coefficient presence indicator based on the transform group and at least one transform coefficient presence indicator based on the transform coefficient with respect to the corresponding

region, and the values of the corresponding transform coefficient presence indicators based on the transform block coefficient and the values of the corresponding transform coefficient presence indicators based on the transform coefficient may be regarded to be "0".

[0359] For example, as shown in FIG. **27**, when the zero out region is in a triangular shape and the minimum transform coefficient group is defined to be in a 4×4 size, at least one transform coefficient presence indicator based on the transform coefficient for the zero out region included in the transform coefficient group is not transmitted by the encoder and is regarded, by the decoder, to have the value of 0. All other transform coefficient presence indicators based on the transform group of the zero out region and transform coefficient presence indicators based on the transform coefficient may be concealed.

[0360] For example, when as the type of transform, the type of transform (for example, DST7 or DCT8) other than DCT2 is used and the current block is in a 64×64 size, the region except for the top left transform region in a 32×32 size is defined as the zero out region. With respect to the transform coefficient groups included in the corresponding region, the encoder may not transmit at least one transform coefficient presence indicator based on the transform group and at least one transform coefficient presence indicator based on the transform coefficient, and the decoder may not entropy decode at least one transform coefficient presence indicator based on the transform group and at least one transform coefficient presence indicator based on the transform coefficient with respect to the corresponding region. Further, the values of the corresponding transform coefficient presence indicators based on the transform coefficient group and the values of the corresponding transform coefficient presence indicators based on the transform coefficient may be regarded to be "0".

[0361] When the transform coefficient presence indicator based on the transform coefficient group or the transform coefficient presence indicator based on the transform coefficient is subjected to binarization (entropy encoding/decoding), the transform coefficient groups uses different probability information (context) in performing entropy encoding/decoding of the transform coefficient presence indicator based on the transform coefficient group or the transform coefficient presence indicator based on the transform coefficient.

[0362] For example, in the case shown in FIG. **26**, the top left 8×8-sized block based on the transform coefficient and the other blocks use different probability information in performing entropy encoding/decoding of a presence indicator based on the transform coefficient block or a presence indicator based on the transform coefficient.

[0363] For example, in the case shown in FIG. **27**, the 4×4-sized transform coefficient group including a part of the zero out region and the 4×4-sized transform coefficient group not including the zero out region uses different probability information in performing entropy encoding/decoding of the transform coefficient presence indicator based on the transform coefficient group or the transform coefficient presence indicator based on the transform coefficient.

[0364] Further, on the basis of the size of each transform coefficient group, different probability information may be used in performing entropy encoding/decoding of the transform coefficient presence indicator based on the transform coefficient group or the transform coefficient presence indicator based on the transform coefficient.

[0365] [D3-2] A Transform Coefficient Sign Binarization

[0366] In performing transform coefficient sign binarization for the current block, binarization of the transform coefficient sign may be performed using at least one method among a transform coefficient sign indicator, transform coefficient sign concealment, and transform coefficient sign prediction.

[0367] In binarization of the transform coefficient sign for the current block, the transform coefficient sign indicator may be used.

[0368] For example, the transform coefficient sign indicator is set to "1" when the transform coefficient is a negative number (−), and is set to "0" when the transform coefficient is a positive number (+). Conversely, the transform coefficient sign indicator is set to "0" when the transform coefficient is a negative number (−), and is set to "1" when the transform coefficient is a positive number (+).

[0369] For example, when the transform coefficient presence indicator based on the transform coefficient or the transform coefficient presence indicator based on the transform coefficient group is "0", the coefficient sign indicator is not encoded/decoded.

[0370] In the meantime, in binarization of the transform coefficient sign for the current block, transform coefficient sign concealment may be used.

[0371] The transform coefficient sign concealment may be to derive N pieces of transform coefficient sign information according to the state of the transform coefficient absolute value, wherein N may be a positive integer, and may be performed on a per transform coefficient group basis.

[0372] For example, when the sum of the transform coefficients of the transform coefficient group is an even number, the first non-zero transform coefficient has a positive sign "+".

[0373] For example, when the sum of the transform coefficients of the transform coefficient group is an even number, the first non-zero transform coefficient has a negative sign "−".

[0374] In the meantime, in binarization of the transform coefficient sign for the current block, transform coefficient sign prediction may be used.

[0375] In the transform coefficient sign prediction, the signs of N non-zero transform coefficients in the transform coefficient block, wherein N is a positive integer, are predicted. The case where the actual sign matches the prediction sign is indicated by a value of "1", and the case where the actual sign does not match the prediction sign is indicated by a value of "0". The transform coefficient sign prediction may be performed on a per transform coefficient group basis.

[0376] For example, transform coefficient sign prediction may be performed with a combination having the most optimum cost among combinations (2N combinations) of N coefficient signs to be predicted. Herein, the cost of the sign combination may be obtained from the similarity between the data, obtained by inverse-transforming the current transform coefficient group with the corresponding combination, and the neighboring reconstructed image.

[D3-3] Transform Coefficient Absolute Value Binarization

[0377] In performing transform coefficient absolute value binarization for the current block, the transform coefficient

absolute value binarization may be performed using at least one method among transform coefficient binarization based on absolute value comparison, transform coefficient binarization based on exponential scale absolute value comparison, and the residual transform coefficient absolute value binarization.

[D3-3-1] Transform Coefficient Binarization Based on Absolute Value Comparison

[0378] The transform coefficient binarization based on absolute value comparison may be performed on a per transform coefficient group basis or a per transform coefficient basis, or may be performed when the transform coefficient presence indicator is 1.

[0379] The transform coefficient binarization based on absolute value comparison may be performed using Expression 1 and a positive integer N as follows.

$$\text{Abs(coeff.)} > K, K = \{1, \ldots, N\}[\quad \text{Expression 1]}$$

[0380] For example, when N is 2, the transform coefficient binarization based on absolute value comparison is performed as shown in FIG. **30**.

[0381] For example, K may be 2N or 2N-1. Herein, when K is 2N, the transform coefficient absolute value binarization is performed by comparing the absolute value of the transform coefficient with the value of 2, 4, 6, . . . , and 2N. In the meantime, when K is 2N-1, the transform coefficient absolute value binarization is performed by comparing the absolute value of the transform coefficient with the value of 1, 3, 5, . . . , and 2N-1.

[0382] In the meantime, according to the above-described scanning method, only I transform coefficients may be subjected to the transform coefficient binarization based on absolute value comparison. Herein, I may be a positive integer.

[D3-3-2] Transform Coefficient Binarization Based on Exponential Scale Absolute Value Comparison

[0383] The transform coefficient binarization based on exponential scale absolute value comparison may be performed on a per transform coefficient group basis or a per transform coefficient basis, or may be performed when the coefficient presence indicator is 1.

[0384] The transform coefficient binarization based on exponential scale absolute value comparison may be performed using Expression 2 and positive integers a and N as follows.

$$\text{Abs(coeff.)} \geq a^K, K = \{1, \ldots, N\}[\quad \text{Expression 2]}$$

[0385] For example, when a is 2 and N is 3, the transform coefficient binarization based on exponential scale comparison is performed as shown in FIG. **31**.

[0386] In the meantime, according to the above-described scanning method, only I transform coefficients may be subjected to the transform coefficient binarization based on absolute value comparison. Herein, I may be a positive integer.

[0387] In FIG. **31**, the value of abs(coeff.) may be the absolute value of the transform coefficient level. Herein, the transform coefficient level may be the result of transform and quantization.

[0388] As shown in FIG. **31**, when a is 2 and N is 3, the transform coefficient presence indicator, an indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 2, an indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 4, and an indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 8 are encoded/decoded.

[0389] In this case, when the transform coefficient presence indicator has a value of true (or a value of "1"), the indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 2 is encoded/decoded. Further, when the indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 2 has a value of true (or a value of "1"), the indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 4 is encoded/decoded. Further, when the indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 4 has a value of true (or a value of "1"), the indicator indicating the case where the absolute value of the transform coefficient level is equal to or larger than 8 is encoded/decoded.

[D3-3-3] Residual Transform Coefficient Binarization

[0390] The residual transform coefficient binarization may be performed on a per transform coefficient group basis or a per transform coefficient basis, or may be performed when the coefficient presence indicator is 1.

[0391] The residual transform coefficient binarization may be performed by various binarization methods.

[0392] For example, the residual transform coefficient binarization may be performed using a truncated rice binarization method.

[0393] For example, the residual transform coefficient binarization may be performed using a K-th order Exp_Golomb binarization method.

[0394] For example, the residual transform coefficient binarization may be performed using a limited binarization method.

[0395] For example, the residual transform coefficient binarization may be performed using a unary binarization method.

[0396] For example, the residual transform coefficient binarization may be performed using a truncated unary or truncated binarization method.

[0397] In the meantime, the binarization method may be determined according to the value of the residual transform coefficient.

[0398] For example, the residual transform coefficient which is equal to or smaller than a value of a positive integer c is subjected to binarization using the truncated unary binarization method, and the residual coefficient which exceeds the value of c is subjected to binarization using the K-th order Exp_Golomb binarization method.

[0399] In the meantime, the binarization method may be determined according to the scanning order.

[0400] For example, when the residual transform coefficient is subjected to binarization using the K-th order Exp_Golomb binarization method, the value of K is increased or decreased according to the scanning order.

[0401] On the basis of the encoding parameter (for example, at least one among the slice type, the tile group type, the encoding mode, the intra prediction mode, the inter prediction mode, the value of the transform coefficient, and

the size/shape of the current block), the transform coefficient binarization method may be determined on a per transform coefficient basis.

[0402] For example, the transform coefficients belonging to the low frequency may be subjected to exponential scale absolute value comparison binarization, and the remaining transform coefficients may be subjected to the transform coefficient binarization based on absolute value comparison.

[0403] For example, the transform coefficient having a low quantization parameter (Q P) may be subjected to the exponential scale absolute value comparison binarization, and the remaining transform coefficients may be subjected to the transform coefficient binarization based on absolute value comparison.

[0404] On the basis of the encoding parameter (for example, at least one among the slice type, the tile group type, the encoding mode, the intra prediction mode, the inter prediction mode, the value of the transform coefficient, and the size/shape of the current block), the transform coefficient binarization method may be determined on a per block basis. Herein, the block may be the transform coefficient group.

[0405] For example, when the current block is in the intra prediction mode, the exponential scale absolute value comparison binarization is performed. When the current block is in the inter prediction mode, the transform coefficient binarization based on absolute value comparison is performed.

[0406] For example, in the case where the current block is in an N×M size, when the minimum value (min(N, M)) among M and N or the maximum value (max(N, M)) among M and N is larger than a predefined K (wherein, K is a positive integer), the exponential scale absolute value comparison binarization is performed, and otherwise, the transform coefficient binarization based on absolute value comparison is performed.

[0407] FIG. **32** is a flowchart illustrating a method of decoding an image according to an embodiment of the present invention.

[0408] Referring to FIG. **32**, an apparatus for decoding an image may determine the zero out region within the current block at step S**3210**.

[0409] Specifically, when the width or height of the current block is larger than a first predefined size, a region of which the size is equal to or larger than the first predefined size within the current block is determined as the zero out region. Herein, the first predefined size may be 32.

[0410] Further, the zero out region may be determined on the basis of the type of frequency transform of the current block. Specifically, when the type of frequency transform of the current block is DST-7 or DCT-8, a region of which the size is equal to or larger than a second predefined size within the current block is determined as the zero out region. Herein, the second predefined size may be 16.

[0411] On the other hand, when the type of frequency transform of the current block is DCT-2, a region of which the size is equal to or larger than a third predefined size within the current block is determined as the zero out region. Herein, the third predefined size may be 32.

[0412] Further, the apparatus for decoding the image may partition the region except for the zero out region within the current block on a per transform coefficient group basis at step S**3220**.

[0413] Herein, the size of the transform coefficient group may be determined on the basis of the width and the height of the current block. Specifically, when the width or height of the current block is smaller than a fourth predefined size, the size of the transform coefficient group is determined to be a fifth predefined size. When the width and the height of the current block is larger than the fourth predefined size, the size of the transform coefficient group is determined to be a sixth predefined size. Herein, the sixth predefined size may be 4, and the fifth predefined size may be 2.

[0414] In the meantime, the shape of the transform coefficient group is determined to be a non-square shape when the area of the current block is larger than a predefined area and the shape of the current block is a non-square shape. Herein, the predefined area may be 8.

[0415] Further, the apparatus for decoding the image may decode the transform coefficient on a per transform coefficient group basis at step S**3230**.

[0416] The method of decoding the image has been described above with reference to FIG. **32**.

[0417] A method of encoding an image of the present invention may be described similarly to the method of decoding the image described with reference to FIG. **32**, so that a redundant description will be omitted.

[0418] The bitstream generated by the method of encoding the image of the present invention may be temporarily stored in a recording medium.

[0419] The above embodiments may be performed in the same method in an encoder and a decoder.

[0420] At least one or a combination of the above embodiments may be used to encode/decode a video.

[0421] A sequence of applying to above embodiment may be different between an encoder and a decoder, or the sequence applying to above embodiment may be the same in the encoder and the decoder.

[0422] The above embodiment may be performed on each luma signal and chroma signal, or the above embodiment may be identically performed on luma and chroma signals.

[0423] A block form to which the above embodiments of the present invention are applied may have a square form or a non-square form.

[0424] The above embodiment of the present invention may be applied depending on a size of at least one of a coding block, a prediction block, a transform block, a block, a current block, a coding unit, a prediction unit, a transform unit, a unit, and a current unit. Herein, the size may be defined as a minimum size or maximum size or both so that the above embodiments are applied, or may be defined as a fixed size to which the above embodiment is applied. In addition, in the above embodiments, a first embodiment may be applied to a first size, and a second embodiment may be applied to a second size. In other words, the above embodiments may be applied in combination depending on a size. In addition, the above embodiments may be applied when a size is equal to or greater that a minimum size and equal to or smaller than a maximum size. In other words, the above embodiments may be applied when a block size is included within a certain range.

[0425] For example, the above embodiments may be applied when a size of current block is 8×8 or greater. For example, the above embodiments may be applied when a size of current block is 4×4 or greater. For example, the above embodiments may be applied when a size of current block is 16×16 or greater. For example, the above embodiments may be applied when a size of current block is equal to or greater than 16×16 and equal to or smaller than 64×64.

[0426] The above embodiments of the present invention may be applied depending on a temporal layer. In order to identify a temporal layer to which the above embodiments may be applied, a corresponding identifier may be signaled, and the above embodiments may be applied to a specified temporal layer identified by the corresponding identifier. Herein, the identifier may be defined as the lowest layer or the highest layer or both to which the above embodiment may be applied, or may be defined to indicate a specific layer to which the embodiment is applied. In addition, a fixed temporal layer to which the embodiment is applied may be defined.

[0427] For example, the above embodiments may be applied when a temporal layer of a current image is the lowest layer. For example, the above embodiments may be applied when a tempor al layer identifier of a current image is 1. For example, the above embodiments may be applied when a temporal layer of a current image is the highest layer.

[0428] A slice type or a tile group type to which the above embodiments of the present invention are applied may be defined, and the above embodiments may be applied depending on the corresponding slice type or tile group type.

[0429] In the above-described embodiments, the methods are described based on the flowcharts with a series of steps or units, but the present invention is not limited to the order of the steps, and rather, some steps may be performed simultaneously or in different order with other steps. In addition, it should be appreciated by one of ordinary skill in the art that the steps in the flowcharts do not exclude each other and that other steps may be added to the flowcharts or some of the steps may be deleted from the flowcharts without influencing the scope of the present invention.

[0430] The embodiments include various aspects of examples. All possible combinations for various aspects may not be described, but those skilled in the art will be able to recognize different combinations. Accordingly, the present invention may include all replacements, modifications, and changes within the scope of the claims.

[0431] The embodiments of the present invention may be implemented in a form of program instructions, which are executable by various computer components, and recorded in a computer-readable recording medium. The computer-readable recording medium may include stand-alone or a combination of program instructions, data files, data structures, etc. The program instructions recorded in the computer-readable recording medium may be specially designed and constructed for the present invention, or well-known to a person of ordinary skilled in computer software technology field. Examples of the computer-readable recording medium include magnetic recording media such as hard disks, floppy disks, and magnetic tapes; optical data storage media such asCD-ROMs or DV D-ROM s; magneto-optimum media such as floptical disks; and hardware devices, such as read-only memory (ROM), random-access memory (RAM), flash memory, etc., which are particularly structured to store and implement the program instruction. Examples of the program instructions include not only a mechanical language code formatted by a compiler but also a high level language code that may be implemented by a computer using an interpreter. The hardware devices may be configured to be operated by one or more software modules or vice versa to conduct the processes according to the present invention.

[0432] Although the present invention has been described in terms of specific items such as detailed elements as well as the limited embodiments and the drawings, they are only provided to help more general understanding of the invention, and the present invention is not limited to the above embodiments. It will be appreciated by those skilled in the art to which the present invention pertains that various modifications and changes may be made from the above description.

[0433] Therefore, the spirit of the present invention shall not be limited to the above-described embodiments, and the entire scope of the appended claims and their equivalents will fall within the scope and spirit of the invention.

INDUSTRIAL APPLICABILITY

[0434] The present invention may be used in an apparatus for encoding/decoding an image.

1. A method of decoding an image, the method comprising:
- determining a non-zeroed region of a current block;
- partitioning the non-zeroed region into transform coefficient groups;
- decoding transform coefficients of the non-zeroed region on the per transform coefficient group basis; and
- generating a residual block of the current block based on the transform coefficients of the non-zeroed region,
- wherein a size of each of the transform coefficient groups is determined based on a width and a height of the non-zeroed region, and
- wherein, for the non-zeroed region of which width or height is smaller than a first predefined size, a shape of each of the transform coefficient groups is determined to be a non-square shape having a second predefined size, when an area of the non-zeroed region is larger than a predefined area and a shape of the non-zeroed region is a non-square shape, and
- wherein, for the non-zeroed region of which width and height is larger than the first predefined size, each of the transform coefficient groups is determined to have a third predefined size which is different from the second predefined size.

2. The method of claim **1**, wherein at the determining of the non-zeroed region, when a width or a height of the current block is larger than a fourth predefined size, a region of which a size is larger than the fourth predefined size within the current block is determined to be a zeroed region,
- a region except for the zeroed region within the current block is determined to be the non-zeroed region.

3. The method of claim **1**, wherein at the determining of the non-zeroed region, the determining is based on a type of frequency transform of the current block.

4. The method of claim **1**, wherein at the determining of the non-zeroed region, when a type of frequency transform of the current block is DST-7 or DCT-8, a region of which a size is equal to or smaller than a fifth predefined size within the current block is determined to be the non-zeroed region.

5. The method of claim **1**, wherein at the determining of the non-zeroed region, when a type of frequency transform of the current block is DCT-2, a region of which a size is equal to or smaller than a sixth predefined size within the current block is determined to be the non-zeroed region.

6. A method of encoding an image, the method comprising:
- generating transform coefficients of a current block;

determining a non-zeroed region within the current block;

partitioning the non-zeroed region into transform coefficient groups; and

encoding transform coefficients of the non-zeroed region on the per transform coefficient group basis,

wherein a size of each of the transform coefficient groups is determined based on a width and a height of the non-zeroed region, and

wherein, for the non-zeroed region of which width or height is smaller than a first predefined size, a shape of each of the transform coefficient groups is determined to be a non-square shape having a second predefined size, when an area of the non-zeroed region is larger than a predefined area and a shape of the non-zeroed region is a non-square shape, and

wherein, for the non-zeroed region of which width and height is larger than the first predefined size, each of the transform coefficient groups is determined to have a third predefined size which is different from the second predefined size.

7. The method of claim **6**, wherein at the determining of the non-zeroed region, when a width or a height of the current block is larger than a fourth predefined size, a region of which a size is larger than the fourth predefined size within the current block is determined to be a zeroed region,

a region except for the zeroed region within the current block is determined to be the non-zeroed region.

8. The method of claim **6**, wherein at the determining of the non-zeroed region, the determining is based on a type of frequency transform of the current block.

9. The method of claim **6**, wherein at the determining of the non-zeroed region, when a type of frequency transform of the current block is DST-7 or DCT-8, a region of which a size is equal to or smaller than a fifth predefined size within the current block is determined to be the non-zeroed region.

10. The method of claim **6**, wherein at the determining of the non-zeroed region, when a type of frequency transform of the current block is DCT-2, a region of which a size is equal to or smaller than a sixth predefined size within the current block is determined to be the non-zeroed region.

11. A non-transitory computer readable recording medium storing a bitstream generated by a method of encoding an image,

wherein the method comprising:

generating transform coefficients of a current block;

determining a non-zeroed region within the current block;

partitioning the non-zeroed region into transform coefficient groups; and

encoding transform coefficients of the non-zeroed region on the per transform coefficient group basis, and

wherein a size of each of the transform coefficient groups is determined based on a width and a height of the non-zeroed region, and

wherein, for the non-zeroed region of which width or height is smaller than a first predefined size, a shape of each of the transform coefficient groups is determined to be a non-square shape having a second predefined size, when an area of the non-zeroed region is larger than a predefined area and a shape of the non-zeroed region is a non-square shape, and

wherein, for the non-zeroed region of which width and height is larger than the first predefined size, each of the transform coefficient groups is determined to have a third predefined size which is different from the second predefined size.

* * * * *