(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 July 2001 (12.07.2001)

PCT

(10) International Publication Number
WO 01/50737 A2

(51) International Patent Classification^{7}: H04N 5/14

(21) International Application Number: PCT/EP00/12864

(22) International Filing Date:
15 December 2000 (15.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/477,085 30 December 1999 (30.12.1999) US

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventors: MCGEE, Thomas; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). DIMITROVA, Nevenka; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: GROENENDAAL, Antonius, W., M.; Internationaal Octrooibureau B.V., Prof Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(national)*: CN, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:

— *Without international search report and to be republished upon receipt of that report.*

WO 01/50737 A2

(54) Title: METHOD AND APPARATUS FOR REDUCING FALSE POSITIVES IN CUT DETECTION

(57) **Abstract**: A video indexing method and device for selecting keyframes from each detected scene in the video. The method and device determines whether a scene change has occurred between two frames of video or whether the change between the two frames is merely a uniform change in luminance values.

Method and apparatus for reducing false positives in cut detection

BACKGROUND OF THE INVENTION

The present invention is in general related to an apparatus that detects significant scenes of a source video and selects representative keyframes therefrom. The present invention in particular relates to determining whether a detected scene change is really a scene change or merely a uniform change in intensity of the image such as when camera flashes occur during a news broadcast etc.

Users will often record home videos or record television programs, movies, concerts, sports events, etc. on a tape for later or repeated viewing. Often, a video will have varied content or be of great length. However, a user may not write down what is on a recorded tape and may not remember what she recorded on a tape or where on a tape particular scenes, movies, events are recorded. Thus, a user may have to sit and view an entire tape to remember what is on the tape.

Video content analysis uses automatic and semi-automatic methods to extract information that describes contents of the recorded material. Video content indexing and analysis extracts structure and meaning from visual cues in the video. Generally, a video clip is taken from a TV program or a home video by selecting frames which reflect the different scenes in a video.

In a scene change detection system described by Hongjiang Zhang, Chien Yong Low and Stephen W. Smoliar in "Video Parsing and Browsing Using Compressed Data", published in Multimedia Tools and Applications in 1995, (pp. 89-111) corresponding blocks between two video frames are compared and the difference between all blocks totaled over the complete video frame without separating out block types. A scene change is detected if a certain number of blocks have changed between the two frames. The system of Zhang, however, may produce skewed results if the differences between respective blocks of two frames are approximately the same with respect to color or intensity. In such a case the system may detect a scene change when in fact the change is only due to camera flashes which occur during a news broadcast.

SUMMARY OF THE PRESENT INVENTION

A system is desired which will create a visual index for a video source that was previously recorded or while being recorded, which is useable and more accurate in selecting significant keyframes, while providing a useable amount of information for a user. This system will detect scene changes and select a key frame from each scene but ignore the detection of scene changes and the selection of key frames where the changes between two frames result from only a substantially uniform change in luminance of substantially all blocks or macroblocks within the frame.

It is an object of the invention to compare two frames of a video to detect a scene change, but if the only difference between the two frames is a substantially uniform change in luminance then the invention determines that a scene change was not detected.

It is another object of the invention to compare the DC coefficients of corresponding blocks of two frames. If the change in the DC coefficients is approximately the same for substantially all blocks within the frames then it is determined that a scene change did not occur and another key frame is not selected.

For a better understanding of the invention, its operating advantages and specific objects attained by its use, reference should be had to the accompanying drawings and descriptive matter in which there are illustrated and described the preferred embodiments of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

For a better understanding of the reference is a made to the following drawings.

Figure 1 illustrates a video archival process;

Figures 2A and 2B are block diagrams of devices used in creating a visual index in accordance with a preferred embodiment of the invention ;

Figure 3 illustrates a frame, a macroblock, and several blocks;

Figure 4 illustrates several DCT coefficients of a block;

Figure 5 illustrates a macroblock and several blocks with DCT coefficients; and

Figures 6 illustrates a stream of video where a change in luminance has occurred.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Two phases exist in the video content indexing process: archival and retrieval. During the archival process, video content is analyzed during a video analysis process and a visual index is created. In the video analysis process, automatic significant scene detection, uniform luminance change detection and keyframe selection occur. Significant scene detection is a process of identifying scene changes, i.e., "cuts" (video cut detection or segmentation detection) and identifying static scenes (static scene detection). For each scene detected, a particular representative frame called a keyframe is extracted. Therefore it is important that correct identification of scene changes occurs otherwise there will be too many keyframes chosen for a single scene or not enough key frames chosen for multiple scene changes. Uniform luminance detection is the process of identifying a change in luminance between two frames and is explained in further detail below. (Reference is to a source tape although clearly, the source video may be from a file, disk, DVD, other storage means or directly from a transmission source (e.g., while recording a home video).)

A video archival process is shown in Figure 1 for a source tape with previously recorded source video, which may include audio and/or text, although a similar process may be followed for other storage devices with previously saved visual information, such as an MPEG file. In this process, a visual index is created based on the source video. A second process, for a source tape on which a user intends to record, creates a visual index simultaneously with the recording.

Figure 1 illustrates an example of the first process (for previously recorded source tape) for a videotape. In step 101, the source video is rewound, if required, by a playback/recording device such as a VCR. In step 102, the source video is played back. Signals from the source video are received by a television, a VCR or other processing device. In step 103, a media processor in the processing device or an external processor, receives the video signals and formats the video signals into frames representing pixel data (frame grabbing).

In step 104, a host processor separates each frame into blocks, and transforms the blocks and their associated data to create DCT (discrete cosine transform) coefficients; performs significant scene detection, uniform change in luminance detection and keyframe selection; and builds and stores keyframes as a data structure in a memory, disk or other storage medium. In step 105, the source tape is rewound to its beginning and in step 106, the source tape is set to record information. In step 107, the data structure is transferred from the memory to the source tape, creating the visual index. The tape may then be rewound to view the visual index. (Instead of a tape, any storage medium can be used or the index could be stored and/or created at the server.)

The above process is slightly altered when a user wishes to create a visual index on a tape while recording. Instead of steps 101 and 102, as shown in step 112 of Figure 1, the frame grabbing process of step 103 occurs as the video (film, etc.) is being recorded.

Steps 103 and 104 are more specifically illustrated in Figures 2A and 2B. Video exists either in analog (continuous data) or digital (discrete data) form. The present example operates in the digital domain and thus uses digital form for processing. The source video or video signal is a series of individual images or video frames displayed at a rate high enough (in this example 30 frames per second) so the displayed sequence of images appears as a continuous picture stream. These video frames may be uncompressed (NTSC or raw video) or compressed data in a format such as MPEG, MPEG 2, MPEG 4, Motion JPEG or such.

The information in an uncompressed video is first segmented into frames in a media processor 202, using a frame grabbing technique 204 such as present on the Intel Smart Video Recorder III. Although other frame sizes are available, in this example shown in Figure 3, a frame 302 represents one television, video, or other visual image and includes 352 x 240 pixels.

The frames 302 are each broken into blocks 304 of, in this example, 8 x 8 pixels in the host processor 210 (Figure 2A). Using these blocks 304 and a popular broadcast standard, CCIR-601, a macroblock creator 206 (Figure 2A) creates luminance blocks and sub samples the color information to create chrominance blocks. The luminance and chrominance blocks form a macroblock 308. In this example, 4:2:0 is being used although other formats such as 4:1:1 and 4:2:2 could easily be used by one skilled in the art. In 4:2:0, a macroblock 308 has six blocks, four luminance, Y1, Y2, Y3, and Y4; and two chrominance Cr and Cb, each block within a macroblock being 8x8 pixels.

The video signal may also represent a compressed image using a compression standard such as Motion JPEG (Joint Photographic Experts Group) and MPEG (Motion Pictures Experts Group). If the signal is an MPEG or other compressed signal, as shown in Figure 2B the MPEG signal is broken into frames using a frame or bitstream parsing technique by a frame parser 205. The frames are then sent to an entropy decoder 214 in the media processor 203 and to a table specifier 216. The entropy decoder 214 decodes the

MPEG signal using data from the table specifier 216, using, for example, Huffman decoding, or another decoding technique.

The decoded signal is next supplied to a dequantizer 218 which dequantizes the decoded signal using data from the table specifier 216. Although shown as occurring in the media processor 203, these steps (steps 214-218) may occur in either the media processor 203, host processor 211 or even another external device depending upon the devices used.

Alternatively, if a system has encoding capability (in the media processor, for example) that allows access at different stages of the processing, the DCT coefficients could be delivered directly to the host processor. In all these approaches, processing may be performed in real time.

In step 104 of Figure 1, the host processor 210, which may be, for example, an Intel© Pentium™ chip or other processor or multiprocessor, a Philips© Trimedia™ chip or any other multimedia processor; a computer; an enhanced VCR, record/playback device, or television; or any other processor, performs significant scene detection, key frame selection, and building and storing a data structure in an index memory, such as, for example, a hard disk, file, tape, DVD, or other storage medium.

Significant Scene Detection/Uniform change in Luminance Detection: For automatic significant scene detection, the present invention attempts to detect when a scene of a video has changed or a static scene has occurred. A scene may represent one or more related images. In significant scene detection, two consecutive frames are compared and, if the frames are determined to be significantly different, a scene change is determined to have occurred between the two frames; and if determined to be significantly alike, processing is performed to determine if a static scene has occurred. In uniform luminance change detection, if a scene change has been detected then the luminance values of the two frames are compared and if a uniform change in luminance is the only major change between the two frames then it is determined that a scene change has not occurred between the two frames.

Fig. 2A shows an example of a host processor 210 with luminance change detector 240. The DCT blocks are provided by macroblock creator 206 and DCT transformer 220. Fig. 2B shows an example of host processor 211 with significant scene detector 230 and luminance charge detector 240. The DCT blocks are provided by dequantizer 218. The significant scene processor 230 detects scene changes between two frames and then the luminance detector 240 determines whether in fact a scene change has occurred or whether the differences between the two f ames are due to a uniform change in luminance. If a scene change occurred a keyframe is se ected and provided to frame memory 234 and then provided to the index memory 260. If a uniform change in luminance is detected, another keyframe is not selected from this same scene.

The present invention addresses the concern where two frames are compared and there is a substantial difference detected between two frames. There are many reasons why this substantial difference may not be due to a scene change. For instance, the video may be a news broadcast where the videographer is taping a press briefing. During this press briefing many camera flashes flash which cause the luminance between two frames to change. Instead of this being detected as a scene change and another keyframe chosen, the present invention detects the uniform change in luminance and treats it as an image from the same scene. Similarly, if the lights are turned on in a room, or the lights flash in a disco, a scene change should not be detected as the difference between the two frames is merely a uniform change in luminance.

The present method and device uses comparisons of DCT (Discrete Cosine Transform) coefficients to detect uniform changes in luminance, but other methods can also be used. First, each received frame 302 is processed individually in the host processor 210 to create 8 x 8 blocks 440. The host processor 210 processes each 8 x 8 block which contains spatial information, using a discrete cosine transformer 220 to extract DCT coefficients and create the macroblock 308.

When the video signal received is in compressed video format such as MPEG, the DCT coefficients may be extracted after dequantization and need not be processed by a discrete cosine transformer. Additionally, as previously discussed, DCT coefficients may be automatically extracted depending upon the devices used.

The DCT transformer provides each of the blocks 440 (Figure 4), Y1, Y2, Y3, Y4, Cr and Cb with DCT coefficient values. According to this standard, the uppermost left hand corner of each block contains DC information (DC value) and the remaining DCT coefficients contain AC information (AC values). The AC values increase in frequency in a zig-zag order from the right of the DC value, to the DCT coefficient just beneath the DC value, as partially shown in Figure 4. The Y values are the luminance values.

In the method to follow, processing is limited to detecting the change in DC values between corresponding blocks of two frames to more quickly produce results and limit processing without a significant loss in efficiency; however, clearly one skilled in the art could compare the difference in the luminance of corresponding macroblocks or any other method which detects a change in luminance.

The method and device in accordance with a preferred embodiment of the instant invention compares the DC values of respective blocks of two frames to determine whether a substantially uniform change in luminance has occurred.

Assume n is the number of blocks within a frame. Assume F_1 is the first frame and F_2 is the second frame where $F_1[i]$ is the ith block in the first frame and $F_2[i]$ is the ith block in the second frame. Also assume diffmin is first set to some high value, such as 1,000,000 and diffmax is first set to some low value, such as -9,000,000 and then a comparison is made as follows:

```
For i=0 to n
        Diff=ABS(F₁[i]- F₂[i])
        If diff < diffmin then diffmin = diff;
        If diff > diffmax then diffmax = diff;
        i=i+1
end
```

If (diffmax - diffmin) < threshold then a scene change has not occurred.

The above computation computes the absolute value of the difference between the DC coefficient of each block in the first frame with its respective DC coefficient in the second frame. This difference is then compared to diffmin and diffmax to find the minimum difference and the maximum difference between corresponding DC coefficients between the two frames. If the difference between the maximum difference (diffmax) and minimum difference (diffmin) is less than a certain threshold then all DC values have changed by approximately the same amount indicating a change in luminance. In a preferred embodiment of the invention the threshold value is chosen anywhere between 0 and 10% of the final diffmax value, but depending on the application this threshold may vary.

If it is determined that a uniform change in luminance has occurred between two frames then a key frame is not chosen from both frame sequences. It should be noted that other methods of detecting changes in luminance can be used such as using histograms and wavelets etc. and the invention is not limited to the embodiment described above. The ratios of the luminance changes compared to the ratios of the chrominance changes could be used to determine the change in luminance, or any other formula for determining luminance change.

Figs. 6A-D illustrates two scenarios where a scene change is detected but the difference between the two frames is merely a change in luminance. Fig. 6A is an example of an image during a camera flash. Fig. 6B shows this same image after the camera flash. Similarly a top view of a disco scene is shown in Fig. 6C during a time period when the lights are off. Fig. 6D shows this same scene when the lights are on.

The present invention is shown using DCT coefficients; however, one may instead use representative values such as wavelet coefficients, histograms etc. or a function which operates on a sub-area of the image to give representative values for that sub-area. In addition, the present invention has been described with reference to a video indexing system, however it pertains in general to detecting a uniform change in luminance between two frames and therefore can be used as a search device to detect scenes where there are camera flashes, or alternatively as an archival method to pick representative frames.

While the invention has been described in connection with preferred embodiments, it will be understood that modifications thereof within the principles outlined above will be evident to those skilled in the art and thus, the invention is not limited to the preferred embodiments but is intended to encompass such modifications.

CLAIMS:

1. A system for detecting a uniform change in luminance between two frames, comprising:

a receiver (210, 202) which receives source video having frames comprised of luminance values; and

a comparator (230, 240) which compares the luminance values of the first frame with respective luminance values of the second frame and detects whether substantially all luminance values in the first frame change by substantially the same amount in the second frame.

2. The system as claimed in claim 1, where the luminance values are in the form of DCT coefficients.

3. The system as claimed in claim 1, where the luminance values are in the form of wavelet coefficients.

4. The system as claimed in claim 1, where the luminance values are in the form of histogram values.

5. The system as claimed in claim 1, further including a comparator (230,240) which computes a maximum difference (diffmax) between all corresponding luminance values in the first and second frames and a minimum difference (diffmin) between all corresponding luminance values of the first and second frames and then compares ABS(diffmax - diffmin) to a threshold value to determine if a uniform change in luminance has occurred.

6. The system as claimed in Claim 5, where the threshold value is approximately zero to ten percent of diffmax.

7. A video indexing s·/stem for detecting scene changes and selecting key frames for each scene, comprising:

a scene ch.ange detector (230) which detects a scene change between two frames of video; and

a uniform luminance change detector (240) which, if a scene change has been detected, receives the two frames of video and determines whether the difference between the two frames is substantially only a uniform change in luminance.

8. The system as claimed in claim 7, where the luminance values are in the form of DCT coefficients.

9. The system as claimed in claim 7, where the luminance values are in the form of wavelet coefficients.

10. The system as claimed in claim 7, where the luminance values are in the form of histogram values.

11. The system as claimed in claim 7, further including a comparator which computes a maximum difference (diffmax) between all corresponding luminance values in the first and second frames and a minimum difference (diffmin) between all corresponding luminance values in the first and second frames and then compares ABS(diffmax - diffmin) to a threshold value to determine if a uniform change in luminance has occurred.

12. The system as claimed in Claim 11, where the threshold value is zero to ten percent of diffmax.

13. A method of detecting false positive scene change detections, comprising:

receiving at least two frames of video which have been detected as having a scene change occur from a first frame to a second frame, each frame having luminance values,

comparing the luminance values of the first frame with corresponding luminance values of the second frame; and computing whether substantially all luminance values in the first frame change by substantially the same amount in the second frame, and if so, determining that a false positive scene change detection has occurred between the two frames.

14. The system as claimed in claim 13, where the luminance values are in the form of DCT coefficients.

15. The system as claimed in claim 13, where the luminance values are in the form of wavelet coefficients.

16. The system as claimed in claim 13, where the luminance values are in the form of histogram values.

17. The system as claimed in claim 13, further including a comparator which computes a maximum difference (diffmax) between all corresponding luminance values in the first and second frames and a minimum difference(diffmin) between all corresponding luminance values of the first and second frames and then compares ABS(diffmax - diffmin) to a threshold value to determine if a uniform change in luminance has occurred.

18. The system as claimed in Claim 17, where the threshold value is zero to ten percent of diffmax.

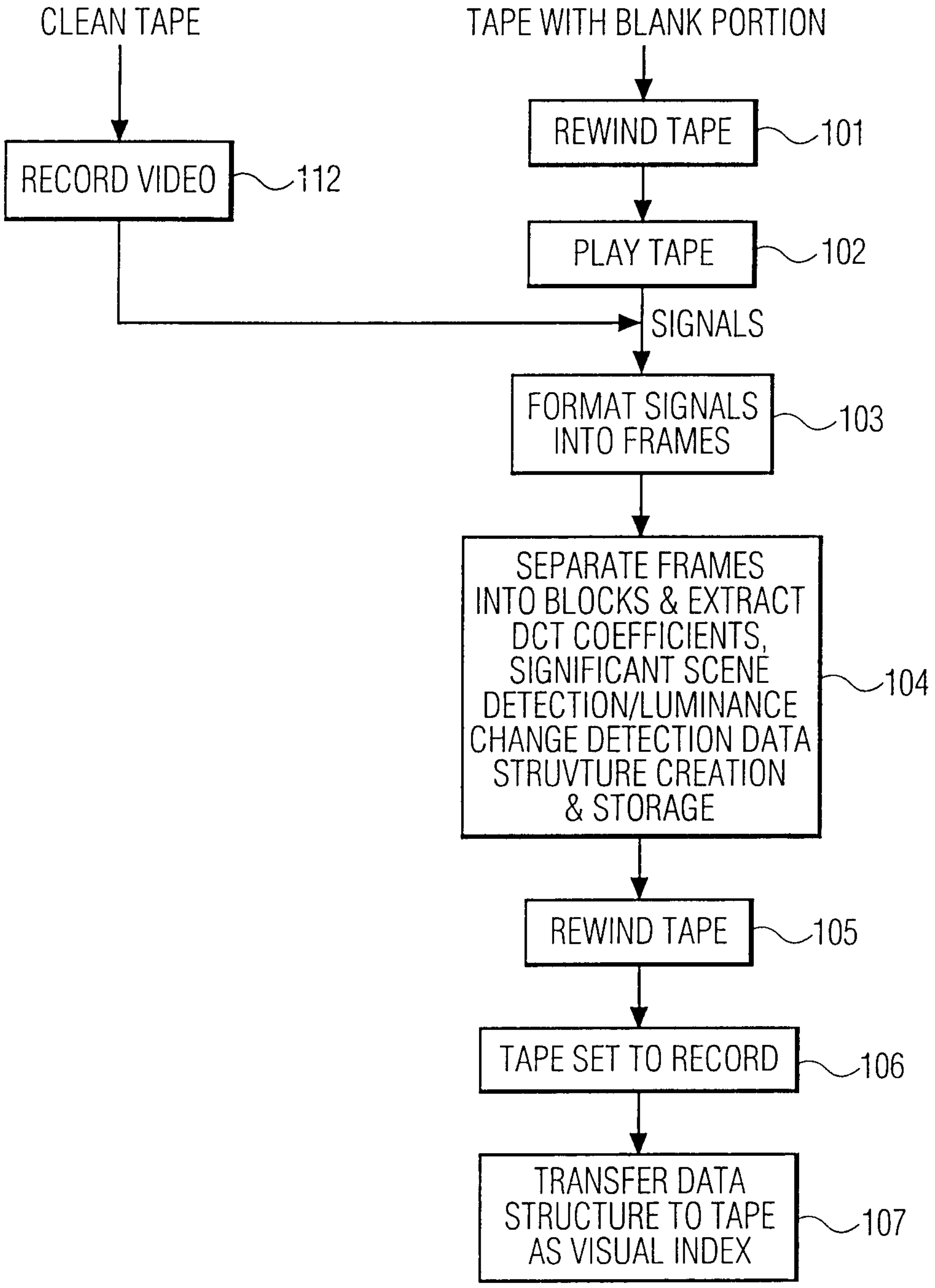
CLEAN TAPE
TAPE WITH BLANK PORTION
REWIND TAPE
101
RECORD VIDEO
112
PLAY TAPE
102
SIGNALS
FORMAT SIGNALS INTO FRAMES
103
SEPARATE FRAMES INTO BLOCKS & EXTRACT DCT COEFFICIENTS, SIGNIFICANT SCENE DETECTION/LUMINANCE CHANGE DETECTION DATA STRUVTURE CREATION & STORAGE
104
REWIND TAPE
105
TAPE SET TO RECORD
106
TRANSFER DATA STRUCTURE TO TAPE AS VISUAL INDEX
107

FIG. 1

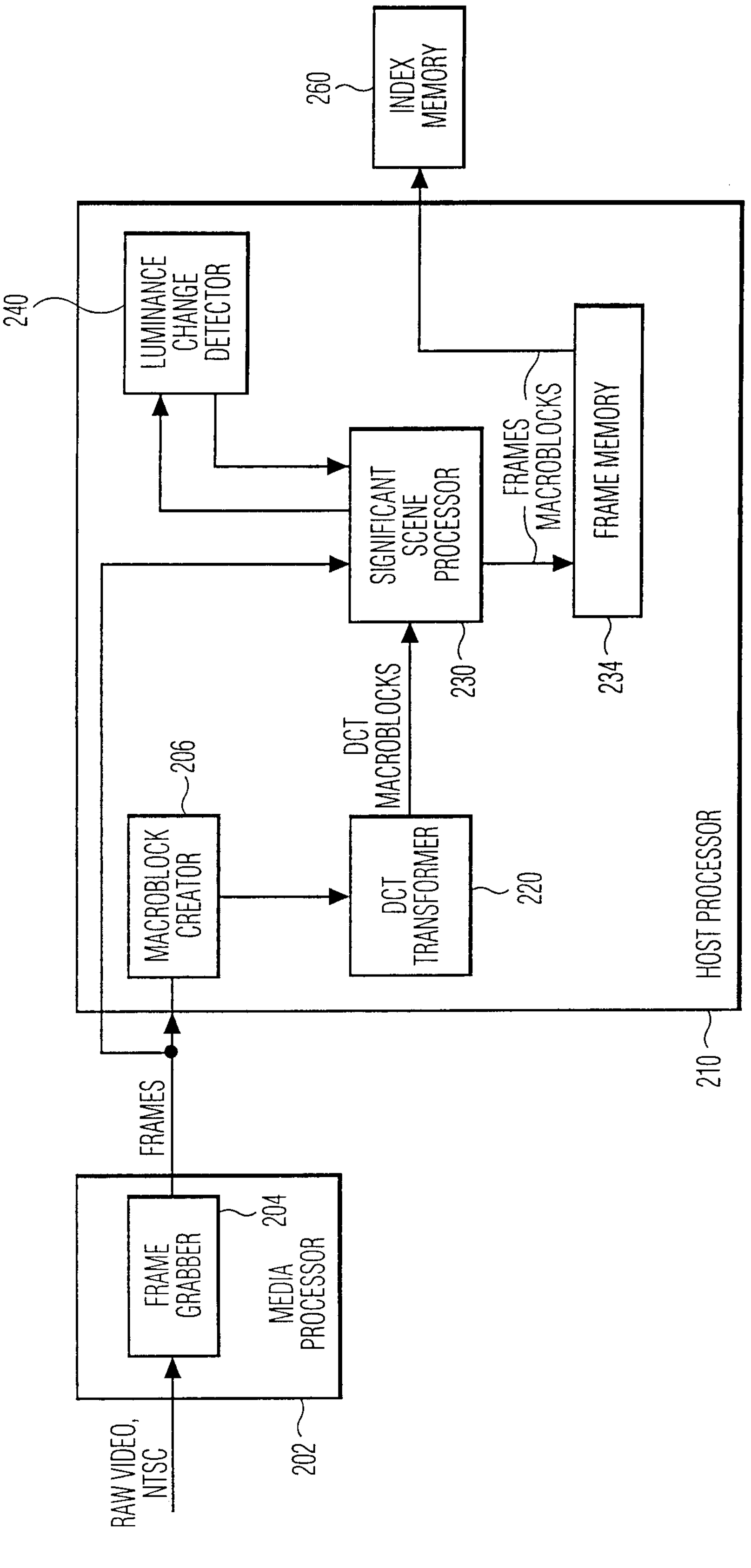
RAW VIDEO, NTSC
202
FRAME GRABBER
204
MEDIA PROCESSOR
FRAMES
210
MACROBLOCK CREATOR
206
DCT TRANSFORMER
220
HOST PROCESSOR
DCT MACROBLOCKS
SIGNIFICANT SCENE PROCESSOR
230
LUMINANCE CHANGE DETECTOR
240
FRAMES MACROBLOCKS
FRAME MEMORY
234
INDEX MEMORY
260

FIG. 2A

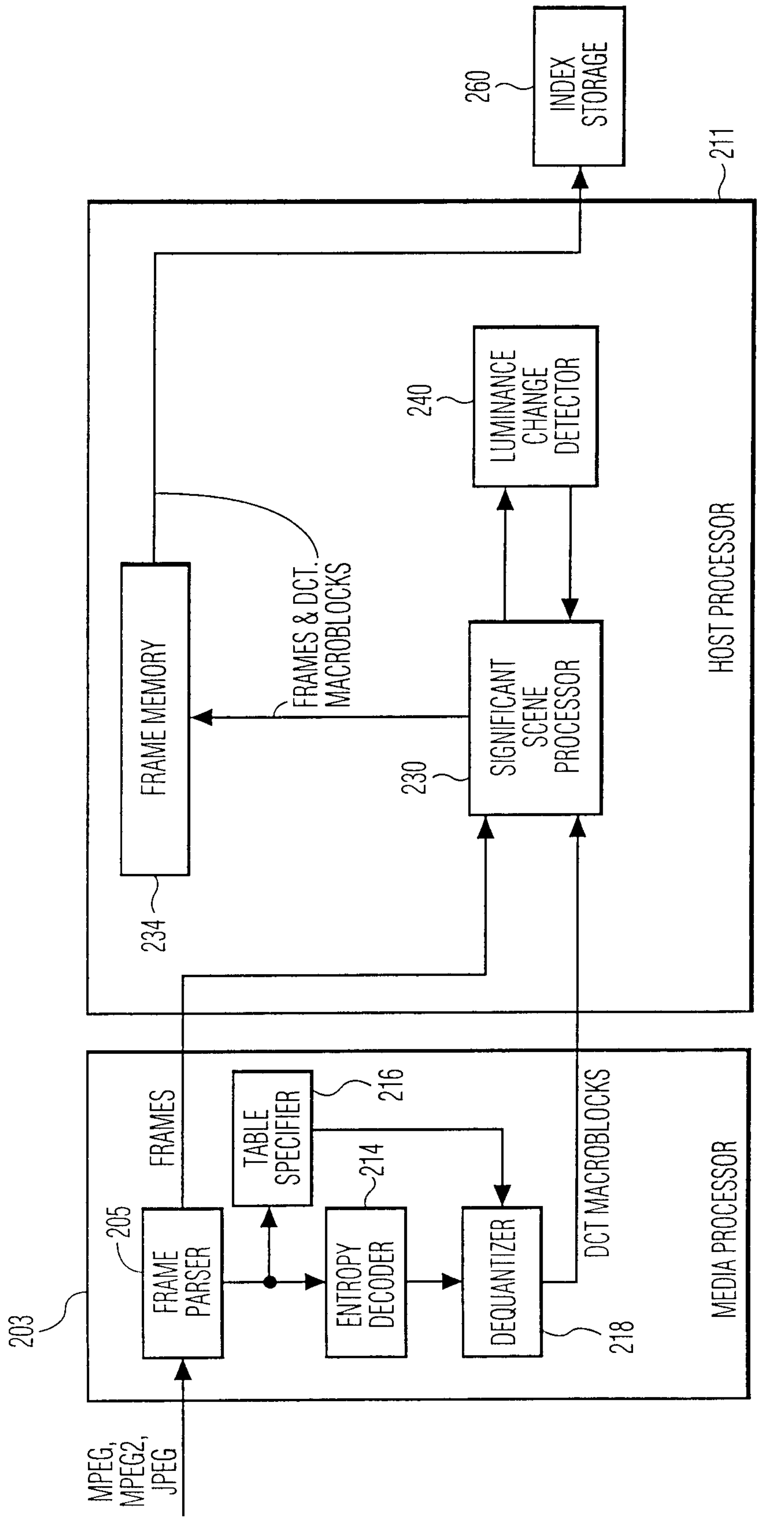
MPEG,
MPEG2,
JPEG
203
205
FRAME PARSER
TABLE SPECIFIER
216
214
ENTROPY DECODER
DEQUANTIZER
218
FRAMES
DCT MACROBLOCKS
MEDIA PROCESSOR
234
FRAME MEMORY
FRAMES & DCT. MACROBLOCKS
230
SIGNIFICANT SCENE PROCESSOR
240
LUMINANCE CHANGE DETECTOR
HOST PROCESSOR
211
260
INDEX STORAGE

FIG. 2B

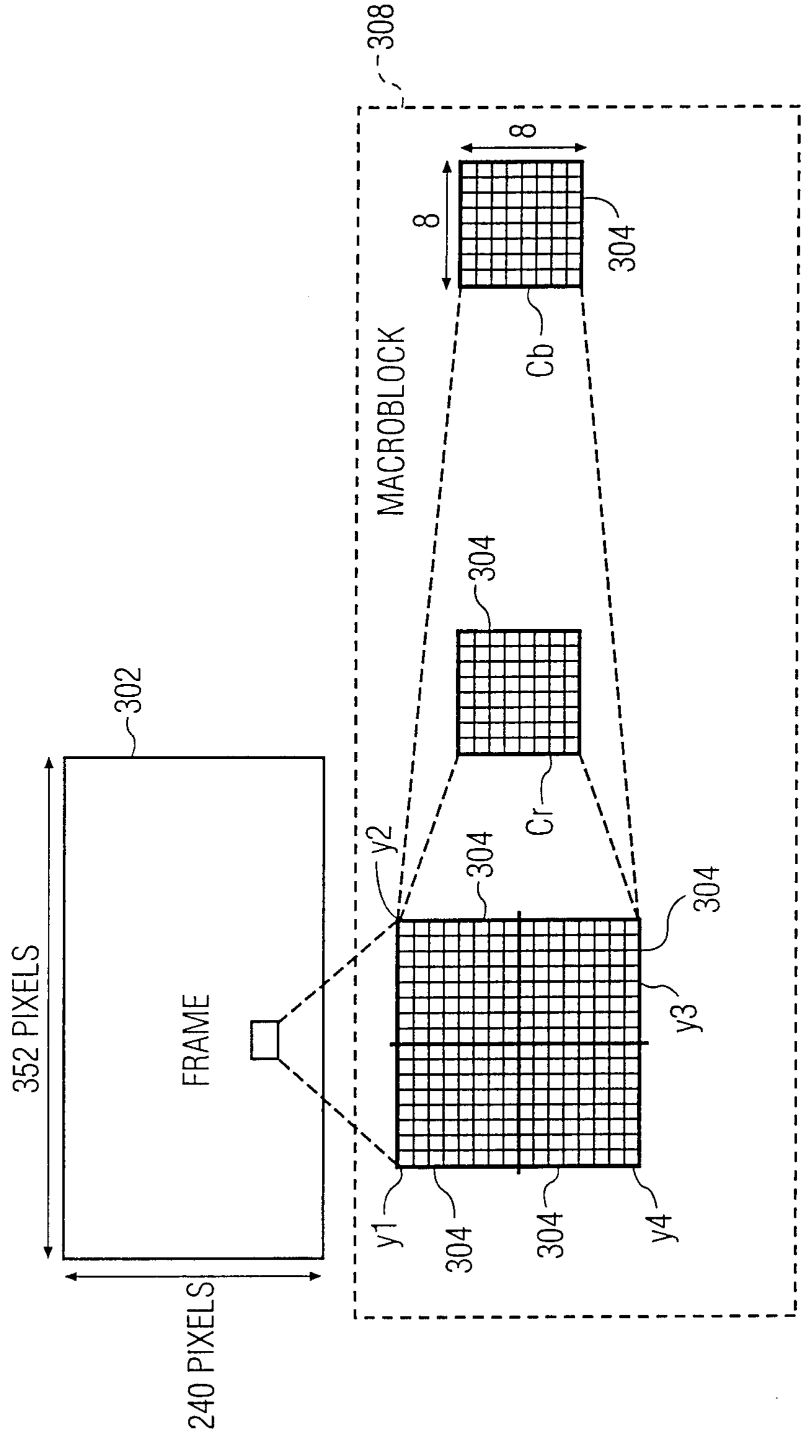
302
FRAME
352 PIXELS
240 PIXELS
308
MACROBLOCK
8
8
304
Cb
304
Cr
y2
304
304
y3
y1
304
304
y4

FIG. 3

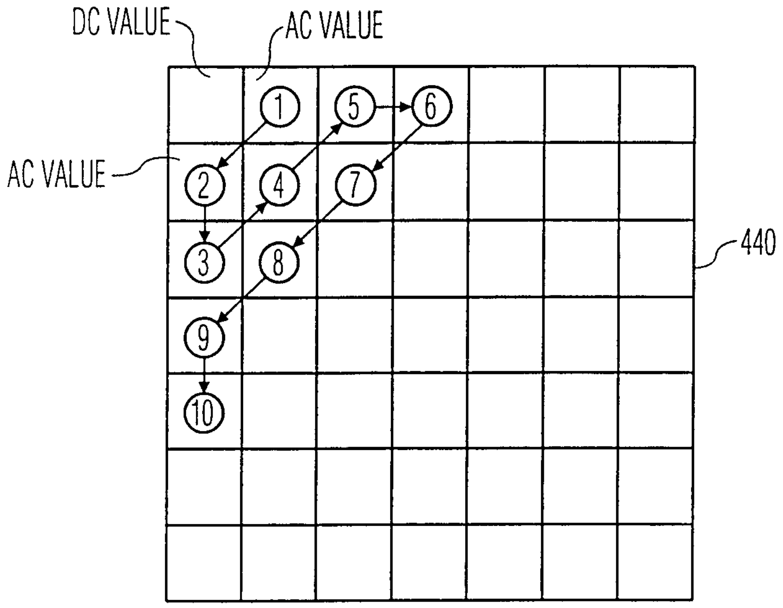
DC VALUE
AC VALUE
AC VALUE
1
2
3
4
5
6
7
8
9
10
440

FIG. 4

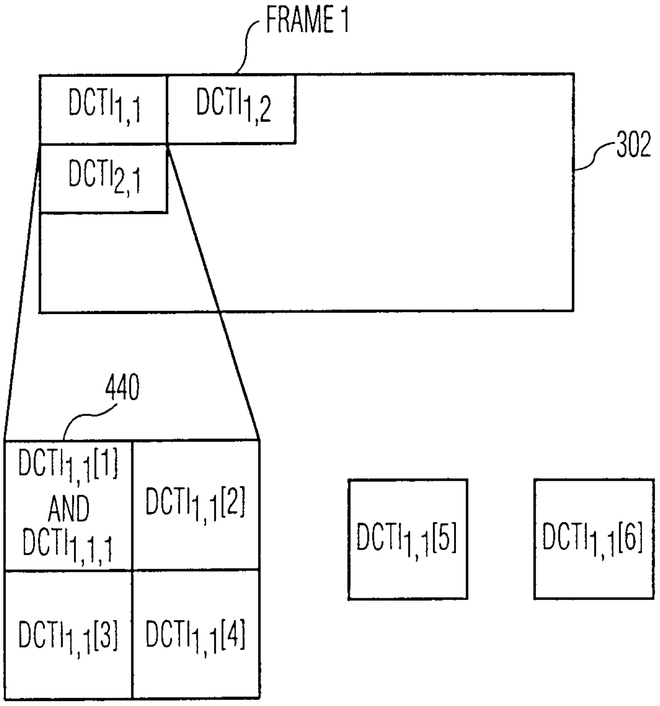
FRAME 1
$DCTI_{1,1}$
$DCTI_{1,2}$
$DCTI_{2,1}$
302
440
$DCTI_{1,1}[1]$ AND $DCTI_{1,1,1}$
$DCTI_{1,1}[2]$
$DCTI_{1,1}[3]$
$DCTI_{1,1}[4]$
$DCTI_{1,1}[5]$
$DCTI_{1,1}[6]$

FIG. 5

FIG. 6A

FIG. 6B

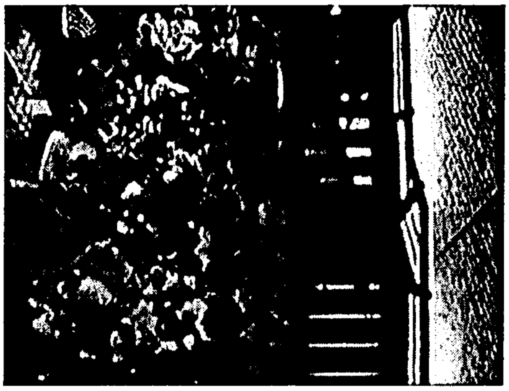

FIG. 6C

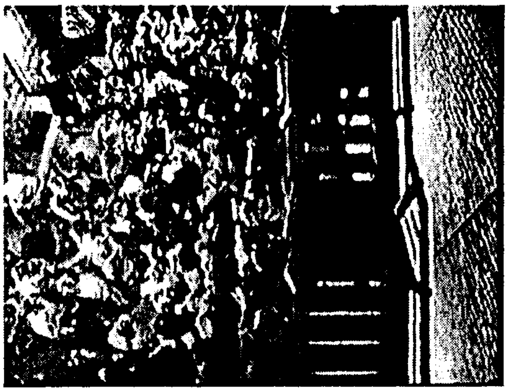

FIG. 6D