

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
3 March 2005 (03.03.2005)

PCT

(10) International Publication Number
WO 2005/020587 A1

(51) International Patent Classification⁷: **H04N 7/32**
(21) International Application Number:
PCT/KR2004/002050
(22) International Filing Date: 16 August 2004 (16.08.2004)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/497,567 26 August 2003 (26.08.2003) US
10-2003-0065863
23 September 2003 (23.09.2003) KR

(71) Applicant (for all designated States except US): **SAM-SUNG ELECTRONICS CO., LTD.** [KR/KR]; 416 Maetan-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do 442-742 (KR).

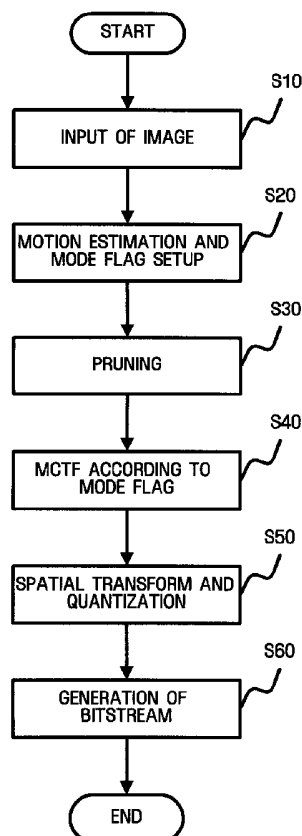
(72) Inventors; and
(75) Inventors/Applicants (for US only): **HA, Ho-Jin** [KR/KR]; 21/2, 994-39 Doksan 3-dong, Geumcheon-gu, Seoul 153-827 (KR). **YIM, Chang-Hoon** [KR/KR]; #301-1506 Cheonggu Apt., Hongje 4-dong, Seodaemun-gu, Seoul 120-786 (KR). **LEE, Bae-Keun** [KR/KR]; 142-10 Chunui-dong, Wonmi-gu, Bucheon-si, Gyeonggi-do 420-858 (KR). **HAN, Woo-Jin** [KR/KR]; #108-703 Jugong 2-danji Apt., Hwanggol Maeul, Yeongtong-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do 442-739 (KR).

(74) Agents: **KIM, Dong-Jin** et al.; 6th Fl. Youngpoong Bldg., 142 Nonhyun-dong, Gangnam-gu, Seoul 135-749 (KR).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,

[Continued on next page]

(54) Title: ADAPTIVE INTERFRAME WAVELET VIDEO CODING METHOD, COMPUTER READABLE RECORDING MEDIUM AND SYSTEM THEREFOR



(57) Abstract: An adaptive interframe wavelet video coding method, a computer readable recording medium and system therefor are provided. The interframe wavelet video coding method includes (a) receiving a group-of-frames including a plurality of frames and determining a mode flag according to a predetermined procedure using motion vectors of boundary pixels, (b) temporally decomposing the frames included in the group-of-frames in predetermined directions in accordance with the determined mode flag, and (c) performing spatial transform and quantization on the frames obtained by performing step (b), thereby generating a bitstream. Since an appropriate temporal filtering is performed in accordance with a boundary condition, efficiency of interframe wavelet video coding is increased.



KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,

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

Description

ADAPTIVE INTERFRAME WAVELET VIDEO CODING METHOD, COMPUTER READABLE RECORDING MEDIUM AND SYSTEM THEREFOR

Technical Field

- [1] The present invention relates to a wavelet video coding method, a computer readable recording medium and system therefor, and more particularly, to an interframe wavelet video coding (IWVC) method which decreases an average temporal distance by changing a temporal filtering direction.

Background Art

- [2] With the development of information communication technology including the Internet, video communication as well as text and voice communication has increased. Conventional text communication cannot satisfy the various demands of users, and thus multimedia services that can provide various types of information such as text, pictures, and music have increased. Multimedia data requires large capacity storage mediums and wide bandwidths for transmission since the amount of multimedia data is usually large. For example, a 24-bit true color image having a resolution of $640 * 480$ needs a capacity of $640 * 480 * 24$ bits, i.e., data of about 7.37 Mbits, per frame. When this image is transmitted at a speed of 30 frames per second, a bandwidth of 221 Mbits/sec is required. When a 90-minute movie based on such an image is stored, a storage space of about 1200 Gbits is required. Accordingly, a compression coding method is a requisite for transmitting multimedia data including text, video, and audio.
- [3] A basic principle of data compression is removing data redundancy. Data can be compressed by removing spatial redundancy in which the same color or object is repeated in an image, temporal redundancy in which there is little change between adjacent frames in a moving image or the same sound is repeated in audio, or mental visual redundancy taking into account human eyesight and limited perception of high frequency. Data compression can be classified into lossy/lossless compression according to whether source data is lost, intraframe/interframe compression according to whether individual frames are compressed independently, and symmetric/asymmetric compression according to whether time required for compression is the same as time required for recovery. In addition, data compression is defined as real-time compression when a compression/recovery time delay does not exceed 50 ms and as scalable compression when frames have different resolutions. For text or medical

data, lossless compression is usually used. For multimedia data, lossy compression is usually used. Meanwhile, intraframe compression is usually used to remove spatial redundancy, and interframe compression is usually used to remove temporal redundancy.

[4] FIG. 1 is a flowchart of a conventional three-dimensional IWVC method.

[5] First, an image is received in group-of-frames (GOF) units in step S1. The GOF includes a plurality of frames, e.g., 16 frames. In IWVC, various operations are performed in GOF units.

[6] Next, motion estimation is performed using hierarchical variable size block matching (HVSBM) in step S2. Referring to FIG. 2, which illustrates a motion estimation using HVSBM, an original image has a size of $N * N$, images of level 0 ($N * N$), of level 1 ($N/2 * N/2$), and of level 2 ($N/4 * N/4$) are obtained using wavelet transform. For the image of level 2, a motion estimation block size is changed from $16 * 16$ to $8 * 8$ and $4 * 4$, and a motion estimation (ME) and a Magnitude of Absolute Distortion (MAD) are obtained with respect to each block. Similarly, for the image of level 1, the motion estimation block size is changed from $32 * 32$ to $16 * 16$, $8 * 8$, and $4 * 4$, and an ME and a MAD are obtained with respect to each block. For the image of level 0, the motion estimation block size is changed from $64 * 64$ to $32 * 32$, $16 * 16$, $8 * 8$, and $4 * 4$, and an ME and a MAD are obtained with respect to each block.

[7] Next, as shown in FIG. 1, an ME tree is pruned to minimize the MAD in step S3.

[8] Motion compensated temporal filtering (MCTF) is performed using a pruned optimal ME in step S4. Referring to FIG. 3, at temporal level 0, MCTF is performed forward with respect to 16 image frames, thereby obtaining 8 low-frequency frames and 8 high-frequency frames. At temporal level 1, MCTF is performed forward with respect to the 8 low-frequency frames, thereby obtaining 4 low-frequency frames and 4 high-frequency frames. At temporal level 2, MCTF is performed forward with respect to the 4 low-frequency frames obtained at temporal level 1, thereby obtaining 2 low-frequency frames and 2 high-frequency frames. Lastly, at temporal level 3, MCTF is performed forward with respect to the 2 low-frequency frames obtained at temporal level 2, thereby obtaining a single low-frequency frame and a single high-frequency frame. Accordingly, as a result of MCTF, a total of 16 subbands H1, H3, H5, H7, H9, H11, H13, H15, LH2, LH6, LH10, LH14, LLH4, LLH12, LLLH8, and LLLL16 including 15 high-frequency frames and a single low-frequency frame at the last level are obtained.

[9] After obtaining the 16 subbands, spatial transform and quantization are performed

on the 16 subbands in step S5. Thereafter, a bitstream including data obtained by performing spatial transform and quantization on the 16 subbands, motion estimation data, and a header is generated in step S6.

[10] Although such conventional IWVC has excellent scalability, it does not have satisfactory performance as compared to other conventional video coding methods. An example of IWVC performance depending upon a boundary condition will be described with reference to FIG. 4.

[11] FIG. 4 is diagram comparing performances of conventional MCTF with respect to a boundary condition.

[12] FIG. 4 illustrates a best case of forward MCTF where an external image comes into a frame and a worst case of forward MCTF where an internal image goes out of the frame. Where MCTF is performed forward, a temporally preceding image is replaced with a filtered high-frequency image, and a temporally succeeding image is replaced with a filtered low-frequency image. For video coding, high-frequency frames and a single low-frequency frame at a highest level are used. In other words, performance of video coding depends on whether a component of a high-frequency frame is large or small.

[13] In a case where the external image comes into the frame, a T-1 frame is replaced with a high-frequency image, and a T frame is replaced with a low-frequency image. All image blocks in the T-1 frame can be exactly matched with image blocks, respectively, in the T-frame, and thus a magnitude of a high frequency component proportional to a difference between two image blocks is less compared to a case where the image blocks are not matched exactly. In other words, a size of the T-1 frame to be replaced with a high-frequency image is small.

Disclosure of Invention

Technical Problem

[14] Conversely, in the worst case where the internal image goes out of the frame, all of the image blocks in the T-1 frame are not exactly matched with the image blocks in the T-frame. Here, image blocks A and N that do not have their matches are coupled with image blocks B and M, respectively, giving a least difference therebetween. Since a difference between the image blocks A and B and a difference between the image blocks N and M are needed to be expressed, the size of the T-1 frame is increased.

Technical Solution

[15] As described above, performance of MCTF greatly changes depending on a boundary condition such as an incoming image or an outgoing image. Therefore, a

video coding method allowing a filtering direction to be adaptively changed according to a boundary condition during MCTF is desired.

[16] The present invention provides an adaptive interframe wavelet video coding (IWVC) method allowing a direction of temporal filtering to be changed according to a boundary condition.

[17] The present invention also provides a computer readable recording medium and a system which can perform the adaptive IWVC method.

[18] According to an aspect of the present invention, there is provided an IWVC method comprising, (a) receiving a group-of-frames including a plurality of frames and determining a mode flag according to a predetermined procedure using motion vectors of boundary pixels; (b) temporally decomposing the frames included in the group-of-frames in predetermined directions in accordance with the determined mode flag; and (c) performing spatial transform and quantization on the frames obtained by performing step (b), thereby generating a bitstream.

[19] Preferably, in step (a), the group-of-frames comprises 16 frames. Step (a) may comprise determining the mode flag according to the predetermined procedure using motion vectors obtained at a boundary having a predetermined thickness among motion vectors of pixels obtained through motion estimation using hierarchical variable size block matching (HVSBM). Meanwhile, the motion vectors used to determine the mode flag may be motion vectors of pixels at left and right boundaries, or motion vectors of pixels at left, right, upper and lower boundaries. In the first case, the mode flag F is preferably determined using the following algorithm:

[20]

if (abs(L)< Threshold)then L=0

if (abs(R)< Threshold)then R=0

if((L < 0 and R==0)or (L==0 and R > 0)or (L < 0 and R > 0))then F=0

else if((L > 0 and R==0)or (L==0 and R < 0)or (L > 0 and R < 0))then F=1

else F=2

[21] where L denotes an average of X components of motion vectors of pixels at the left boundary having the predetermined thickness, and R denotes an average of X components of motion vectors of pixels at the right boundary having the predetermined thickness, wherein step (b) comprises temporally decomposing the frames included in

the group-of-frames in a forward direction when $F=0$, temporally decomposing the frames included in the group-of-frames in a backward direction when $F=1$, and temporally decomposing the frames included in the group-of-frames in forward and backward directions combined in a predetermined sequence when $F=2$. In the latter case, the mode flag F is preferably determined using the following algorithm:

[22]

if ($\text{abs}(L) < \text{Threshold}$) then $L=0$

if ($\text{abs}(R) < \text{Threshold}$) then $R=0$

if ($\text{abs}(U) < \text{Threshold}$) then $U=0$

if ($\text{abs}(D) < \text{Threshold}$) then $D=0$

if ((($L < 0$ and $R==0$) or ($L==0$ and $R > 0$) or ($L < 0$ and $R > 0$)) and (($D < 0$ and $U==0$) or ($D==0$ and $U > 0$) or ($D < 0$ and $U > 0$) or ($D==0$ and $U==0$))) then $F=0$

else if ((($L > 0$ and $R==0$) or ($L==0$ and $R < 0$) or ($L > 0$ and $R < 0$)) and (($D > 0$ and $U==0$) or ($D==0$ and $U < 0$) or ($D > 0$ and $U < 0$) or ($D==0$ and $U==0$))) then $F=1$

else $F=2$

[23]

where L denotes an average of X components of motion vectors of pixels at the left boundary having the predetermined thickness, R denotes an average of X components of motion vectors of pixels at the right boundary having the predetermined thickness, U denotes an average of Y components of motion vectors of pixels at the upper boundary having the predetermined thickness, and D denotes an average of Y components of motion vectors of pixels at the lower boundary having the predetermined thickness, wherein step (b) comprises temporally decomposing the frames included in the group-of-frames in a forward direction when $F=0$, temporally decomposing the frames included in the group-of-frames in a backward direction when $F=1$, and temporally decomposing the frames included in the group-of-frames in forward and backward directions combined in a predetermined sequence when $F=2$.

[24]

In either case, when $F=2$ in step (b), the frames are preferably decomposed such

that an average temporal distance between frames is minimized.

- [25] Programs executing the adaptive IWVC method may be recorded onto a computer readable recording medium to be used in a computer.
- [26] According to another aspect of the present invention, there is provided an IWVC system which receives a group-of-frames including a plurality of frames and generates a bitstream. The IWVC system comprises a motion estimation/mode determination block which receives the group-of-frames, obtains motion vectors of pixels in each of the frames using a predetermined procedure, and determines a mode flag using motion vectors of boundary pixels among the obtained motion vectors; and a motion compensation temporal filtering block which decomposes the frames into low- and high-frequency frames in a predetermined temporal direction in accordance with the mode flag determined by the motion estimation/mode determination block using the motion vectors.
- [27] The interframe wavelet video coding system may further comprise a spatial transform block which wavelet-decomposes the low- and high-frequency frames generated by the motion compensation temporal filtering block into spatial low- and high-frequency components.

Description of Drawings

- [28] The above and other features and advantages of the present invention will become more apparent by describing in detail exemplary embodiments thereof with reference to the attached drawings in which:
- [29] FIG. 1 is a flowchart of a conventional three-dimensional interframe wavelet video coding (IWVC) method;
- [30] FIG. 2 illustrates conventional motion estimation using hierarchical variable size block matching (HVSBM);
- [31] FIG. 3 illustrates conventional motion compensated temporal filtering (MCTF);
- [32] FIG. 4 is diagram comparing performances of conventional MCTF with respect to a boundary condition;
- [33] FIG. 5 is a flowchart of an adaptive IWVC method according to an embodiment of the present invention;
- [34] FIG. 6 illustrates a reference for determining an MCTF direction according to a boundary condition;
- [35] FIGS. 7 and 8 illustrate boundary pixels used to determine a mode flag;
- [36] FIG. 9 illustrates MCTF directions according to a mode flag representing a boundary condition; and

- [37] FIG. 10 is a functional block diagram of a system for adaptive IWVC according to an embodiment of the present invention.

Mode for Invention

- [38] Exemplary, non-limiting, embodiments of the present invention will now be described with reference to the accompanying drawings.
- [39] FIG. 5 is a flowchart of an adaptive interframe wavelet video coding (IWVC) method according to an embodiment of the present invention.
- [40] An image is received in group-of-frames (GOF) units in step S10. A single GOF includes a plurality of frames and preferably includes 2^n frames (where 'n' is a natural number), e.g., 2, 4, 8, 16, or 32 frames, to facilitate computation and management. When the number of frames included in a GOF increases, video coding efficiency increases while buffering time and coding time also increases unfavorably. As the number of frames included in a GOF decreases, video coding efficiency decreases. In the embodiment of the present invention, a single GOF includes 16 frames.
- [41] After receiving the image, motion estimation is performed and a mode flag is set in step S20. Preferably, the motion estimation is performed using hierarchical variable size block matching (HVSBM) as described with reference to FIG. 1. The mode flag is used to determine a direction of temporal filtering according to a boundary condition. A reference for determining a mode flag will be described with reference to FIGS. 6, 7 and 8.
- [42] After the motion estimation and mode flag setup, pruning is performed in the same manner as in conventional technology in step S30.
- [43] Next, motion compensated temporal filtering (MCTF) is performed using a pruned motion vector in step S40. An MCTF direction in accordance with the mode flag will be described with reference to FIG. 9.
- [44] After completing the MCTF, 16 subbands resulting from the MCTF are subjected to spatial transform and quantization in step S50. Thereafter, a bitstream including data resulting from the spatial transform and quantization, motion vector data, and the mode flag is generated in step S60.
- [45] FIG. 6 illustrates a reference for determining an MCTF direction according to a boundary condition, and FIGS. 7 and 8 illustrate boundary pixels used to determine a mode flag.
- [46] FIGS. 6 illustrates cases where an internal image goes out of the frame. FIG. 6 illustrates forward MCTF and backward MCTF. In other words, image blocks B and N flow out of the frame when a T-1 frame is converted into a T frame. In the worst case

of forward MCTF shown in FIG. 6, the image blocks B and N in the T-1 frame do not have their matches in the T frame. Thus, the image blocks B and N in the T-1 frame are compared with image blocks C and M, respectively, in the T frame. In this situation, a difference between the image blocks B and C and a difference between the image blocks N and M are large, which increases the amount of information of the T-1 frame to be replaced with a high-frequency frame. Conversely, in the best case of backward MCTF shown in FIG. 6, each image block in the T frame to be replaced with a high-frequency frame has its matches in the T-1 frame, and therefore, the amount of information of the high-frequency frame, i.e., the T frame, may be decreased.

[47] In a comprehensive conception, forward MCTF is more efficient in a case where a new image comes into a frame through a boundary while backward MCTF is more efficient in a case where an image goes out of the frame through a boundary. In other cases, it is efficient to properly combine forward MCTF and backward MCTF. In other words, video coding efficiency and performance can be increased by properly selecting either forward or backward MCTF according to a boundary condition of an input GOF. In setting a mode flag, a basic principle is made that forward MCTF is used when a new image comes into a frame, backward MCTF is used when an image goes out of a frame, and forward MCTF and backward MCTF are properly combined in other cases.

[48] The mode flag can be determined using a motion vector for pixels at a boundary of a frame. As shown in FIG. 7, pixels at right and left boundaries of a frame may be used in a first embodiment. Alternatively, as shown in FIG. 8, pixels at right, left, upper, and lower boundaries of a frame may be used in a second embodiment. Video coding performance depends on a thickness of a boundary used to determine the mode flag. Where the boundary is too thin, information regarding output/input of a particular image may be missed. Conversely, where the boundary is too thick, a boundary condition may not be sharply identified. Accordingly, the thickness of the boundary needs to be appropriately determined. In embodiments of the present invention, the boundary has a thickness of 32 pixels.

[49] In determining the mode flag, motion vectors of pixels in each frames are obtained using HVSBM. A mode flag is determined based on the motion vectors of pixels in the frames. The mode flag may be different according to a temporal level, but it is preferable to determine the mode flag at temporal level 0.

[50] In the first embodiment shown in FIG. 7, a mode flag is determined using motion vectors at left and right boundaries of each frame because a new image usually comes into or goes out of a frame of a moving picture in an X direction. An average of

motion vectors of pixels at the left boundary of each of all frames included in a single GOF is obtained. An X component of the average motion vector at the left boundary is denoted by 'L.' Similarly, an average of motion vectors of pixels at the right boundary of each of all frames included in a single GOF is obtained. An X component of the average motion vector at the right boundary is denoted by 'R.' Where an L value less than 0 indicates that an image comes into the frame through the left boundary, an R value less than 0 indicates that an image goes out of the frame through the right boundary. Similarly, the L value greater than 0 and the R value greater than 0 account for the opposite cases, respectively. Actually, the L or R value may not be 0 even if an image does not come in or go out of the frame. Accordingly, it is preferable that the L and R values not exceeding a predetermined threshold are determined as 0. When an image comes into the frame through the left or right boundary, the L value is less than 0 and the R value is equal to or greater than 0, or the L value is less than 0 and the R value is greater than 0. In this case, it is preferable to use forward MCTF. Conversely, when an image goes out of the frame through the left or right boundary, the L value is greater than 0 and the R value is equal to or less than 0, or the L value is greater than 0 and the R value is less than 0. In this case, it is preferable to use backward MCTF. When an image comes into the frame through the left boundary and an image goes out of the frame through the right boundary, it is preferable to appropriately combine forward MCTF and backward MCTF.

[51] As such, a mode flag F can be determined by the following algorithm:

[52]

if (abs(L)< Threshold)then L=0

if (abs(R)< Threshold)then R=0

if((L < 0 and R==0)or (L==0 and R > 0)or (L < 0 and R > 0))then F=0

else if((L > 0 and R==0)or (L==0 and R < 0)or (L > 0 and R < 0))then F=1

else F=2 ,

[53] Here, F=0 indicates a forward mode, F=1 indicates a backward mode, and F=2 indicates a bi-directional mode.

[54] In a second embodiment shown in FIG. 8, left, right, upper, and lower boundaries are used. L and R values are obtained in the same manner as described in the first embodiment, and U and D values are obtained using averages of Y components of

motion vectors. Like the first embodiment, where an image comes into a frame through at least one boundary and an image does not go out of the frame through any of the boundaries, it is preferable to use forward MCTF. Where an image goes out of the frame through at least one boundary and an image does not come into the frame through any of the boundaries, it is preferable to use backward MCTF. In other cases, it is preferable to appropriately combine forward MCTF and backward MCTF.

[55] As such, a mode flag F can be determined by the following algorithm:

[56]

if (abs(L)< Threshold)then L=0

if (abs(R)< Threshold)then R=0

if (abs(U)< Threshold)then U=0

if (abs(D)< Threshold)then D=0

if(((L < 0 and R==0)or (L==0 and R > 0)or (L < 0 and R > 0))and ((D < 0 and U==0)or (D==0 and U > 0)or (D < 0 and U > 0)or (D ==0 and U==0)))then F=0

else if(((L > 0 and R==0)or (L==0 and R < 0)or (L > 0 and R < 0))and ((D > 0 and U==0) or (D==0 and U < 0)or (D > 0 and U < 0)or (D ==0 and U==0)))then F=1

else F=2.

[57] Here, F=0 indicates a forward mode, F=1 indicates a backward mode, and F=2 indicates a bi-directional mode. The first and second embodiments are exemplary, and the spirit of the present invention is not restricted thereto. In other words, a direction of MCTF is appropriately determined using information regarding image input/output at a boundary. Accordingly, the present invention will be considered as including a case where a mode flag is determined to be different among two or some frames more than two in a GOF in addition to the first and second embodiments where a mode flag is determined using average motion vectors obtained with respect to all of the frames in a GOF.

[58] FIG. 9 illustrates MCTF directions according to a mode flag representing a

boundary condition.

[59] In a forward mode, MCTF directions are depicted as + + + + + + +. In a backward mode, MCTF directions are depicted as - - - - - - -. In a bi-directional mode, MCTF directions may be depicted in various ways, but FIG. 9 illustrates an example where MCTF directions are depicted as + - + - + - + - at temporal level 0. Here, '+' indicates a forward direction, and '-' indicates a backward direction.

[60] In each of the forward and backward modes, MCTF is performed in the same direction. However, in the bi-directional mode, video coding performance changes depending on a combination of forward and backward directions. In other words, in the bi-directional mode, a sequence of forward and backward directions may be determined in various ways. Representative examples of a sequence of MCTF directions in the forward, backward, and bi-directional modes are shown in Table 1.

[61]

Table 1

Mode flag	Level 0	Level 1	Level 2	Level 3
Forward direction (F=0)	+++++++	++++	++	+
Backward direction (F=1)	-----	----	--	-
Bi-direction(F=2) a	+--+--+	++--	+-	+(-)
b	+--+--+	+--+	+-	+(-)
c	+++++++	++--	+-	-
d	++++----	++--	+-	-

[62] Various combinations of forward and backward directions may be made in the bi-directional mode, but four cases 'a', 'b', 'c', and 'd' are shown as examples. The cases 'c' and 'd' are characterized in that a low-frequency frame (hereinafter, referred to as a reference frame) at a last level is positioned at a center (i.e., an 8th frame) among 1st through 16th frames. The reference frame is a most essential frame in video coding. The other frames are recovered based on the reference frame. As a temporal distance between a frame and the reference frame increases, recovery performance decreases. Accordingly, in the cases 'c' and 'd', a combination of forward MCTF and backward MCTF is made such that the reference frame is positioned at the center, i.e., the 8th frame, to minimize a temporal distance between the reference frame and each of the other frames.

[63] In the cases 'a' and 'b', an average temporal distance (ATD) is minimized. To calculate an ATD, temporal distances are calculated. A temporal distance is defined as

a positional difference between two frames. Referring to FIG. 3, a temporal distance between a first frame and a second frame is defined as 1, and a temporal distance between the frame L2 and the frame L4 is defined as 2. An ATD is obtained by dividing the sum of temporal distances between frames subjected to an operation for motion estimation in pairs by the number of pairs of frames defined for the motion estimation. In the case 'a',

$$ATD = \frac{8 \times 1 + 4 \times 1 + 2 \times 4 + 1 \times 3}{15} = 1.53.$$

In the case 'b',

$$ATD = \frac{8 \times 1 + 4 \times 1 + 2 \times 3 + 1 \times 5}{15} = 1.53.$$

In the forward mode and the backward mode shown in Table 1,

$$ATD = \frac{8 \times 1 + 4 \times 2 + 2 \times 4 + 1 \times 8}{15} = 2.13.$$

In the case 'c',

$$ATD = \frac{8 \times 1 + 4 \times 2 + 2 \times 4 + 1 \times 2}{15} = 1.73.$$

In the case 'd',

$$ATD = \frac{8 \times 1 + 4 \times 1 + 2 \times 4 + 1 \times 1}{15} = 1.67.$$

In actual simulations, as an ATD was decreased, a PSNR value was increased so that performance of video coding was increased.

[64] FIG. 10 is a functional block diagram of a system for adaptive IWVC according to an embodiment of the present invention.

[65] The system for adaptive IWVC includes a motion estimation/mode determination block 10 which obtains a motion vector and determines a mode using the motion vector, a motion compensation temporal filtering block 40 which removes temporal redundancy using the motion vector and the determined mode, a spatial transform block 50 which removes spatial redundancy, a motion vector encoding block 20 which encodes the motion vector using a predetermined algorithm, a quantization block 60 which quantizes wavelet coefficients of respective components generated by the spatial transform block 50, and a buffer 30 which temporarily stores an encoded bitstream received from the quantization block 60.

- [66] The motion estimation/mode determination block 10 obtains a motion vector used by the motion compensation temporal filtering block 40 using a hierarchical method such as HVSBM. In addition, the motion estimation/mode determination block 10 determines a mode flag for determining temporal filtering directions.
- [67] The motion compensation temporal filtering block 40 decomposes frames into low- and high-frequency frames in a temporal direction using the motion vector obtained by the motion estimation/mode determination block 10. A direction of the decomposition is determined according to the mode flag. Frames are decomposed in GOF units. Through such decomposition, temporal redundancy is removed.
- [68] The spatial transform block 50 wavelet-decomposes frames that have been decomposed in the temporal direction by the motion compensation temporal filtering block 40 into spatial low- and high-frequency components, thereby removing spatial redundancy.
- [69] The motion vector encoding block 20 encodes the motion vector and the mode flag hierarchically obtained by the motion estimation/mode determination block 10 and then transmits the encoded motion vector and the encoded mode flag to the buffer 30.
- [70] The quantization block 60 quantizes and encodes wavelet coefficients of components generated by the spatial transform block 50.
- [71] The buffer 30 stores a bitstream including encoded data, the encoded motion vector, and the encoded mode flag before transmission and is controlled by a rate control algorithm.
- [72] In experiments, performance was increased by about 0.8 dB. In the experiments, Mobile, Tempete, Canoa, and Bus were used, and results of the experiments are shown in Tables 2 through 5.

[73]

Table 2: Mobile, CIF, Frames: 0-299

Bit rate	Forward direction	Backward direction
400	26.1	26.2
600	28.0	28.0
800	29.3	29.2

[74]

Table 3: Tempete CIF, Frames: 0-259

Bit rate	Forward direction	Backward direction
400	29.2	29.2
600	30.7	30.7
800	31.8	31.8

[75]

Table 4: Canoa, CIF, Frames: 0-208

Bit rate	Forward direction	Backward direction
400	23.3	24.8
600	25.2	26.2
800	26.2	27.2

[76]

Table 5: Bus, CIF, Frames: 0-150

Bit rate	Forward direction	Backward direction
400	25.5	26.5
600	27.3	28.2
800	28.6	29.4

Industrial Applicability

[77] According to the present invention, IWVC can be adaptively performed in accordance with a boundary condition. In other words, as compared to conventional methods, a PSNR is increased in the present invention.

[78] It will be understood by those of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope of the present invention as defined by the following claims. Therefore, it is to be appreciated that the above described embodiment is for purposes of illustration only and not to be construed as a limitation of the invention. The scope of the invention is given by the appended claims, rather than the preceding description, and all variations and equivalents which fall within the range of the claims are intended to be embraced therein.

Claims

- [1] An interframe wavelet video coding method comprising:
 - (a) receiving a group-of-frames including a plurality of frames and determining a mode flag according to a predetermined procedure using motion vectors of boundary pixels;
 - (b) temporally decomposing the frames included in the group-of-frames in predetermined directions in accordance with the determined mode flag; and
 - (c) performing spatial transform and quantization on the frames obtained by performing step (b), thereby generating a bitstream.
- [2] The interframe wavelet video coding method of claim 1, wherein in step (a), the group-of-frames comprises 16 frames.
- [3] The interframe wavelet video coding method of claim 1, wherein step (a) comprises determining the mode flag according to the predetermined procedure using motion vectors obtained at a boundary having a predetermined thickness among motion vectors of pixels obtained through motion estimation using hierarchical variable size block matching (HVSBM).
- [4] The interframe wavelet video coding method of claim 3, wherein the motion vectors used to determine the mode flag are motion vectors of pixels at left and right boundaries.
- [5] The interframe wavelet video coding method of claim 4, wherein the mode flag F is determined using the following algorithm:


```

if (abs(L)< Threshold)then L=0

if (abs(R)< Threshold)then R=0

if((L < 0 and R==0)or (L==0 and R > 0)or (L < 0 and R > 0))then F=0

else if((L > 0 and R==0)or (L==0 and R < 0)or (L > 0 and R < 0))then F=1

else F=2

```

where L denotes an average of X components of motion vectors of pixels at the left boundary having the predetermined thickness, and R denotes an average of X components of motion vectors of pixels at the right boundary having the predetermined thickness,

wherein step (b) comprises temporally decomposing the frames included in the group-of-frames in a forward direction when F=0, temporally decomposing the

frames included in the group-of-frames in a backward direction when $F=1$, and temporally decomposing the frames included in the group-of-frames in forward and backward directions combined in a predetermined sequence when $F=2$.

- [6] The interframe wavelet video coding method of claim 5, wherein when $F=2$ in step (b), the frames are decomposed such that an average temporal distance between frames is minimized.
- [7] The interframe wavelet video coding method of claim 3, wherein the motion vectors used to determine the mode flag are motion vectors of pixels at left, right, upper, and lower boundaries.
- [8] The interframe wavelet video coding method of claim 7, wherein the mode flag F is determined using the following algorithm:

if ($\text{abs}(L) < \text{Threshold}$) then $L=0$

if ($\text{abs}(R) < \text{Threshold}$) then $R=0$

if ($\text{abs}(U) < \text{Threshold}$) then $U=0$

if ($\text{abs}(D) < \text{Threshold}$) then $D=0$

if ((($L < 0$ and $R==0$) or ($L==0$ and $R > 0$) or ($L < 0$ and $R > 0$)) and (($D < 0$ and $U==0$) or ($D==0$ and $U > 0$) or ($D < 0$ and $U > 0$) or ($D==0$ and $U==0$))) then $F=0$

else if ((($L > 0$ and $R==0$) or ($L==0$ and $R < 0$) or ($L > 0$ and $R < 0$)) and (($D > 0$ and $U==0$) or ($D==0$ and $U < 0$) or ($D > 0$ and $U < 0$) or ($D==0$ and $U==0$))) then $F=1$

else $F=2$

where L denotes an average of X components of motion vectors of pixels at the left boundary having the predetermined thickness, R denotes an average of X components of motion vectors of pixels at the right boundary having the predetermined thickness, U denotes an average of Y components of motion vectors of pixels at the upper boundary having the predetermined thickness, and D denotes an average of Y components of motion vectors of pixels at the lower boundary having the predetermined thickness,

wherein step (b) comprises temporally decomposing the frames included in the

- group-of-frames in a forward direction when $F=0$, temporally decomposing the frames included in the group-of-frames in a backward direction when $F=1$, and temporally decomposing the frames included in the group-of-frames in forward and backward directions combined in a predetermined sequence when $F=2$.
- [9] The interframe wavelet video coding method of claim 8, wherein when $F=2$ in step (b), the frames are decomposed such that an average temporal distance between frames is minimized.
- [10] A recording medium comprising commands which can be executed in a computer, the commands executing:
- (a) receiving a group-of-frames including a plurality of frames and determining a mode flag according to a predetermined procedure using motion vectors of boundary pixels;
 - (b) temporally decomposing the frames included in the group-of-frames in predetermined directions in accordance with the determined mode flag; and
 - (c) performing spatial transform and quantization on the frames obtained by performing step (b), thereby generating a bitstream.
- [11] The recording medium of claim 10, wherein in step (a), the group-of-frames comprises 16 frames.
- [12] The recording medium of claim 9, wherein step (a) comprises determining the mode flag according to the predetermined procedure using motion vectors obtained at a boundary having a predetermined thickness among motion vectors of pixels obtained through motion estimation using hierarchical variable size block matching (HVSBM).
- [13] The recording medium of claim 12, wherein the motion vectors used to determine the mode flag are motion vectors of pixels at left and right boundaries.
- [14] The recording medium of claim 13, wherein the mode flag F is determined using the following algorithm:
- ```

if (abs(L) < Threshold) then L=0
if (abs(R) < Threshold) then R=0
if ((L < 0 and R==0) or (L==0 and R > 0) or (L < 0 and R > 0)) then F=0
else if ((L > 0 and R==0) or (L==0 and R < 0) or (L > 0 and R < 0)) then F=1
else F=2

```

where L denotes an average of X components of motion vectors of pixels at the left boundary having the predetermined thickness, and R denotes an average of X components of motion vectors of pixels at the right boundary having the predetermined thickness,

wherein step (b) comprises temporally decomposing the frames included in the group-of-frames in a forward direction when  $F=0$ , temporally decomposing the frames included in the group-of-frames in a backward direction when  $F=1$ , and temporally decomposing the frames included in the group-of-frames in forward and backward directions combined in a predetermined sequence when  $F=2$ .

[15] The recording medium of claim 14, wherein when  $F=2$  in step (b), the frames are decomposed such that an average temporal distance between frames is minimized.

[16] The recording medium of claim 12, wherein the motion vectors used to determine the mode flag are motion vectors of pixels at left, right, upper, and lower boundaries.

[17] The recording medium of claim 16, wherein the mode flag F is determined using the following algorithm:

if (abs(L)< Threshold)then L=0

if (abs(R)< Threshold)then R=0

if (abs(U)< Threshold)then U=0

if (abs(D)< Threshold)then D=0

if(((L < 0 and R==0)or (L==0 and R > 0)or (L < 0 and R > 0))and ((D < 0 and U==0)or (D==0 and U > 0)or (D < 0 and U > 0)or (D ==0 and U==0)))then F=0

else if(((L > 0 and R==0)or (L==0 and R < 0)or (L > 0 and R < 0))and ((D > 0 and U==0) or (D==0 and U < 0)or (D > 0 and U < 0)or (D ==0 and U==0)))then F=1

else F=2

where L denotes an average of X components of motion vectors of pixels at the left boundary having the predetermined thickness, R denotes an average of X

components of motion vectors of pixels at the right boundary having the predetermined thickness, U denotes an average of Y components of motion vectors of pixels at the upper boundary having the predetermined thickness, and D denotes an average of Y components of motion vectors of pixels at the lower boundary having the predetermined thickness,

wherein step (b) comprises temporally decomposing the frames included in the group-of-frames in a forward direction when  $F=0$ , temporally decomposing the frames included in the group-of-frames in a backward direction when  $F=1$ , and temporally decomposing the frames included in the group-of-frames in forward and backward directions combined in a predetermined sequence when  $F=2$ .

[18] The recording medium of claim 17, wherein when  $F=2$  in step (b), the frames are decomposed such that an average temporal distance between frames is minimized.

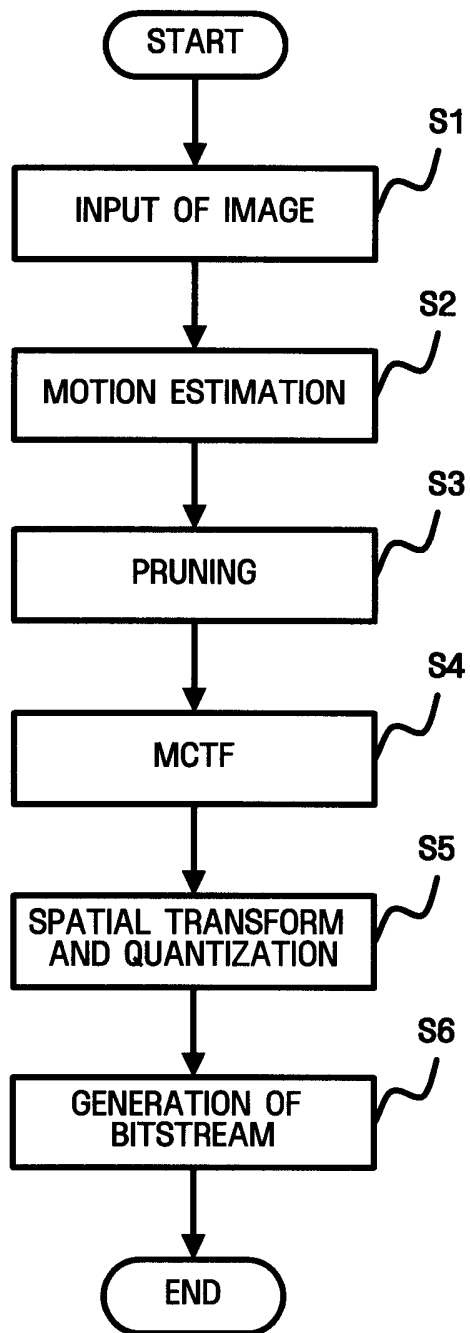
[19] An interframe wavelet video coding system which receives a group-of-frames including a plurality of frames and generates a bitstream, the interframe wavelet video coding system comprising:

a motion estimation/mode determination block which receives the group-of-frames, obtains motion vectors of pixels in each of the frames using a predetermined procedure, and determines a mode flag using motion vectors of boundary pixels among the obtained motion vectors; and

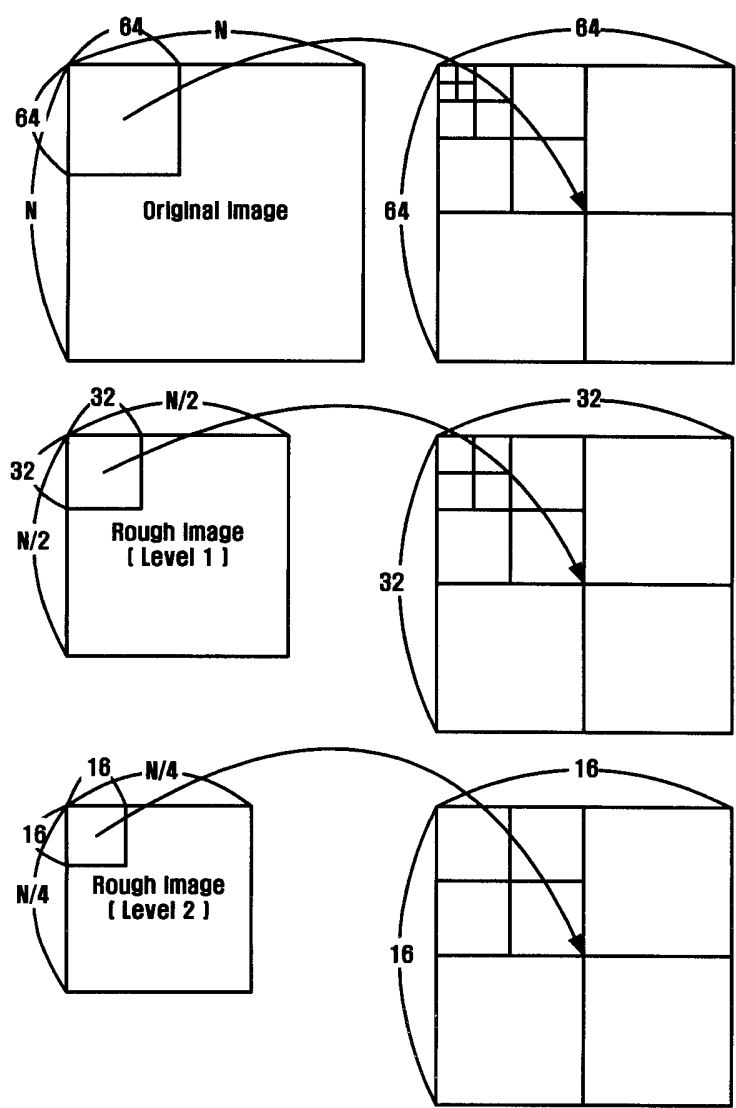
a motion compensation temporal filtering block which decomposes the frames into low- and high-frequency frames in a predetermined temporal direction in accordance with the mode flag determined by the motion estimation/mode determination block using the motion vectors.

[20] The interframe wavelet video coding system of claim 19, further comprising a spatial transform block which wavelet-decomposes the low- and high-frequency frames generated by the motion compensation temporal filtering block into spatial low- and high-frequency components.

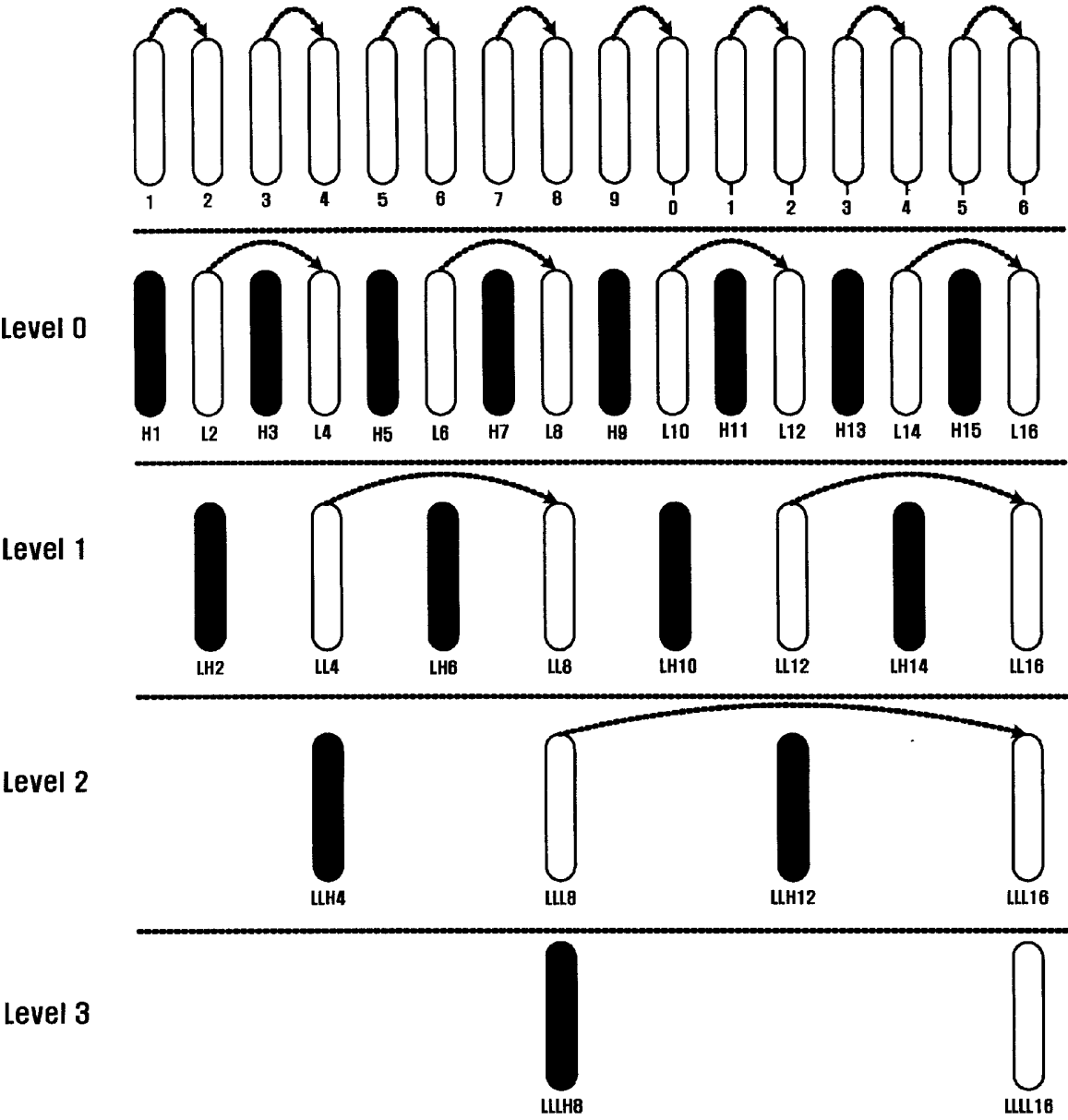
[Fig. 1]



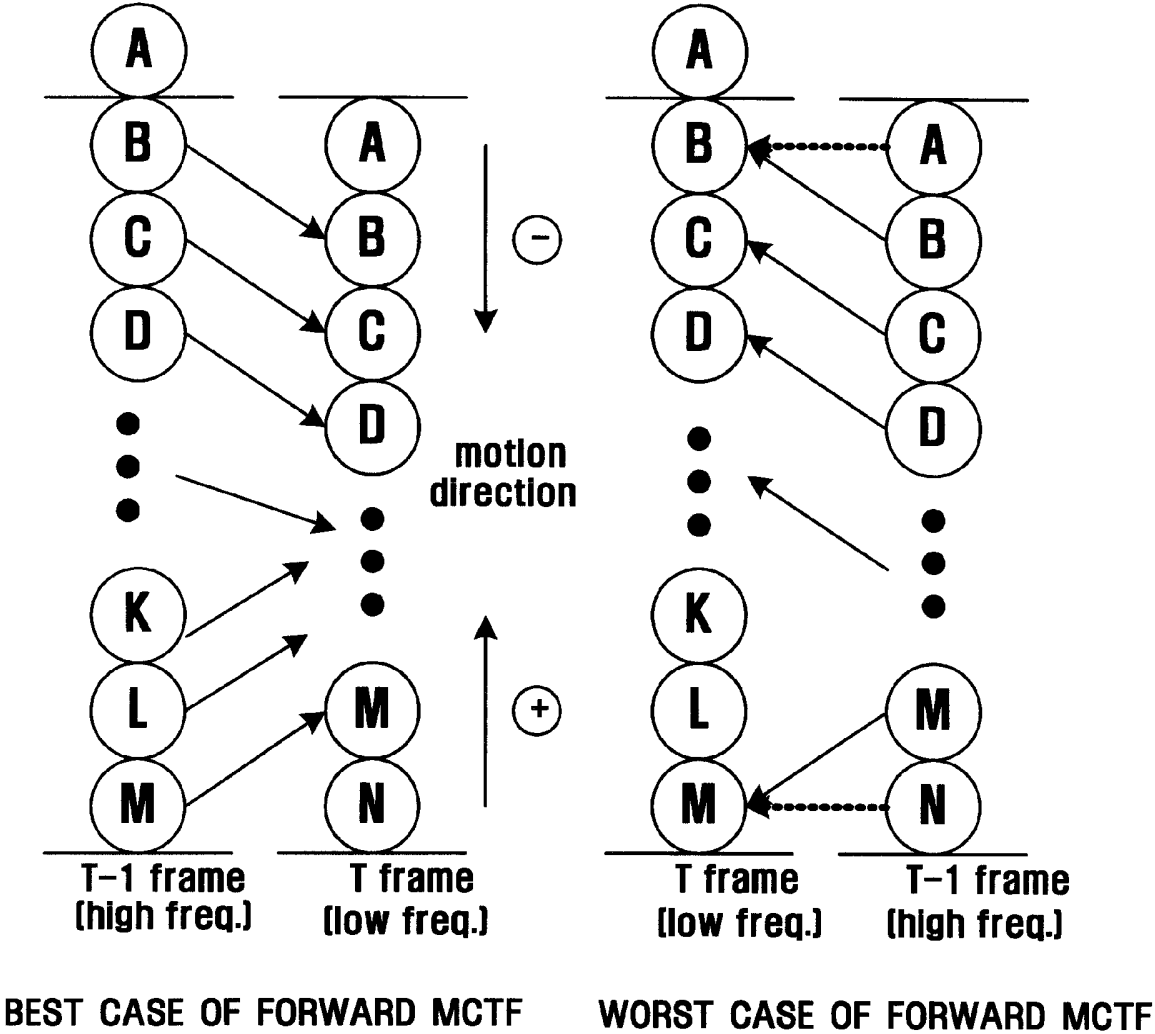
[Fig. 2]



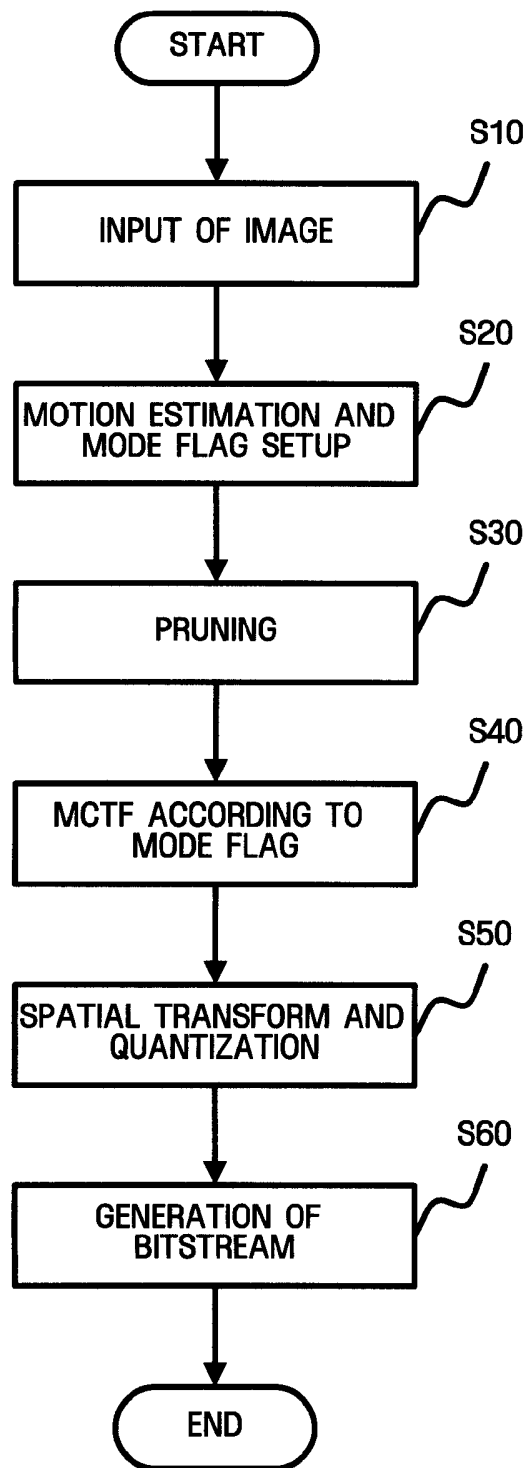
[Fig. 3]



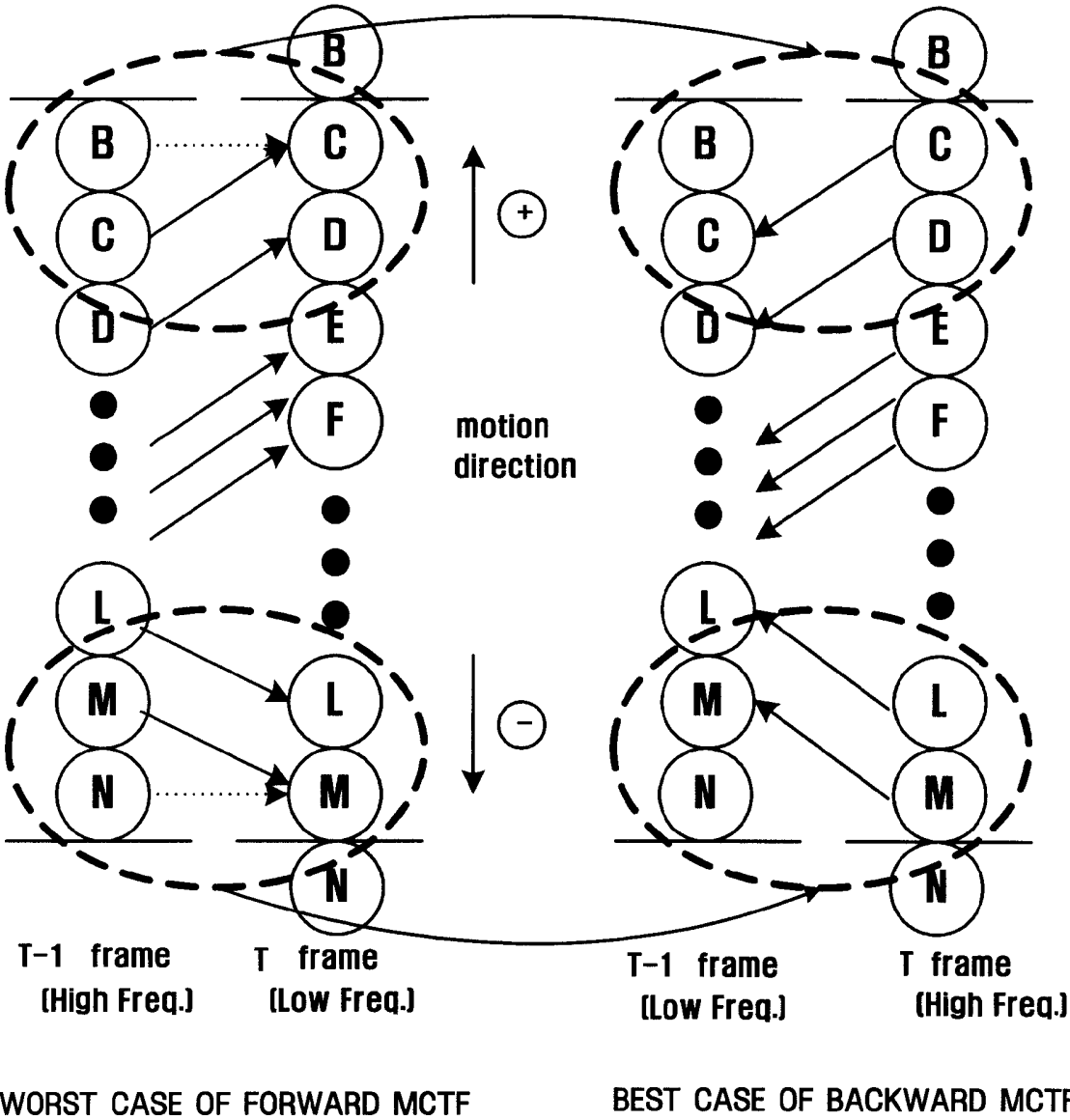
[Fig. 4]



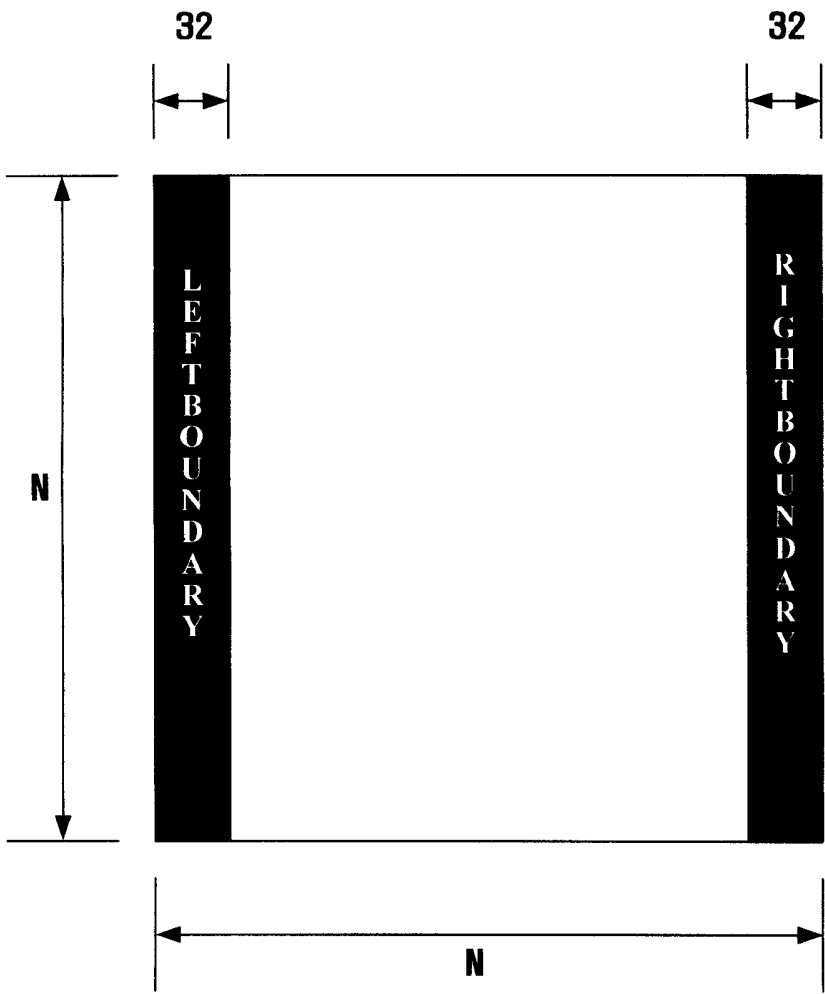
[Fig. 5]



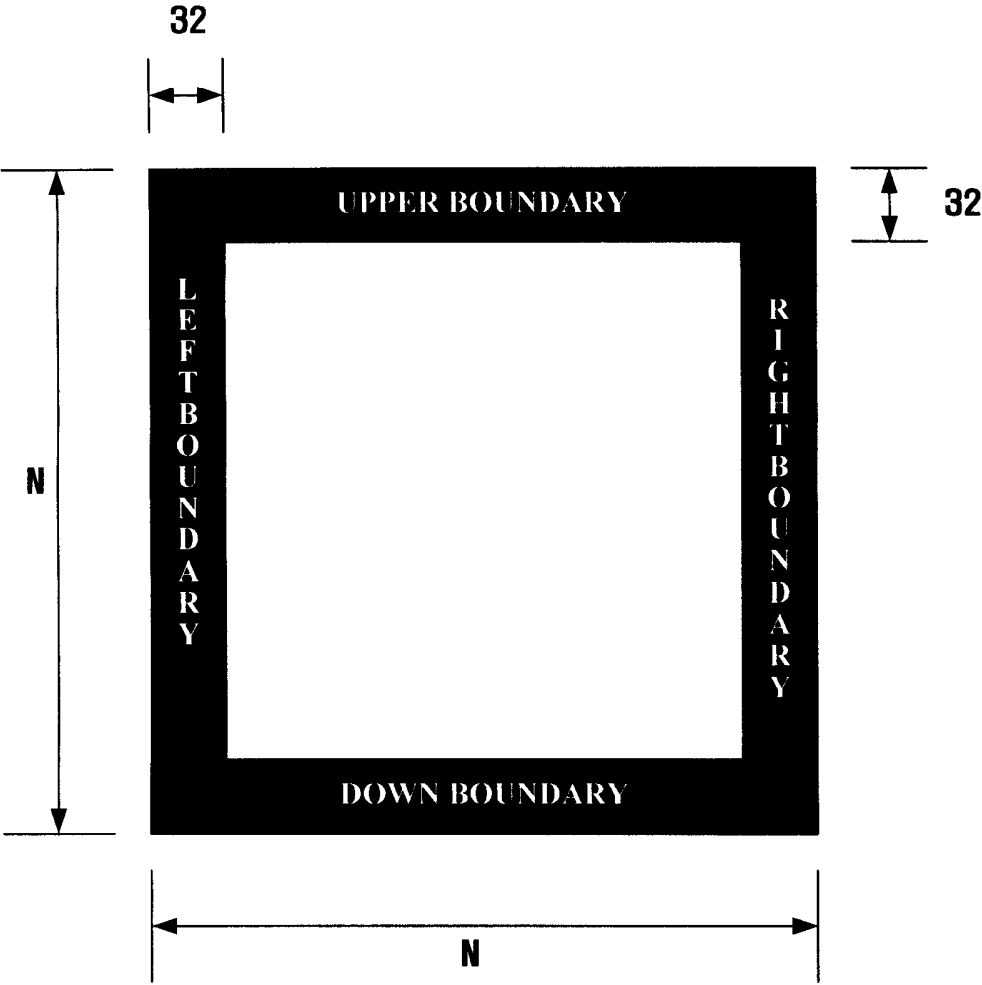
[Fig. 6]



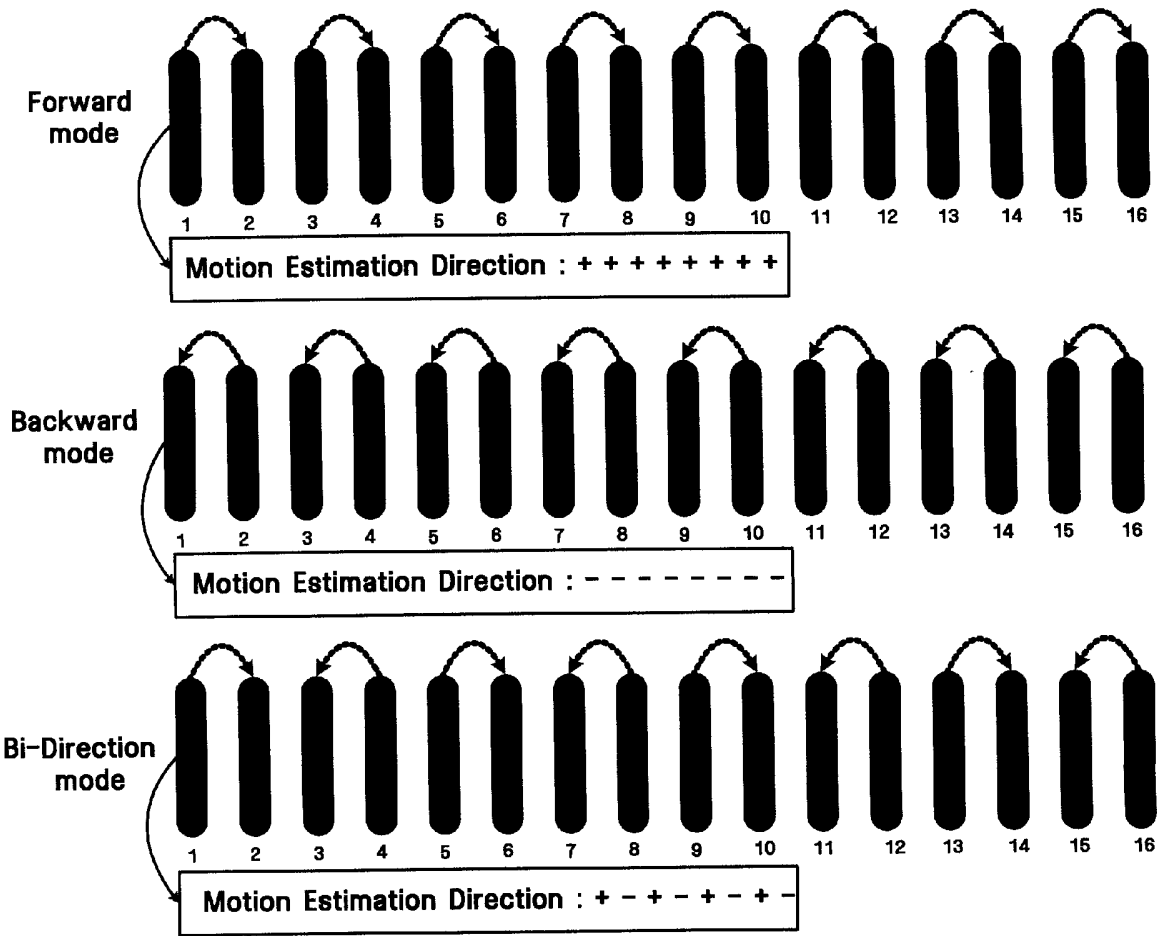
[Fig. 7]



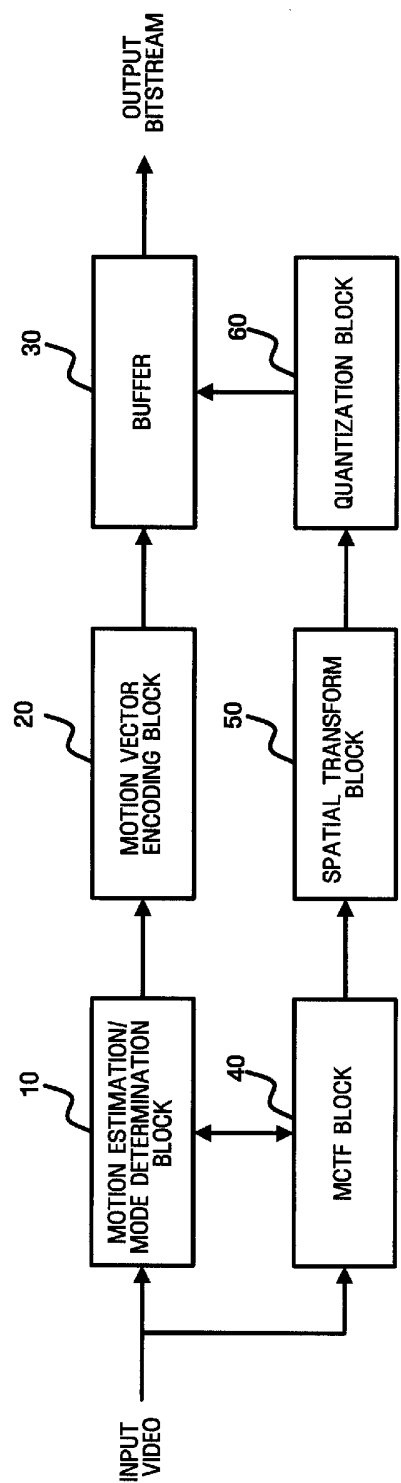
[Fig. 8]



[Fig. 9]



[Fig. 10]



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/KR2004/002050

## A. CLASSIFICATION OF SUBJECT MATTER

**IPC7 H04N 7/32**

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
Korean Patents and Applications for Inventions since 1975

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
NPS: "wavelet video coding, interframe, motion-compensated temporal filtering, motion, direction, trajectory"

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                               | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| A         | Shih-Ta Hsiang and John W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank", Signal Processing: Image Communication, Volume 16, Issue 8, Pages 705-724, May 2001.<br>See the whole document                    | 1-20                  |
| A         | V. Botreau, M. Benetiere, B. Felts, B. Pesquet-Popescu, "A fully scalable 3D subband video codec", International Conference on Image Processing, Volume: 2, Pages:1017 - 1020, 7-10 Oct. 2001.<br>See the whole document                                         | 1-20                  |
| A         | B. Pesquet-Popescu, M. Benetiere, "Three-dimensional lifting schemes for motion compensated video compression" IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume: 3, Pages:1793 - 1796, 7-11 May 2001.<br>See the whole document | 1-20                  |

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

19 NOVEMBER 2004 (19.11.2004)

Date of mailing of the international search report

**19 NOVEMBER 2004 (19.11.2004)**

Name and mailing address of the ISA/KR



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

Facsimile No. 82-42-472-7140

Authorized officer

KIM, Kyeoun Soo

Telephone No. 82-42-481-8174

