



STATE OF NEW YORK

COUNTY OF NEW YORK

)
)
)
)

ss

CERTIFICATION

This is to certify that the attached translation is to the best of my knowledge and belief a true and accurate translation from Korean into English of the attached Patent Publication Number 2007-0017068 dated February 08, 2007.

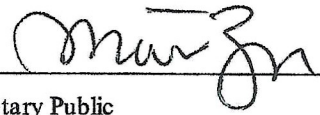


Michael Stypa
Divergent Language Solutions, LLC

State of New York

County of New York

Subscribed to and sworn before me this 17 day of April, 2021,
by Michael Stypa.



Notary Public



(19) The Korean Intellectual Property Office (KR)
(12) Notice of Publication of Unexamined Patent Application (A)

(51) Int. Cl.
H04N 5/93 (2006.01)

(11) Publication Number: 10-2007-0017068
(43) Publication Date: February 08, 2007

(21) Application Number:	10-2006-0073770		
(22) Filing Date:	August 04, 2006		
Examination request date:	None		

(30) Priority Rights Claim	1020050071507	August 04, 2005	Republic of Korea (KR)
----------------------------	---------------	-----------------	------------------------

(71) Applicant:	Samsung Electronics Co., Ltd. 416 Maetan-dong, Yeongtong-gu, Suwon-si, Gyeonggi-do
(72) Inventors:	Jeong-bae Kim 102-304 Samik Apartment, Pungdeokcheon-dong, Yongin-si, Gyeonggi-do Doo-seon Hwang 103-606 Jinheung Apartment, Susaek-dong, Eunpyeong-gu, Seoul Ji-yeon Kim 5-102 Hanyang Apartment, Myeongil-dong, Gangdong-gu, Seoul
(74) Agent	Y.P. Lee, Mock and Partners

Total Number of claims: Total 21 claims

(54) Apparatus and method for partitioning videos according to topic

Abstract

An apparatus and method for partitioning a video according to topic are disclosed. The apparatus for partitioning the video according to topic consists of a start shot determining unit that detects a plurality of key frames using character information in a video sequence composed of a plurality of frames and determines the detected key frames as the start shot of each subject, and a topic list generating unit that generates a topic list using the starting shot of each topic.

Representative Drawing

FIG. 2

Claims

Claim 1

An apparatus for partitioning a video according to topic, comprising a start shot determining unit that detects a plurality of key frames using character information in a video sequence composed of a plurality of frames and determines the detected key frames as the starting shot of each subject, and

a topic list generating unit that generates a topic list using the starting shot of each topic.

Claim 2

The apparatus for partitioning a video according to topic according to claim 1, wherein the starting shot determining unit detects the key frames using clothing information of a key character comprised in the video sequence.

Claim 3

The apparatus for partitioning a video according to topic according to claim 1, wherein the topic list generating unit composes frames existing between the starting shot of the current topic and the starting shot of the next topic as episodes of the current topic, and adds them to each starting shot of the topic list.

Claim 4

The apparatus for partitioning a video according to topic according to claim 1, wherein the starting shot determining unit consists of:
a preprocessing unit that detects scene transitions among the frames comprised in the video sequence, determines frames belonging to each scene, and calculates the number of main characters appearing in the video sequence;
a face detecting unit that detects faces from frames belonging to each scene determined by the preprocessing unit and determines frames in which faces are detected; and
a key frame determining unit that determines key frames by clustering the frames in which the face is detected according to the main character, considering the number of main characters.

Claim 5

The apparatus for partitioning a video according to topic according to claim 4, wherein the preprocessing unit detects the scene transformation using the calculated similarity between the current frame and the previous frame.

Claim 6

The apparatus for partitioning a video according to topic according to claim 4, wherein the preprocessing unit calculates the number of main characters from the EPG signal.

Claim 7

The apparatus for partitioning a video according to topic according to claim 4, wherein the preprocessing unit consists of:
a thumbnail image generating unit that generates a thumbnail image of an input frame,
a scene change detecting unit that detects a scene change using similarity of color histograms between the thumbnail images of adjacent frames, and

a main character number determining unit that determines the number of main characters from the interpreted EPG signal.

Claim 8

The apparatus for partitioning a video according to topic according to claim 4,
wherein the face detecting unit consists of a thumbnail image reconstructing unit that generates an integral image for the thumbnail image of the input frame and reconstructs the thumbnail image using the integral image,
a sub-window generating unit that generates a sub-window for the reconstructed thumbnail image, and
a classifying unit that classifies whether the subwindow comprises a face.

Claim 9

The apparatus for partitioning a video according to topic according to claim 4,
wherein the face detecting unit partitions the thumbnail image of the input frame into a plurality of areas comprising the area most likely to comprise a face, and sequentially provides the thumbnail image to the thumbnail image reconstructing unit, starting from the area most likely to comprise the face.

Claim 10

The apparatus for partitioning a video according to topic according to claim 4,
wherein the keyframe determining unit consists of a clothing information extracting unit that extracts clothing information from a frame in which a face is detected;
a character clustering unit that performs character clustering based on the clothing information extracted from the clothing information extracting unit; and
a key character determining unit that selects, considering the number of key characters above, a cluster corresponding to the main character among the plurality of clusters clustered in the character clustering unit and provides frames comprised in each selected cluster as keyframes for each subject.

Claim 11

The apparatus for partitioning a video according to topic according to claim 4,
wherein the clothing information is the color histogram of the clothing.

Claim 12

A method for partitioning videos according to topic, comprising:
a step of detecting a plurality of key frames using character information in a video sequence consisting of a plurality of frames, and determining the detected key frames as the starting shot of each subject;
and
a step of creating a topic list using the starting shot of each topic.

Claim 13

The method for partitioning videos according to topic according to claim 12, wherein the starting shot determining step involves detecting the key frames using clothing information of key characters comprised in the video sequence.

Claim 14

The method for partitioning videos according to topic according to claim 12, wherein, in the topic list generating step, frames existing between the starting shot of the current topic and the starting shot of the next topic are composed of episodes of the current topic and added to each starting shot of the topic list.

Claim 15

The method for partitioning videos according to topic according to claim 12, wherein the starting shot determination step consists of:

a step of detecting scene transitions among frames comprised in the video sequence, determining frames belonging to each scene, and calculating the number of main characters appearing in the video sequence;

a step of detecting a face from frames belonging to each scene and determining frames in which a face is detected; and

a step of clustering the frames in which the face is detected according to the main character to determine key frames, considering the above number of key characters.

Claim 16

The method for partitioning videos according to topic according to claim 15, wherein the scene change generates a thumbnail image of an input frame and detects it using the similarity of color histograms between the thumbnail images of adjacent frames.

Claim 17

The method for partitioning videos according to topic according to claim 15, wherein the number of key characters is obtained by interpreting the EPG signal.

Claim 18

The method for partitioning videos according to topic according to claim 15, wherein the face detecting step consists of:

a step of generating an integral image for the thumbnail image of the input frame and reconstructing the thumbnail image using the integral image;

a step of generating a subwindow for the reconstructed thumbnail image; and

a step of classifying whether the subwindow comprises a face.

Claim 19

The method for partitioning videos according to topic according to claim 18, wherein the face detecting step is further equipped with a step of partitioning the thumbnail image of the input frame into a plurality of areas comprising the area most likely to comprise a face, and sequentially providing the thumbnail image to the thumbnail image reconstruction step, starting from the area most likely to contain the face.

Claim 20

The method for partitioning videos according to topic according to claim 15, wherein the keyframe determination step consists of:

- a step of extracting clothing information from a frame in which a face is detected;
- a step of performing character clustering based on the extracted clothing information; and
- a step of selecting, considering the number of key characters, a cluster corresponding to the main character from among the plurality of clusters, and providing frames comprised in each selected cluster as keyframes for each subject.

Claim 21

A computer-readable recording medium that records a program that can execute a method of partitioning a video according to topic, comprising:

- a step of detecting a plurality of key frames using character information in a video sequence composed of a plurality of frames, and determining the detected key frames as the starting shot of each subject; and
- a step of creating a topic list using the starting shot of each topic.

Specification

Detailed Description of the Invention

Purpose of the Invention

Technical Field of the Invention and Prior Art

The present invention relates to video partitioning, and in particular to an apparatus and method for high-speed video partitioning according to topic based on detected main characters.

Due to the development of digital signal processing technology, comprising video and audio compression technology, users can now search and browse only the desired portion of multimedia content at the desired time. The most basic technologies for nonlinear multimedia content browsing and search are shot partitioning and shot clustering; these two technologies are the most core technologies in structurally analyzing multimedia content.

In video, a “shot” refers to a sequence of video frames obtained from a single camera without interruption, and is the most basic unit for analyzing or composing video. Here, the shot can be a single frame or multiple frames; here, a single frame is taken as an example. Additionally, a scene exists in a video, where a scene is a meaningful component in the development of a story or video composition, and a plurality of shots usually exist within one scene. These concepts of shots and scenes can be equally applied to audio programs as well as video.

Ultimately, the purpose of multimedia content indexing techniques is to extract structural information on a shot/scene basis that exists in the content, index structural information regarding multimedia content by extracting key special elements such as key frames that can represent the partition for each structural unit, and index semantic information about the occurrence of events, the appearance of audio-visual objects, the status of objects, and the background in a way that describes them on a time axis, allowing users to easily browse or search for a desired part.

However, in the prior art related to multimedia content indexing techniques, difficulties have been experienced in delivering summary results because partitions have been partitioned into too many partitions when partitioning the partition by scene change. In addition, there has been a disadvantage in that it is difficult to detect the exact start of a partition by summarizing multimedia content using single information, such as color similarity, rather than partitioning partitions based on similarity of content. Furthermore, because multimedia content is summarized using the characteristics of a specific genre, there is a disadvantage in that it is difficult to generate a summary when the broadcast type or genre changes. Moreover, due to the excessive amount of calculation when summarizing multimedia content, the disadvantage exists of it being difficult to apply to embedded systems such as mobile phones, personal digital assistants (PDAs), or digital cameras equipped with low-specification processors.

Problem to be Solved by the Invention

The technical problem to be solved by the present invention is to provide an apparatus and method for high-speed partition of a video according to topic based on detected main characters.

In order to achieve the above technical tasks, the apparatus for partitioning a video by subject according to the present invention comprises a start shot determining unit that detects a plurality of key frames using character information in a video sequence composed of a plurality of frames and determines the detected key frames as the starting shot of each subject, and a topic list generating unit that generates a topic list using the starting shot of each topic.

In order to achieve the above technical task, the method of partitioning a video by subject according to the present invention comprises a step of detecting a plurality of key frames using character information in a video sequence consisting of a plurality of frames, and determining the detected key frames as the starting shot of each subject; and a step of creating a topic list using the starting shot of each topic.

The method of partitioning the video according to topic can preferably be implemented with a computer-readable recording medium that records a program for execution on a computer.

Configuration of the Invention

--

Hereinafter, preferred embodiments of the present invention will be described in detail with reference to the attached drawings.

FIG. 1 is a diagram showing an example of division by subject in a news video. Looking at chapters 1 to 25 partitioned according to topic, each chapter consists of a starting shot set as a key frame in which the main characters appear and a data screen supporting the content, that is, an episode part.

FIG. 2 is a block diagram showing the configuration of an apparatus for partitioning a video by subject according to the present invention; it comprises a starting shot determining unit 210 and a topic list generating unit 230, and is used to partition the video into topics based on the main characters detected.

Referring to FIG. 2, the starting shot determining unit 210 detects a plurality of key frames using character information in a video sequence consisting of a plurality of frames, and determines the detected key frames as the starting shot of each subject. In particular, it is desirable for a key character to appear in each key frame. Additionally, it is preferable that starting shot detection be performed on a per-scene basis, but it is not necessarily limited thereto.

The topic list generating unit 230 generates a topic list using the starting shot of each topic determined in the starting shot determining unit 210. At this time, it is desirable to generate a topic list by collecting the starting shots detected in each scene. Additionally, frames existing between the starting shot of the current topic and the starting shot of the next topic are preferably configured as episodes of the current topic and added to each starting shot in the topic list.

FIG. 3 is a block diagram showing the detailed configuration of the start shot determining unit 210 shown in FIG. 2; it is constituted to comprise a preprocessing unit 310, a face detecting unit 330, and a key frame determining unit 350.

Referring to FIG. 3, the preprocessing unit 310 takes the video sequence constituting one video as an input, detects the scene transition, determines the frames belonging to the current scene, and uses the EPG (Electronic Program Guide) signal of the video as an input to determine the number of main characters. As shown in FIG. 4A, the EPG signal may comprise information such as the program's genre, title, director's name, character's name, plot, etc., as well as the broadcast time.

The face detecting unit 330 detects a face in each frame belonging to the current scene determined by the preprocessing unit 310. Since the main character usually faces the front, it is preferable to detect the front face. At this time, regardless of the number of faces in each frame, a determination of only whether or not a face exists can be made. For face detection, various known algorithms can be applied.

The key frame determining unit 350 extracts clothing information from frames in which faces are detected in the face detecting unit 330, clusters the frames by characters corresponding to the clothing information, and determines the frame comprising the main character as the key frame, that is, the starting shot of the subject. Since the clothing information of the main characters in one video rarely changes, the clothing information is used for clustering by character. Considering the number of main characters determined in the preprocessing unit 310, clusters containing a relatively small number of frames are removed from among the plurality of clusters generated as a result of clustering. The reason for this is that the number of appearances of the main characters is greater than that of other actors. The key frame determining unit 350 determines the character clustering result, for example, the key frame shown in FIG. 4B and is used to generate a topic list as shown in FIG. 4C. Meanwhile, examples of key frames determined by the key frame determining unit 350 for the news genre, education genre, and information/culture genre are shown in FIGS. 13 to 15.

FIG. 5 is a block diagram showing the detailed configuration of the preprocessing unit 310 in FIG. 3; it is constituted by comprising a frame image input unit 510, a thumbnail image generating unit 530, a scene change detecting unit 550, an EPG analysis unit 570, and a main character number determining unit 590.

Referring to FIG. 5, the frame image input unit 510 sequentially inputs each frame image extracted from the video sequence.

The thumbnail image generating unit 530 samples pixels at regular intervals from the original frame image of $W \times H$ size provided from the frame image input unit 510 to generate a thumbnail image of reduced size $w \times h$. According to this, when using a thumbnail image, faster processing is possible when detecting a face in the face detecting unit 230 than when using an original frame image.

The scene change detecting unit 550 stores the previous frame image, calculates the similarity of the color histogram between two consecutive frame images, that is, the current frame image and the previous frame image, and if the calculated similarity is less than a certain threshold, it determines that a scene change has been detected in the current frame. Here, the similarity ($\text{Sim}(H_t, H_{t+1})$) can be calculated as in Formula 1 below:

$$\text{Sim}(H_t, H_{t+1}) = \sum_{n=1}^N \min[H_t(n), H_{t+1}(n)]$$

Here, H_t represents the color histogram of the previous frame image, H_{t+1} represents the color histogram of the current frame image, and N represents the histogram level.

The EPG analysis unit 570 analyzes the EPG signal comprised in one video. The main character number determining unit 59 determines the main character number based on the analysis results from the EPG analysis unit 570.

FIG. 6A is a block diagram illustrating the detailed configuration of the face detecting unit 330 shown in FIG. 3, which comprises a thumbnail image reconstructing unit 610, a subwindow generating unit 630, and a classifying unit 650.

Referring to FIG. 6A, the thumbnail image reconstructing unit 610 obtains an integral image at each point from the thumbnail image for each frame image belonging to the current scene provided by the preprocessing unit 310 and reconstructs the thumbnail image. Here, a method for obtaining the integral image will be described with reference to FIG. 6B.

Referring to FIG. 6B, for example, a thumbnail image comprises four areas (A, B, C, and D) and four points (a, b, c, and d) are designated. The integral image of point (a) refers to the sum of pixel values present in the upper left part of point (a). In other words, the integrated image of point (a) refers to the sum of pixel values of pixels comprised in area (A). Here, the pixel value may be the luminance level of the pixel. Additionally, the integral square image of point (a)

refers to the sum of the squares of pixel values present in the upper left part of point (a). In other words, the integral square image of point (a) refers to the sum of the squares of the pixel values of the pixels comprised in the area (A). When introducing this concept of integral image, it becomes convenient to calculate the sum of pixel values for an arbitrary area of the image. That is, when using an integral image, the classification operation performed in the classifying unit 670, which will be described later, can be processed at high speed. For example, the sum (S(D)) of pixel values existing within the area (D) can be obtained as in Formula 2 below.

$$S(D) = ii(d) - ii(b) - ii(c) + ii(a)$$

Here, ii(d) refers to the integral image of point (d), ii(b) refers to the integral image of point (b), ii(c) refers to the integral image of point (c), and ii(a) represents the integral image of point (a), respectively.

The thumbnail image reconstructing unit 610 reconstructs the thumbnail image using the integral image at each point calculated as in Formula 2 above. Here, the thumbnail reconstructing unit 610 may be provided as an option.

The sub-window generating unit 630 partitions the thumbnail image reconstructed by the thumbnail image reconstructing unit 610 and generates a sub-window. Here, the size of the subwindow may be determined in advance and may be increased at a constant rate. That is, the subwindow is initially set to a predetermined size, for example, 20×20 pixels, and the entire image is partitioned by subwindows having the set size. Afterwards, the size of the subwindow is increased by, for example, 1.2 times, and the entire image is partitioned again by the subwindow with the increased size. In this way, the image can be partitioned by increasing the size of the subwindow until the size of the subwindow becomes the same as the size of the entire image. Subwindows generated by the subwindow generating unit 630 may overlap each other. Reference numbers 710, 730, and 750 in FIG. 7 indicate examples of subwindows generated by the subwindow generating unit 630.

The classifying unit 650 may have a structure in which n stages (S_1 to S_n) are connected in cascade. Here, each stage (S_1 to S_n) performs face detection using a classifier based on simple features. For example, the first stage (S_1) uses 4 to 5 classifiers, the second stage (S_2) uses 15 to 20 classifiers, etc., and the number of classifiers used increases toward the later stages.

Each stage consists of a weighted sum for a plurality of classifiers, and determines whether face detection is successful according to the sign of the weight sum. Each stage can be expressed as Formula 3 below.

Here, c_m represents the weight of the classifier, and $f_m(x)$ represents the output of the classifier. Each classifier consists of one simple feature and a threshold value, and outputs the value of -1 or 1 as the value of $f_m(x)$.

--

In the classifying unit 650 the first stage (S_1) attempts to detect a face using the k-th sub-window provided from the sub-window generating unit 630 as an input; if face detection fails, the k-th sub-window is determined to be a non-face and is used for face detection. If successful, the kth subwindow image is provided to the second stage (S_2). In the last stage (S_n), if face detection is successful in the k-th sub-window, the k-th sub-window is determined as the face. Meanwhile, a learning algorithm based on Adaboost can be used to select each classifier and weight. According to the Adaboost algorithm, a very efficient classifier is created by selecting a few important visual features from a large feature set. According to this cascaded stage structure, non-face cases can be determined using a small number of simple features. Since it is possible to reject immediately at an initial stage such as the first stage or the second stage, and then receive the next (k+1)th subwindow and attempt face detection, the overall face detection processing speed can be improved.

FIG. 8A shows edge simple features (811, 812) and line simple features (813, 814, 815, 816) used in each classifier comprised in the classifying unit 650. Each simple feature consists of two or three white or black squares. Each classifier subtracts the sum of pixel values located within the white square from the sum of pixel values located within the black square according to the simple feature, and compares the subtracted result with a threshold value corresponding to the simple feature. A value of 1 or -1 is output depending on the comparison result between the subtracted result and the threshold value. FIG. 8B shows an example of detecting the eye area using a line simple feature (821) consisting of one white square and two black squares or an edge simple feature (823) consisting of one white square and one black square. When using the line simple feature, the difference in pixel values between the eye area and the bridge of the nose is measured, taking into account the fact that the eye area is darker than the bridge of the nose area. When using the

edge simple feature, the difference in grayscale value between the eye area and the upper cheek area is measured, taking into account the fact that the eye area is darker than the cheek area. In this way, a wide variety of simple features for detecting faces can exist.

FIG. 9 is a diagram showing an example of partitioning a frame image for high-speed face detection in the face detecting unit 330 shown in FIG. 3. Before inputting the thumbnail image to the thumbnail image reconstructing unit 610, the frame image is partitioned into areas I to V (910, 930, 950, 970, 990) according to the possibility of a face being present. Here, the division position of each area can be statistically obtained through experiment or simulation. Since area I (910) generally corresponds to the area where the face 900 is most likely to be located, the thumbnail image is sequentially provided to the thumbnail image reconstructing unit 610 from area I (910) to area IV (970).

FIG. 10 is a flowchart describing the operation of the face detecting unit 330 shown in FIG. 3.

Referring to FIG. 10, in step 1011, area I (910) is reconstructed by generating an integral image for area I (910). In step 1013, a subwindow is generated from the image reconstructed as an integral image for area I (910). In step 1015, it is determined whether face detection was successful in the generated subwindow, and a frame image containing a face is constructed using the subwindow in which face detection was successful. In step 1017, if face detection fails in the generated sub-window as a result of the determination in step 1015, it is determined whether the generation of the sub-window for area I (910) has ended; if the generation of the subwindow for area I (910) has not ended, the process returns to step 1013. If the generation of the subwindow for area I (910) has ended, the process proceeds to step 1031.

In step 1031, area II (930) is reconstructed by generating an integral image for area II (930). In step 1033, a subwindow is generated from the image reconstructed as an integral image for area I (910) and area II (930). At this time, it is desirable to exclude subwindows located only in area I (910).

In step 1035, it is determined whether face detection was successful in the generated subwindow, and a frame image containing a face is constructed using the subwindow in which face detection was successful. In step 1037, if face detection fails in the subwindow generated as a result of the determination in step 1035, it is determined whether the generation of subwindows for area I (910) and area II (930) has ended; if the generation of subwindows for area I (910) and area II (930) has not ended, a return is made to step 1033, but if the generation of subwindows for area I (910) and area II (930) has ended, the process proceeds to step 1051.

In step 1051, area III (950) is reconstructed by generating an integral image for area III (950). In step 1053, subwindows are generated from images reconstructed as integral images for area I (910), area II (930), and area III (950). At this time, it is desirable to exclude subwindows located only in area I (910) and area II (930). In step 1055, it is determined whether face detection was successful in the generated subwindow, and a frame image containing a face is constructed using the subwindow in which face detection was successful. In step 1057, if face detection fails in the subwindow generated as a result of the determination in step 1055, it is determined whether the generation of subwindows for area I (910), area II (930), and area III (950) has ended; if the generation of subwindows for area I (910), area II (930), and area III (950) has not ended, a return is made to step 1053, but if the generation of subwindows for area I (910), area II (930), and area III (950) has ended, the process proceeds to step 1071.

In step 1071, the integral image for area IV (970) is generated to reconstruct area IV (970).

At this time, it is desirable to exclude subwindows located only in area I (910), area II (930), and area III (950).

In step 1073, subwindows of images reconstructed as integral images are generated for area I (910), area II (930), area III (950), and area IV (970).

In step 1075, it is determined whether face detection was successful in the generated subwindow, and a frame image containing the face is constructed using the subwindow in which face detection was successful. In step 1077, if face detection fails in the generated subwindow as a result of the determination in step 1075, a determination is made of whether the generation of subwindows for area I (910), area II (930), area III (950), and area IV (970) has ended. If the generation of subwindows for area I (910), area II (930), area III (950), and area IV (970) has not ended, a return is made to step 1073; if the generation of subwindows for area I (910), area II (930), area III (950), and area IV (970) has ended, the corresponding thumbnail image is determined to be a frame image that does not comprise a face.

Steps 1011, 1031, 1051, and 1071 are performed in the thumbnail image reconstructing unit 610, and steps 1013, 1033, 1053, and 1073 are performed in the subwindow generating unit 630; steps 1015 and 1017, steps 1035 and 1037, steps 1055 and 1057, and steps 1075 and 1077 are preferably performed in the classifying unit 650, respectively.

FIG. 11 is a block diagram showing the detailed configuration of the key frame determining unit 350 in FIG. 3; it comprises a clothing information extracting unit 1110, a character clustering unit 1130, and a main character determining unit 1150.

Referring to FIG. 11, the clothing information extracting unit 1110 extracts clothing information, for example, a histogram for the color of the clothing, from the frame image in which the face is detected by the face detecting unit 330. Here, a histogram for the color of clothing is extracted for a certain detecting area 1230 that is set based on the face area (1210 in FIG. 12) detected by the face detecting unit 330. The reason for extracting a histogram for the color of clothing is because major characters in a video usually appear wearing the same clothing.

The character clustering unit 1130 receives the histogram for the color of clothing extracted from the clothing information extracting unit 1110 as input, and clusters frame images in which faces are detected to create a plurality of clusters. At this time, a frame image having a histogram for the color of the same clothing is judged to be a frame image containing the same character.

The main character determining unit 1150 selects a cluster containing the main character among the plurality of clusters provided by the character clustering unit 1130, taking into account the number of main characters provided by the pre-processing unit 310. Here, when the number of main characters is 1, one cluster containing the largest number of frame images is selected, and when there is a plurality of main characters, clusters are selected in the order of the number of frame images comprised. The frame images comprised in the selected cluster constitute the starting shot as the key frame of each subject.

The present invention can also be implemented as computer-readable code on a computer-readable recording medium. Computer-readable recording media comprise all types of recording apparatuses that store data that can be read by a computer system. Examples of computer-readable recording media include ROM, RAM, CD-ROM, magnetic tapes, floppy disks, and optical data storage apparatuses, and are also implemented in the form of a carrier wave (e.g., transmission via the Internet). Additionally, computer-readable recording media can be distributed across computer systems connected to a network, so that computer-readable code can be stored and executed in a distributed manner. Further, functional programs, codes, and code partitions for implementing the present invention can be easily deduced by programmers in the technical field pertaining to the present invention.

Effects of the Invention

As described above, By adopting a character-based high-speed calculation algorithm to partition videos according to topic, the apparatus and method for partitioning a video according to topic according to the present invention has the advantage of being easily applicable to embedded systems such as mobile phones, PDAs, and digital cameras equipped with low-specification processors.

In addition, the apparatus and method for partitioning videos according to topic according to the present invention has the advantage of being applicable to all videos edited by multiple topics, such as news genre, education genre, information/culture genre, etc.

Moreover, the apparatus and method for partitioning a video by subject according to the present invention can be applied not only to embedded systems but also to various products such as Blue-ray disc recorders or home servers for Digital Contents Management (DCM), thereby increasing user convenience.

The present invention has been described with reference to the above embodiments, but this is merely illustrative; those skilled in the art will understand that various modifications and equivalent embodiments are possible therefrom. Therefore, the true technical protection scope of the present invention should be determined by the technical spirit of the attached patent claims.

Brief Description of the Drawings

FIG. 1 is a diagram showing an example of topical division in a news video;

FIG. 2 is a block diagram showing the configuration of an apparatus for partitioning a video according to topic according to the present invention;

FIG. 3 is a block diagram showing the detailed configuration of the start shot determining unit shown in FIG. 2;

FIGS. 4A to 4C are diagrams for describing the operation of each component shown in FIG. 3;

FIG. 5 is a block diagram showing the detailed configuration of the preprocessing unit in FIG. 3;

FIG. 6A is a block diagram describing the detailed configuration of the face detecting unit shown in FIG. 3, and FIG. 6B is a diagram describing a method of constructing an integral image;

FIG. 7 is a diagram showing an example of a sub-window used in the face detecting unit shown in FIG. 3;

FIGS. 8A and 8B are diagrams showing examples of features used in the classifying unit shown in FIG. 6A;

FIG. 9 is a diagram showing an example of partitioning a frame image for face detection in the face detecting unit shown in FIG. 3;

FIG. 10 is a flowchart describing the operation of the face detecting unit shown in FIG. 3;

FIG. 11 is a block diagram showing the detailed configuration of the key frame determining unit in FIG. 3;

FIG. 12 is a diagram describing the operation of the clothing information extracting unit in FIG. 11;

FIG. 13 is a diagram showing examples of key frames determined as starting shots according to the present invention in the case of a news genre;

FIG. 14 is a diagram showing examples of key frames determined as starting shots according to the present invention in the case of an educational genre; and

FIG. 15 is a diagram showing examples of key frames determined as starting shots according to the present invention in the case of the information/culture genre.

Drawings

FIG. 1



FIG. 2

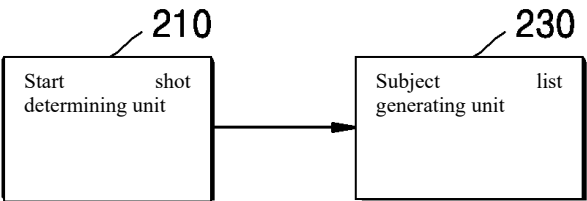


FIG. 3

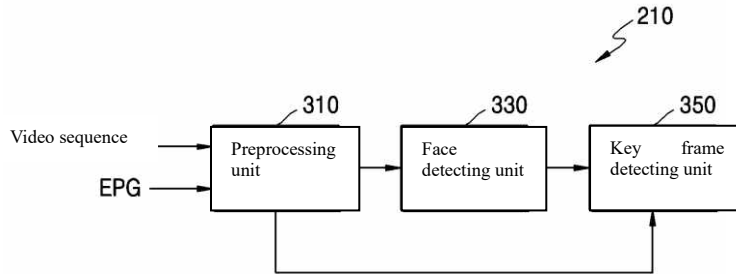


FIG. 4A

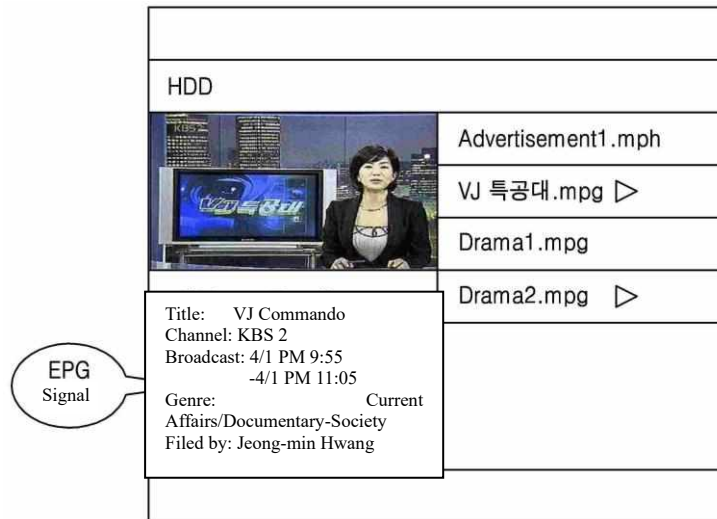


FIG. 4B



FIG. 4C



FIG. 5

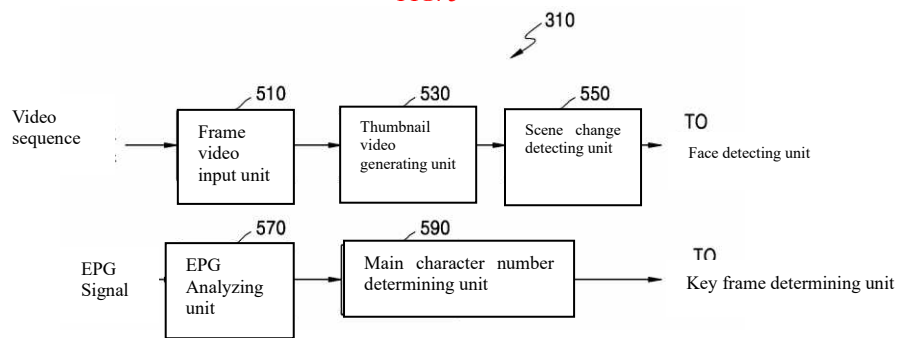


FIG. 6A

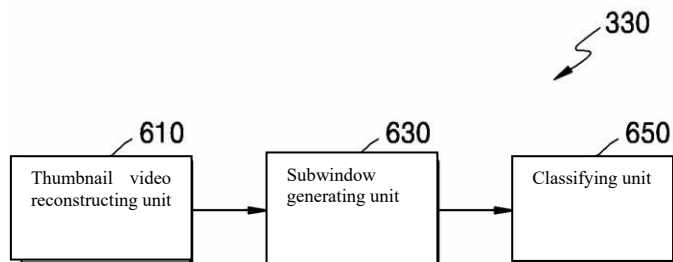


FIG. 6B

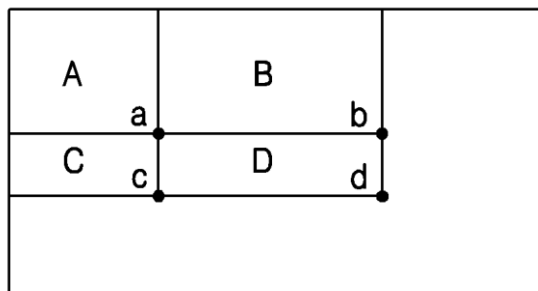


FIG. 7



FIG. 8A

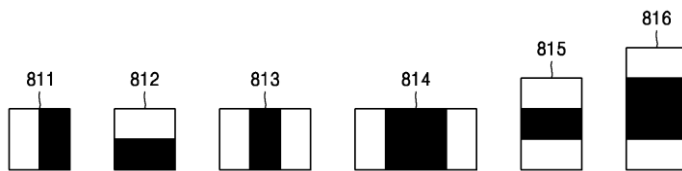


FIG. 8B

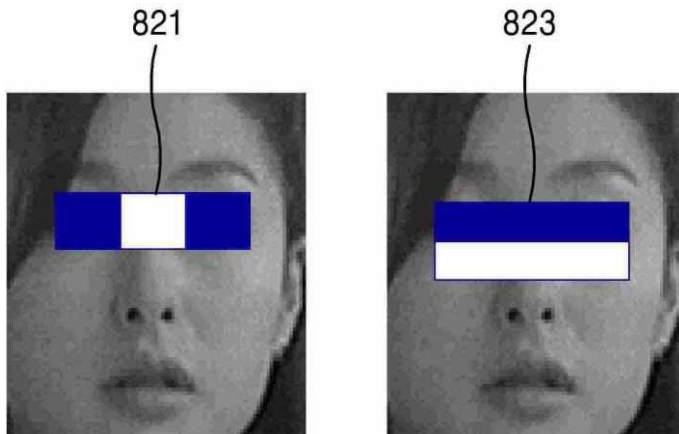


FIG. 9

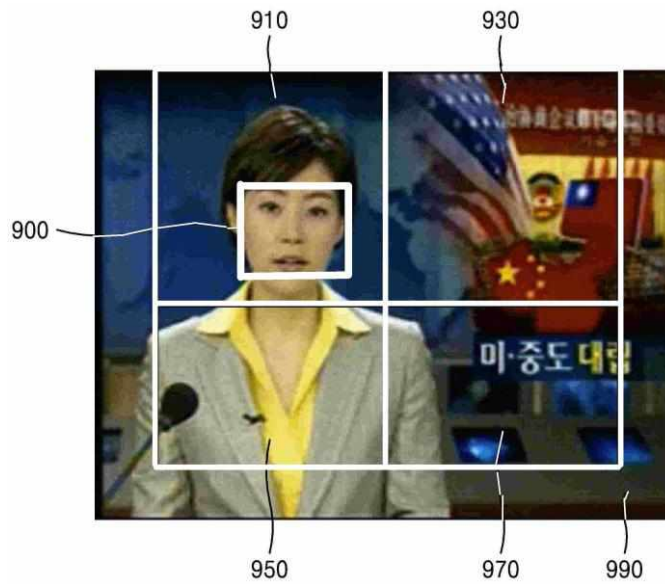


FIG.10

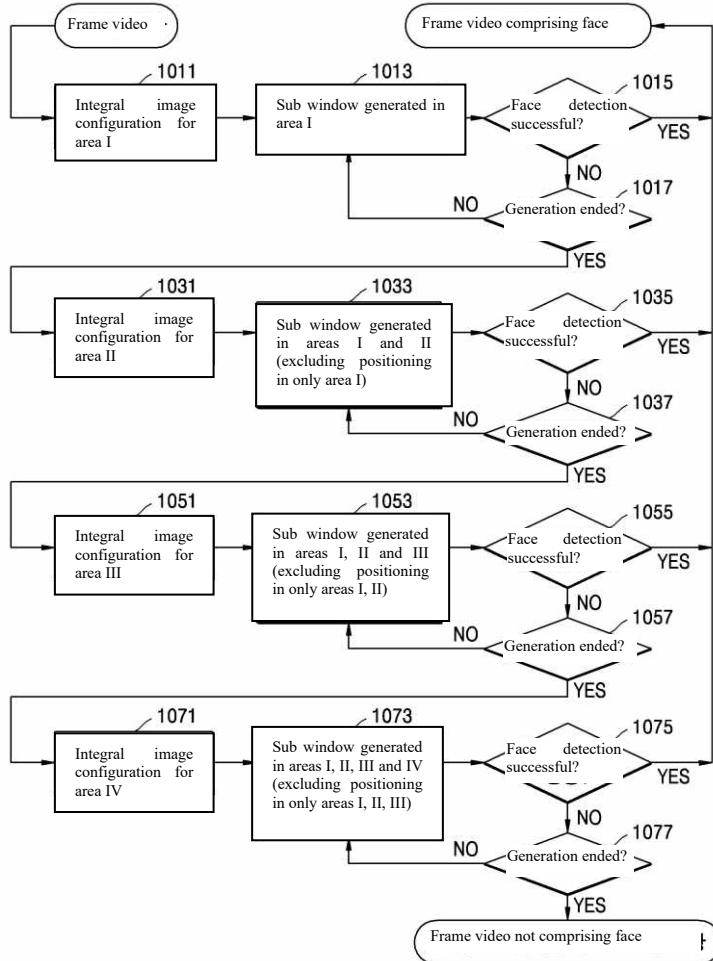


FIG.11

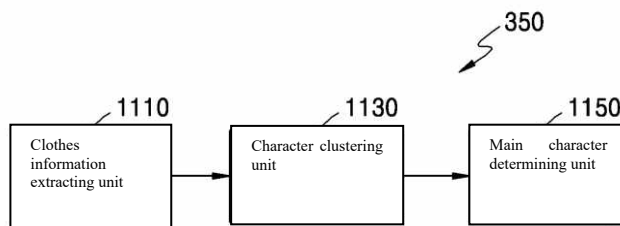


FIG.12

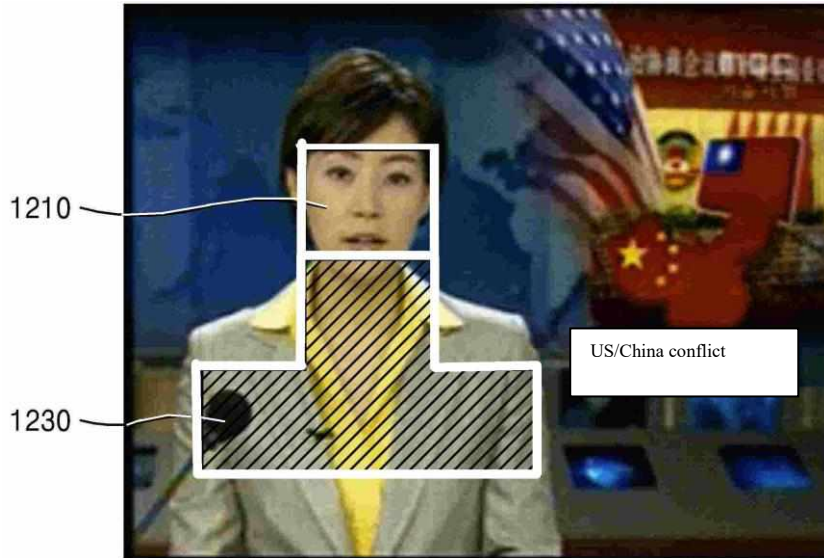


FIG.13

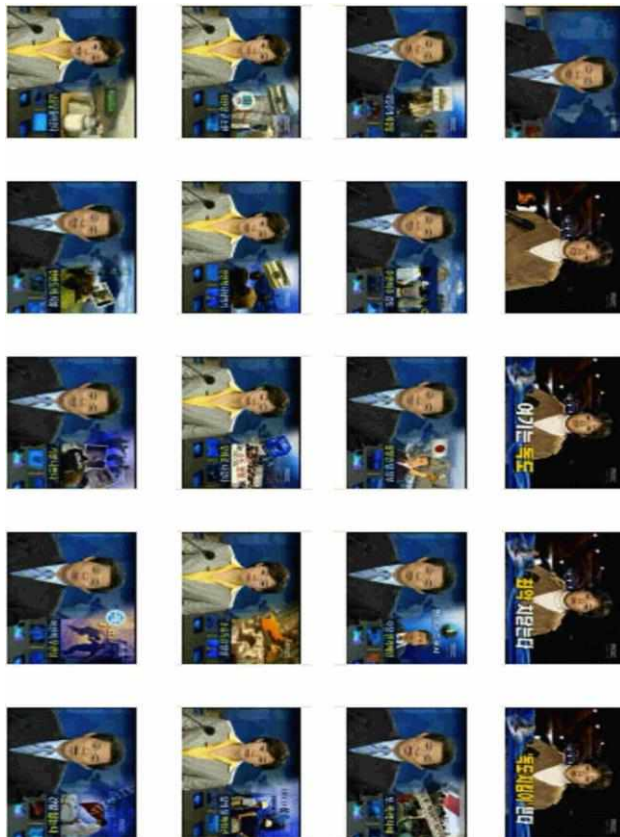


FIG.14

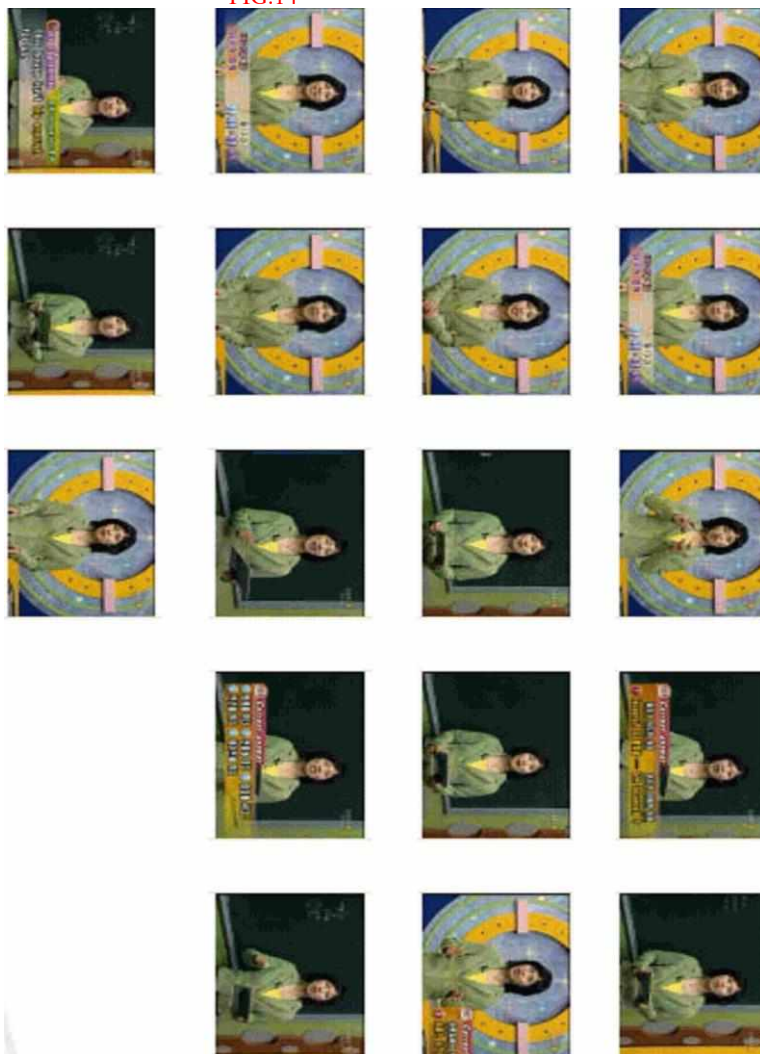


FIG.15





(19)대한민국특허청(KR)
(12) 공개특허공보(A)

(51) 。 Int. Cl.
H04N 5/93 (2006.01)

(11) 공개번호 10-2007-0017068
(43) 공개일자 2007년02월08일

(21) 출원번호 10-2006-0073770
(22) 출원일자 2006년08월04일
심사청구일자 없음

(30) 우선권주장 1020050071507 2005년08월04일 대한민국(KR)

(71) 출원인 삼성전자주식회사
경기도 수원시 영통구 매탄동 416

(72) 발명자 김정배
경기 용인시 풍덕천동 삼익아파트 102동 304호
황두선
서울 은평구 수색동 진흥아파트 103-606
김지연
서울 강동구 명일동 한양아파트 5-102

(74) 대리인 리앤목특허법인

전체 청구항 수 : 총 21 항

(54) 동영상의 주제별 분할장치 및 방법

(57) 요약

동영상의 주제별 분할장치 및 방법이 개시된다. 동영상의 주제별 분할장치는 복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키 프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정하는 시작 샷 결정부; 및 상기 각 주제의 시작 샷을 이용하여 주제 리스트를 생성하는 주제 리스트 생성부로 이루어진다.

대표도

도 2

특허청구의 범위

청구항 1.

복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키 프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정하는 시작 샷 결정부; 및

상기 각 주제의 시작 샷을 이용하여 주제 리스트를 생성하는 주제 리스트 생성부를 포함하는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 2.

제1 항에 있어서, 상기 시작 샷 결정부는 상기 비디오 시퀀스에 포함된 주요 인물의 옷 정보를 이용하여 상기 키 프레임들을 검출하는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 3.

제1 항에 있어서, 상기 주제 리스트 생성부는 현재 주제의 시작 샷과 다음 주제의 시작 샷 사이에 존재하는 프레임들은 현재 주제의 에피소드로 구성하여, 주제 리스트의 각 시작 샷에 추가하는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 4.

제1 항에 있어서, 상기 시작 샷 결정부는

상기 비디오 시퀀스에 포함된 프레임들 중에서 장면 변환을 검출하여 각 장면에 속하는 프레임들을 결정하고, 상기 비디오 시퀀스에 등장하는 주요 인물의 수를 구하는 전처리부;

상기 전처리부에서 결정된 각 장면에 속하는 프레임들로부터 얼굴을 검출하고, 얼굴이 검출되는 프레임들을 결정하는 얼굴검출부; 및

상기 주요 인물의 수를 고려하여, 상기 얼굴이 검출된 프레임들을 상기 주요 인물에 따라서 클러스터링하여 키프레임들을 결정하는 키프레임 결정부로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 5.

제4 항에 있어서, 상기 전처리부는 현재 프레임과 이전 프레임간의 산출되는 유사도를 이용하여 상기 장면 변환을 검출하는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 6.

제4 항에 있어서, 상기 전처리부는 EPG 신호로부터 상기 주요 인물의 수를 구하는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 7.

제4 항에 있어서, 상기 전처리부는

입력되는 프레임의 썸네일 영상을 생성하는 썸네일 영상 생성부;

인접하는 프레임의 상기 썸네일 영상들간의 칼라 히스토그램의 유사도를 이용하여 장면 전환을 검출하는 장면전환 검출부; 및

해석된 EPG 신호로부터 상기 주요인물의 수를 결정하는 주요 인물수 결정부로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 8.

제4 항에 있어서, 상기 얼굴검출부는

입력되는 프레임의 썸네일 영상에 대하여 적분영상을 생성하고, 상기 적분영상을 이용하여 썸네일영상을 재구성하는 썸네일영상 재구성부;

재구성된 썸네일영상에 대하여 서브윈도우를 발생시키는 서브윈도우 발생부;

상기 서브윈도우가 얼굴을 포함하는지 여부를 분류하는 분류부로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 9.

제8 항에 있어서, 상기 얼굴검출부는

입력되는 프레임의 썸네일 영상을 얼굴이 포함될 가능성이 가장 많은 영역을 포함하여 복수개의 영역으로 분할하고, 상기 얼굴이 포함될 가능성이 가장 많은 영역부터 순차적으로 상기 썸네일영상 재구성부로 제공하는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 10.

제4 항에 있어서, 상기 키프레임 결정부는

얼굴이 검출된 프레임에서 옷정보를 추출하는 옷정보 추출부;

상기 옷정보 추출부에서 추출된 옷정보를 기반으로 하여 인물 클러스터링을 수행하는 인물 클러스터링부; 및

상기 주요 인물의 수를 고려하여, 상기 인물 클러스터링부에서 클러스터링된 복수의 클러스터들 중에서 상기 주요 인물에 해당하는 클러스터를 선택하고, 선택된 각 클러스터에 포함된 프레임들을 각 주제의 키프레임으로 제공하는 주요 인물 결정부로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 11.

제10 항에 있어서, 상기 옷정보는 옷의 칼라 히스토그램인 것을 특징으로 하는 동영상의 주제별 분할장치.

청구항 12.

복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키 프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정하는 단계; 및

상기 각 주제의 시작 샷을 이용하여 주제 리스트를 생성하는 단계를 포함하는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 13.

제12 항에 있어서, 상기 시작 샷 결정단계는 상기 비디오 시퀀스에 포함된 주요 인물의 옷 정보를 이용하여 상기 키 프레임들을 검출하는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 14.

제12 항에 있어서, 상기 주제 리스트 생성단계는 현재 주제의 시작 샷과 다음 주제의 시작 샷 사이에 존재하는 프레임들은 현재 주제의 에피소드로 구성하여, 주제 리스트의 각 시작 샷에 추가하는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 15.

제12 항에 있어서, 상기 시작 샷 결정단계는

상기 비디오 시퀀스에 포함된 프레임들 중에서 장면 변환을 검출하여 각 장면에 속하는 프레임들을 결정하고, 상기 비디오 시퀀스에 등장하는 주요 인물의 수를 구하는 단계;

상기 각 장면에 속하는 프레임들로부터 얼굴을 검출하고, 얼굴이 검출되는 프레임들을 결정하는 단계; 및

상기 주요 인물의 수를 고려하여, 상기 얼굴이 검출된 프레임들을 상기 주요 인물에 따라서 클러스터링하여 키프레임들을 결정하는 단계로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 16.

제15 항에 있어서, 상기 장면 전환은, 입력되는 프레임의 썸네일 영상을 생성하고, 인접하는 프레임의 상기 썸네일 영상들 간의 칼라 히스토그램의 유사도를 이용하여 검출하는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 17.

제15 항에 있어서, 상기 주요 인물의 수는 EPG 신호를 해석하여 구하는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 18.

제15 항에 있어서, 상기 얼굴검출단계는

입력되는 프레임의 썸네일 영상에 대하여 적분영상을 생성하고, 상기 적분영상을 이용하여 썸네일영상을 재구성하는 단계;

재구성된 썸네일영상에 대하여 서브윈도우를 발생시키는 단계;

상기 서브윈도우가 얼굴을 포함하는지 여부를 분류하는 단계로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 19.

제18 항에 있어서, 상기 얼굴검출단계는

입력되는 프레임의 썸네일 영상을 얼굴이 포함될 가능성이 가장 많은 영역을 포함하여 복수개의 영역으로 분할하고, 상기 얼굴이 포함될 가능성이 가장 많은 영역부터 순차적으로 상기 썸네일영상 재구성단계로 제공하는 단계를 더 구비하는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 20.

제15 항에 있어서, 상기 키프레임 결정단계는

얼굴이 검출된 프레임에서 웃정보를 추출하는 단계;

상기 추출된 웃정보를 기반으로 하여 인물 클러스터링을 수행하는 단계; 및

상기 주요 인물의 수를 고려하여, 상기 클러스터링된 복수의 클러스터들 중에서 상기 주요 인물에 해당하는 클러스터를 선택하고, 선택된 각 클러스터에 포함된 프레임들을 각 주제의 키프레임으로 제공하는 단계로 이루어지는 것을 특징으로 하는 동영상의 주제별 분할방법.

청구항 21.

복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정하는 단계; 및

상기 각 주제의 시작 샷을 이용하여 주제 리스트를 생성하는 단계를 포함하는 동영상의 주제별 분할방법을 실행시킬 수 있는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 동영상 분할에 관한 것으로서, 특히 검출되는 주요 인물에 기반하여 주제별로 고속으로 동영상을 분할하기 위한 장치 및 방법에 관한 것이다.

비디오 및 오디오 압축 기술을 포함하는 디지털 신호처리 기술의 발달로 인하여 사용자들은 원하는 멀티미디어 콘텐츠를 원하는 시점에 원하는 부분만을 검색하고 브라우징할 수 있게 되었다. 비선형적인 멀티미디어 콘텐츠 브라우징(non-linear multimedia content browsing)과 검색을 위하여 가장 기본이 되는 기술은 샷 분할(shot segmentation) 기법과 샷 클러스터링(shot clustering) 기법이며, 이 두가지 기술은 멀티미디어 콘텐츠를 구조적으로 분석하는데 있어서 가장 핵심이 되는 기술이다.

비디오에서 샷은 방해(interruption)없이 하나의 카메라로부터 얻어진 비디오 프레임들의 시퀀스를 의미하며, 이는 비디오를 분석하거나 구성하는 가장 기본이 되는 단위이다. 여기서, 샷은 단일 프레임 또는 복수의 프레임이 될 수 있으나, 여기서는 단일 프레임을 예로 들기로 한다. 또한, 비디오에는 장면(scene)이 존재하는데, 여기서 장면은 이야기 전개 또는 비디오 구성에 있어서 의미적인 구성요소이며, 통상 하나의 장면 내에는 복수의 샷이 존재한다. 이러한 샷과 장면의 개념은 비디오 뿐만 아니라 오디오 프로그램에도 동일하게 적용할 수 있다.

궁극적으로 멀티미디어 콘텐츠 인덱싱 기법은 콘텐츠에 존재하는 샷/장면 단위의 구조적인 정보를 추출하고, 각각의 구조 단위별로 그 세그먼트를 대표할 수 있는 키 프레임 등과 같은 주요 특징요소를 추출하여 멀티미디어 콘텐츠에 대한 구조적인 정보를 인덱싱하고, 이벤트의 발생, 시청각 오브젝트의 등장, 오브젝트의 상태, 배경 등에 대한 의미적인 정보를 시간축 상에 기술하는 방식으로 인덱싱하여 사용자가 원하는 부분을 쉽게 브라우징하거나 검색할 수 있도록 하는데 그 목적이 있다.

그런데, 멀티미디어 콘텐츠 인덱싱 기법과 관련된 종래기술에서는 장면 전환에 의한 세그먼트 분할시, 너무 많은 세그먼트로 분할하기 때문에 요약의 결과를 전달하는데 어려움이 있었다. 또한, 내용의 유사성에 기반하여 세그먼트들을 분할하는 것이 아니라, 단일 정보, 예를 들면 칼라의 유사성을 가지고 멀티미디어 콘텐츠를 요약함으로써 정확한 세그먼트의 시작을 검출하기 어려운 단점이 있다. 또한, 특정 장르의 특징을 이용하여 멀티미디어 콘텐츠를 요약하기 때문에 방송형태 혹은 장르가 바뀌는 경우 요약 생성이 어려운 단점이 있다. 또한, 멀티미디어 콘텐츠 요약시 계산량의 과다로 인하여 낮은 사양의 프로세서를 탑재하는 모바일 폰, PDA(Personal Digital Assistant), 혹은 디지털 카메라와 같은 임베디드 시스템(embedded system)에는 적용하기 어려운 단점이 있다.

발명이 이루고자 하는 기술적 과제

본 발명이 이루고자 하는 기술적 과제는 검출되는 주요 인물에 기반하여 주제별로 고속으로 동영상을 분할하기 위한 장치 및 방법을 제공하는데 있다.

상기 기술적 과제를 달성하기 위하여 본 발명에 따른 동영상의 주제별 분할장치는 복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키 프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정하는 시작 샷 결정부; 및 상기 각 주제의 시작 샷을 이용하여 주제 리스트를 생성하는 주제 리스트 생성부를 포함한다.

상기 기술적 과제를 달성하기 위하여 본 발명에 따른 동영상의 주제별 분할방법은 복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키 프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정하는 단계; 및 상기 각 주제의 시작 샷을 이용하여 주제 리스트를 생성하는 단계를 포함한다.

상기 동영상의 주제별 분할방법은 바람직하게는 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체로 구현할 수 있다.

발명의 구성

이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예에 대하여 상세하게 설명하기로 한다.

도 1은 뉴스 동영상에서 주제별 분할의 예를 보여주는 도면이다. 주제별로 분할된 챕터 1 내지 챕터 25를 살펴보면, 각 챕터는 주요 인물이 등장하는 키 프레임으로 설정되는 시작 샷과 내용을 뒷받침하는 자료화면 즉, 에피소드 부분으로 구성된다.

도 2는 본 발명에 따른 동영상의 주제별 분할장치의 구성을 나타내는 블록도로서, 시작 샷 결정부(210) 및 주제 리스트 생성부(230)를 포함하여 이루어지며, 검출되는 주요 인물에 기반하여 동영상을 주제별로 분할하기 위한 것이다.

도 2를 참조하면, 시작 샷 결정부(210)에서는 복수의 프레임으로 구성되는 비디오 시퀀스에서 인물 정보를 이용하여 복수의 키 프레임을 검출하고, 검출된 키프레임들을 각 주제의 시작 샷으로 결정한다. 특히, 각 키 프레임에는 주요 인물이 등장하는 것이 바람직하다. 또한, 시작 샷 검출은 매 장면 단위로 수행되는 것이 바람직하나, 반드시 이에 한정되지는 않는다.

주제 리스트 생성부(230)에서는 시작 샷 결정부(210)에서 결정된 각 주제의 시작 샷을 이용하여 주제 리스트를 생성한다. 이때, 각 장면에서 검출된 시작 샷들을 취합하여 주제 리스트를 생성하는 것이 바람직하다. 또한, 바람직하게는 현재 주제의 시작 샷과 다음 주제의 시작 샷 사이에 존재하는 프레임들은 현재 주제의 에피소드로 구성하여, 주제 리스트의 각 시작 샷에 부가된다.

도 3은 도 2에 도시된 시작 샷 결정부(210)의 세부적인 구성을 보여주는 블록도로서, 전처리부(310), 얼굴검출부(330), 및 키 프레임 결정부(350)를 포함하여 이루어진다.

도 3을 참조하면, 전처리부(310)에서는 하나의 동영상상을 구성하는 비디오 시퀀스를 입력으로 하여 장면전환을 검출하여 현재 장면에 속하는 프레임들을 결정하고, 해당 동영상의 EPG(Electronic Program Guide) 신호를 입력으로 하여 주요 인물의 수를 결정한다. EPG 신호에는 도 4a에 도시된 바와 같이 방송시간 뿐만 아니라 프로그램의 장르, 제목, 감독 이름, 등장 인물 이름, 줄거리 등과 같은 정보가 포함될 수 있다.

얼굴검출부(330)에서는 전처리부(310)로부터 결정되는 현재 장면에 속하는 각 프레임에서 얼굴을 검출한다. 주요 인물은 통상 정면을 바라보고 있으므로 바람직하게는 정면 얼굴을 검출한다. 이때, 각 프레임에서 얼굴의 갯수와는 상관없이 얼굴이 존재하는지 유무만 판단할 수 있다. 얼굴검출을 위해서는 공지된 다양한 알고리즘을 적용할 수 있다.

키 프레임 결정부(350)에서는 얼굴검출부(330)에서 얼굴이 검출된 프레임에 대하여 옷 정보를 추출하여 옷 정보에 대응하는 인물별로 프레임들을 클러스터링하고, 주요 인물이 포함된 프레임을 키 프레임 즉, 주제의 시작 샷으로 결정한다. 하나의 동영상에서 주요 인물의 옷 정보는 거의 변하지 않으므로 옷 정보를 인물별 클러스터링에 사용한다. 전처리부(310)에서 결정된 주요 인물의 수를 고려하여, 클러스터링 결과 생성되는 다수의 클러스터 중에서 포함된 프레임의 수가 상대적으로 적은 클러스터는 제거한다. 그 이유는 주요 인물의 등장 횟수는 다른 출연자들에 비하여 많기 때문이다. 키 프레임 결정부(350)에서는 인물 클러스터링 결과, 예를 들면 도 4b에 도시된 키 프레임을 결정하여 도 4c에 도시된 바와 같은 주제 리스트를 생성하는데 사용된다. 한편, 뉴스 장르, 교육 장르, 정보/교양 장르에 대하여 키 프레임 결정부(350)에 의해 결정된 키 프레임의 예는 도 13 내지 도 15에 도시된 바와 같다.

도 5는 도 3에 있어서 전처리부(310)의 세부적인 구성을 보여주는 블록도로서, 프레임영상 입력부(510), 썸네일영상 생성부(530), 장면전환 검출부(550), EPG 해석부(570) 및 주요 인물수 결정부(590)를 포함하여 이루어진다.

도 5를 참조하면, 프레임영상 입력부(510)에서는 비디오 시퀀스에서 추출되는 각 프레임영상을 순차적으로 입력한다.

썸네일영상 생성부(530)에서는 프레임영상 입력부(510)로부터 제공되는 W×H 사이즈의 원본 프레임영상에 대하여 일정한 간격으로 화소를 샘플링함으로써 사이즈가 축소된 w×h 사이즈의 썸네일영상(thumbnail image)을 생성한다. 이에 따르면, 원본 프레임영상을 사용하는 경우보다 썸네일영상을 사용할 경우 얼굴검출부(230)에서의 얼굴검출시 고속 처리가 가능해진다.

장면전환 검출부(550)에서는 이전 프레임영상을 저장하고 있으며, 연속되는 두 프레임영상, 즉 현재 프레임영상과 이전 프레임영상간의 칼라 히스토그램에 대한 유사도를 산출하고, 산출된 유사도가 일정한 문턱치보다 작을 경우 현재 프레임에서 장면전환이 검출된 것으로 판단한다. 여기서, 유사도($Sim(H_t, H_{t+1})$)는 다음 수학적 식 1에서와 같이 산출될 수 있다.

$$Sim(H_t, H_{t+1}) = \sum_{n=1}^N \min[H_t(n), H_{t+1}(n)]$$

여기서, H_t 는 이전 프레임영상의 칼라 히스토그램, H_{t+1} 은 현재 프레임영상의 칼라 히스토그램을 나타내고, N은 히스토그램 레벨을 나타낸다.

EPG 해석부(570)에서는 하나의 동영상에 포함된 EPG 신호를 해석한다. 주요 인물수 결정부(59)에서는 EPG 해석부(570)에서의 해석결과에 근거하여 주요 인물수를 결정한다.

도 6a는 도 3에 도시된 얼굴검출부(330)의 세부적인 구성을 설명하는 블록도로서, 썸네일영상 재구성부(610), 서브원도우 발생부(630), 및 분류부(650)를 포함하여 이루어진다.

도 6a를 참조하면, 썸네일영상 재구성부(610)는 전처리부(310)로부터 제공되는 현재 장면에 속하는 각 프레임영상에 대한 썸네일영상에서 각 점에서의 적분영상(integral image)을 구하여 썸네일영상을 재구성한다. 여기서, 적분영상을 구하는 방법들 도 6b를 참조하여 설명하기로 한다.

도 6b를 참조하면, 썸네일영상에서 네 개의 영역들(A, B, C 및 D)을 포함하고, 네 개의 점들(a, b, c 및 d)이 지정되는 경우를 예를 들기로 한다. 점(a)의 적분영상이란, 점(a)의 왼쪽 윗 부분에 존재하는 화소값들의 합을 의미한다. 즉, 점(a)의 적분영상은 영역(A)에 포함되는 화소들의 화소값들의 합을 의미한다. 여기서, 화소값이란, 화소의 휘도레벨이 될 수 있다. 또

한, 점(a)의 적분제곱영상은 점(a)의 왼쪽 윗 부분에 존재하는 화소값들의 제곱들의 합을 의미한다. 즉, 점(a)의 적분제곱 영상은 영역(A)에 포함되는 화소들의 화소값들의 제곱들의 합을 의미한다. 이와 같은 적분영상이란 개념을 도입할 경우, 영상의 임의의 영역에 대한 화소값들의 합을 구하는 것이 편리해진다. 즉, 적분영상을 이용할 경우, 후술할 분류부(670)에서 수행되는 분류동작을 고속으로 처리할 수 있다. 예를 들어, 영역(D) 내부에 존재하는 화소값들의 합(S(D))은 다음 수학적 식 2와 같이 구해질 수 있다.

$$S(D)=ii(d)-ii(b)-ii(c)+ii(a)$$

여기서, ii(d)는 점(d)의 적분영상을 의미하고, ii(b)는 점(b)의 적분영상을 의미하고, ii(c)는 점(c)의 적분영상을 의미하고, ii(a)는 점(a)의 적분영상을 각각 나타낸다.

썸네일영상 재구성부(610)는 상기 수학적 식 2에서와 같이 산출되는 각 점에서의 적분영상을 이용하여 썸네일영상을 재구성한다. 여기서, 썸네일 재구성부(610)는 옵션으로 구비될 수 있다.

서브윈도우 발생부(630)는 썸네일영상 재구성부(610)에서 재구성된 썸네일영상을 분할하여 서브윈도우를 발생시킨다. 여기서, 서브윈도우의 사이즈는 미리 결정될 수 있으며 일정한 비율로 증가될 수도 있다. 즉, 서브윈도우는 초기에 소정 사이즈, 예를 들면 20×20 화소의 사이즈로 설정되고, 설정된 사이즈를 갖는 서브 윈도우에 의해 전체 영상을 분할한다. 그 후, 서브윈도우의 사이즈를 예를 들면 1.2배 증가시키고, 증가된 사이즈를 갖는 서브윈도우에 의해 전체 영상을 다시 분할한다. 이와 같은 방식으로 서브윈도우의 사이즈가 전체 영상의 사이즈와 동일해질 때까지, 서브윈도우의 사이즈를 증가시키면서 영상을 분할할 수 있다. 서브윈도우 발생부(630)에서 발생하는 서브윈도우들은 서로 중첩될 수도 있다. 도 7의 참조번호 710, 730, 및 750은 서브윈도우 발생부(630)에서 발생하는 서브윈도우의 예를 나타낸다.

분류부(650)는 n개의 스테이지(S_1 내지 S_n)가 캐스케이드로 연결된 구조일 수 있다. 여기서, 각 스테이지(S_1 내지 S_n)는 심플특징에 기반한 분류기를 이용하여 얼굴검출을 수행하는데, 예를 들면 제1 스테이지(S_1)는 4~5개의 분류기를 사용하고, 제2 스테이지(S_2)는 15~20개의 분류기를 사용하는 등, 후단으로 갈수록 사용되는 분류기의 수가 증가되는 구조이다.

각 스테이지는 복수개의 분류기에 대한 가중치 합(weighted sum)으로 이루어져 가중치 합의 부호에 따라서 얼굴검출 성공여부를 판단하는데, 각 스테이지는 다음 수학적 식 3과 같이 나타낼 수 있다.

$$H(x) = \text{sign} \left[\sum_{m=1}^M c_m \cdot f_m(x) \right]$$

여기서, c_m 은 분류기의 가중치를 나타내고, $f_m(x)$ 는 분류기의 출력을 나타내는 것으로서, 각 분류기는 하나의 심플 특징과 문턱값으로 구성되어, $f_m(x)$ 의 값으로 -1 또는 1의 값을 출력한다.

분류부(650)에 있어서 제1 스테이지(S_1)는 서브윈도우 발생부(630)로부터 제공되는 k번째 서브윈도우를 입력으로 하여 얼굴검출을 시도하고, 얼굴검출에 실패하면 k번째 서브윈도우는 비얼굴로 결정하고, 얼굴검출에 성공하면 k번째 서브윈도우영상을 제2 스테이지(S_2)로 제공한다. 마지막 스테이지(S_n)에서 k번째 서브윈도우에서 얼굴검출에 성공하면 k번째 서브윈도우를 얼굴로 결정한다. 한편, 각 분류기와 가중치 선택에는 아다부스트(Adaboost)에 기반한 학습알고리즘을 사용할 수 있다. 아다부스트 알고리즘에 의하면 대형 특징 세트로부터 몇개의 중요한 시각적인 특징을 선택하여 매우 효율적인 분류기를 생성한다. 이와 같은 캐스케이드로 연결된 스테이지 구조에 따르면 비얼굴인 경우에는 작은 수의 심플 특징을 사용하여 결정할 수 있으므로 제1 스테이지 또는 제2 스테이지 등과 같은 초기 단계에서 바로 리젝트시키고, 다음 (k+1)번째 서브윈도우를 입력받아 얼굴검출을 시도할 수 있으므로 전체적인 얼굴검출 처리속도를 향상시킬 수 있다.

도 8a는 분류부(650)에 포함되는 각 분류기에서 사용되는 에지 심플특징(811,812) 및 라인 심플특징(813,814,815,816)을 나타낸다. 각 심플특징은 2개 또는 3개의 흰색 또는 검은색 사각형으로 이루어진다. 각 분류기는 심플특징에 따라서, 흰색 사각형내에 위치하는 화소값의 합을 검은색 사각형내에 위치하는 화소값의 합으로부터 감산하여, 감산된 결과값을 심플특징에 대응하는 문턱값과 비교한다. 감산된 결과값과 문턱값과의 비교결과에 따라서 1 또는 -1의 값을 출력한다. 도 8b는 1개의 흰색 사각형과 2개의 검은색 사각형으로 이루어진 라인 심플특징(821) 혹은 1개의 흰색 사각형과 1개의 검은색 사각형으로 이루어진 에지 심플특징(823)을 이용하여 눈 부분을 검출하는 예를 보여준다. 라인 심플특징을 이용하는

경우, 눈 영역이 콧등 영역보다 더 어둡다는 측면을 고려하여 눈 영역과 콧등 영역간의 화소값의 차이를 측정한다. 예지 심플특징을 이용하는 경우, 눈 영역이 뺨 영역보다 더 어둡다는 측면을 고려하여 눈 영역과 뺨 위쪽 영역간의 계조값의 차이를 측정한다. 이와 같이 얼굴을 검출하기 위한 심플특징은 매우 다양하게 존재할 수 있다.

도 9는 도 3에 도시된 얼굴검출부(330)에서 고속 얼굴검출을 위하여 프레임영상을 분할한 예를 보여주는 도면으로서, 썸네일영상 재구성부(610)에 썸네일영상을 입력하기 이전에 얼굴이 존재할 가능성에 따라서 프레임영상을 영역 I 내지 영역 V(910, 930, 950, 970, 990)로 분할한다. 여기서, 각 영역의 분할 위치는 실험 혹은 시뮬레이션을 통해 통계적으로 구해질 수 있다. 대체적으로 영역 I(910)이 얼굴(900)이 위치할 가능성이 가장 큰 영역에 해당하므로, 썸네일영상 재구성부(610)에 영역 I(910)에서부터 영역 IV(970)까지 순차적으로 제공한다.

도 10은 도 3에 도시된 얼굴검출부(330)의 동작을 설명하는 흐름도이다.

도 10을 참조하면, 1011 단계에서는 영역 I(910)에 대한 적분영상을 생성하여 영역 I(910)을 재구성한다. 1013 단계에서는 영역 I(910)에 대하여 적분영상으로 재구성된 영상에서 서브윈도우를 발생시킨다. 1015 단계에서는 발생된 서브윈도우에서 얼굴검출이 성공하였는지를 판단하고, 얼굴검출이 성공한 서브윈도우로 얼굴이 포함된 프레임영상을 구성한다. 1017 단계에서는 1015 단계에서의 판단결과 발생된 서브윈도우에서 얼굴검출이 실패한 경우 영역 I(910)에 대한 서브윈도우의 발생이 종료되었는지를 판단하고, 영역 I(910)에 대한 서브윈도우의 발생이 종료되지 않은 경우 1013 단계로 복귀하고, 영역 I(910)에 대한 서브윈도우의 발생이 종료된 경우 1031 단계로 이행한다.

1031 단계에서는 영역 II(930)에 대한 적분영상을 생성하여 영역 II(930)를 재구성한다. 1033 단계에서는 영역 I(910) 및 영역 II(930)에 대하여 적분영상으로 재구성한 영상에서 서브윈도우를 발생시킨다. 이때, 영역 I(910)에만 위치한 서브윈도우는 제외시킴이 바람직하다. 1035 단계에서는 발생된 서브윈도우에서 얼굴검출이 성공하였는지를 판단하고, 얼굴검출이 성공한 서브윈도우로 얼굴이 포함된 프레임영상을 구성한다. 1037 단계에서는 1035 단계에서의 판단결과 발생된 서브윈도우에서 얼굴검출이 실패한 경우 영역 I(910) 및 영역 II(930)에 대한 서브윈도우의 발생이 종료되었는지를 판단하고, 영역 I(910) 및 영역 II(930)에 대한 서브윈도우의 발생이 종료되지 않은 경우 1033 단계로 복귀하고, 영역 I(910) 및 영역 II(930)에 대한 서브윈도우의 발생이 종료된 경우 1051 단계로 이행한다.

1051 단계에서는 영역 III(950)에 대한 적분영상을 생성하여 영역 III(950)을 재구성한다. 1053 단계에서는 영역 I(910), 영역 II(930), 및 영역 III(950)에 대하여 적분영상으로 재구성한 영상에서 서브윈도우를 발생시킨다. 이때, 영역 I(910) 및 영역 II(930)에만 위치한 서브윈도우는 제외시킴이 바람직하다. 1055 단계에서는 발생된 서브윈도우에서 얼굴검출이 성공하였는지를 판단하고, 얼굴검출이 성공한 서브윈도우로 얼굴이 포함된 프레임영상을 구성한다. 1057 단계에서는 1055 단계에서의 판단결과 발생된 서브윈도우에서 얼굴검출이 실패한 경우 영역 I(910), 영역 II(930), 및 영역 III(950)에 대한 서브윈도우의 발생이 종료되었는지를 판단하고, 영역 I(910), 영역 II(930), 및 영역 III(950)에 대한 서브윈도우의 발생이 종료되지 않은 경우 1053 단계로 복귀하고, 영역 I(910), 영역 II(930), 및 영역 III(950)에 대한 서브윈도우의 발생이 종료된 경우 1071 단계로 이행한다.

1071 단계에서는 영역 IV(970)에 대한 적분영상을 생성하여 영역 IV(970)를 재구성한다. 이때, 영역 I(910), 영역 II(930) 및 영역 III(950)에만 위치한 서브윈도우는 제외시킴이 바람직하다. 1073 단계에서는 영역 I(910), 영역 II(930), 영역 III(950), 및 영역 IV(970)에 대하여 적분영상으로 재구성한 영상의 서브윈도우를 발생시킨다. 1075 단계에서는 발생된 서브윈도우에서 얼굴검출이 성공하였는지를 판단하고, 얼굴검출이 성공한 서브윈도우로 얼굴이 포함된 프레임영상을 구성한다. 1077 단계에서는 1075 단계에서의 판단결과 발생된 서브윈도우에서 얼굴검출이 실패한 경우 영역 I(910), 영역 II(930), 영역 III(950), 및 영역 IV(970)에 대한 서브윈도우의 발생이 종료되었는지를 판단하고, 영역 I(910), 영역 II(930), 영역 III(950), 및 영역 IV(970)에 대한 서브윈도우의 발생이 종료되지 않은 경우 1073 단계로 복귀하고, 영역 I(910), 영역 II(930), 영역 III(950), 및 영역 IV(970)에 대한 서브윈도우의 발생이 종료된 경우 해당 썸네일영상을 얼굴이 포함되지 않은 프레임영상으로 결정한다.

상기 1011 단계, 1031 단계, 1051 단계 및 1071 단계는 썸네일영상 재구성부(610)에서, 상기 1013 단계, 1033 단계, 1053 단계 및 1073 단계는 서브윈도우 발생부(630)에서, 상기 1015 및 1017 단계, 1035 및 1037 단계, 1055 및 1057 단계, 1075 및 1077 단계는 분류부(650)에서 각각 수행됨이 바람직하다.

도 11은 도 3에 있어서 키 프레임 결정부(350)의 세부적인 구성을 보여주는 블록도로서, 옷정보 추출부(1110), 인물 클러스터링부(1130), 및 주요인물 결정부(1150)를 포함하여 이루어진다.

도 11을 참조하면, 옷정보 추출부(1110)는 얼굴검출부(330)에서 얼굴이 검출된 프레임영상에 대하여 옷정보, 예를 들면 옷의 칼라에 대한 히스토그램을 추출한다. 여기서, 얼굴검출부(330)에서 검출된 얼굴영역(도 12의 1210)을 기준으로 하여 설정되는 일정한 검출영역(1230)에 대하여 옷의 칼라에 대한 히스토그램을 추출한다. 옷의 칼라에 대한 히스토그램을 추출하는 이유는 하나의 동영상에서 주요 인물인 경우 통상적으로 동일한 옷을 입고 등장하기 때문이다.

인물 클러스터링부(1130)는 옷정보 추출부(1110)에서 추출되는 옷의 칼라에 대한 히스토그램을 입력으로 하여, 얼굴이 검출된 프레임영상들을 클러스터링하여 복수의 클러스터를 생성한다. 이때, 동일한 옷의 칼라에 대한 히스토그램을 가지는 프레임영상을 동일한 인물이 포함된 프레임영상으로 판단한다.

주요인물 결정부(1150)는 인물 클러스터링부(1130)로부터 제공되는 복수의 클러스터 중에서, 전처리부(310)로부터 제공되는 주요인물의 수를 고려하여 주요인물이 포함된 클러스터를 선택한다. 여기서, 주요인물의 수가 1인 경우에는 가장 많은 프레임영상이 포함된 하나의 클러스터를 선택하고, 주요인물의 복수인 경우에는 포함된 프레임영상이 많은 순서대로 클러스터들을 선택한다. 선택된 클러스터에 포함된 프레임영상들은 각 주제의 키 프레임으로서 시작 샷을 구성하게 된다.

본 발명은 또한 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플라피디스크, 광데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다. 그리고 본 발명을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있다.

발명의 효과

상술한 바와 같이 본 발명에 따른 동영상의 주제별 분할장치 및 방법은, 인물 기반의 고속 연산 알고리즘을 채택하여 동영상을 주제별로 분할함으로써 낮은 사양의 프로세서를 탑재하는 모바일폰, PDA, 디지털 카메라 등과 같은 임베디드 시스템에서도 용이하게 적용할 수 있는 장점이 있다.

또한, 본 발명에 따른 동영상의 주제별 분할장치 및 방법은, 뉴스 장르, 교육 장르, 정보/교양 장르 등과 같이 다수개의 주제별로 편집된 모든 동영상에도 적용할 수 있는 장점이 있다.

또한, 본 발명에 따른 동영상의 주제별 분할장치 및 방법은 임베디드 시스템 뿐만 아니라 블루 레이(Blue-ray) 디스크 레코더 혹은 DCM(Digital Contents Management)용 홈 서버 등과 같은 다양한 제품에 적용되어 사용자의 편의성을 증대시킬 수 있다.

본 발명에 대해 상기 실시예를 참고하여 설명하였으나, 이는 예시적인 것에 불과하며, 본 발명에 속하는 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면의 간단한 설명

도 1은 뉴스 동영상에서 주제별 분할의 예를 보여주는 도면,

도 2는 본 발명에 따른 동영상의 주제별 분할장치의 구성을 나타내는 블록도,

도 3은 도 2에 도시된 시작 샷 결정부의 세부적인 구성을 보여주는 블록도,

도 4a 내지 도 4c는 도 3에 도시된 각 구성요소의 동작을 설명하기 위한 도면,

도 5는 도 3에 있어서 전처리부의 세부적인 구성을 보여주는 블록도,

도 6a는 도 3에 도시된 얼굴검출부의 세부적인 구성을 설명하는 블록도이고, 도 6b는 적분 영상을 구성하는 방법을 설명하는 도면,

도 7은 도 3에 도시된 얼굴검출부에서 사용되는 서브 윈도우의 예를 설명하는 도면,

도 8a 및 도 8b는 도 6a에 도시된 분류부에서 사용되는 특징의 예를 보여주는 도면,

도 9는 도 3에 도시된 얼굴검출부에서 얼굴검출을 위하여 프레임영상을 분할한 예를 보여주는 도면,

도 10은 도 3에 도시된 얼굴검출부의 동작을 설명하는 흐름도,

도 11은 도 3에 있어서 키 프레임 결정부의 세부적인 구성을 보여주는 블록도,

도 12는 도 11에 있어서 옷 정보 추출부의 동작을 설명하는 도면,

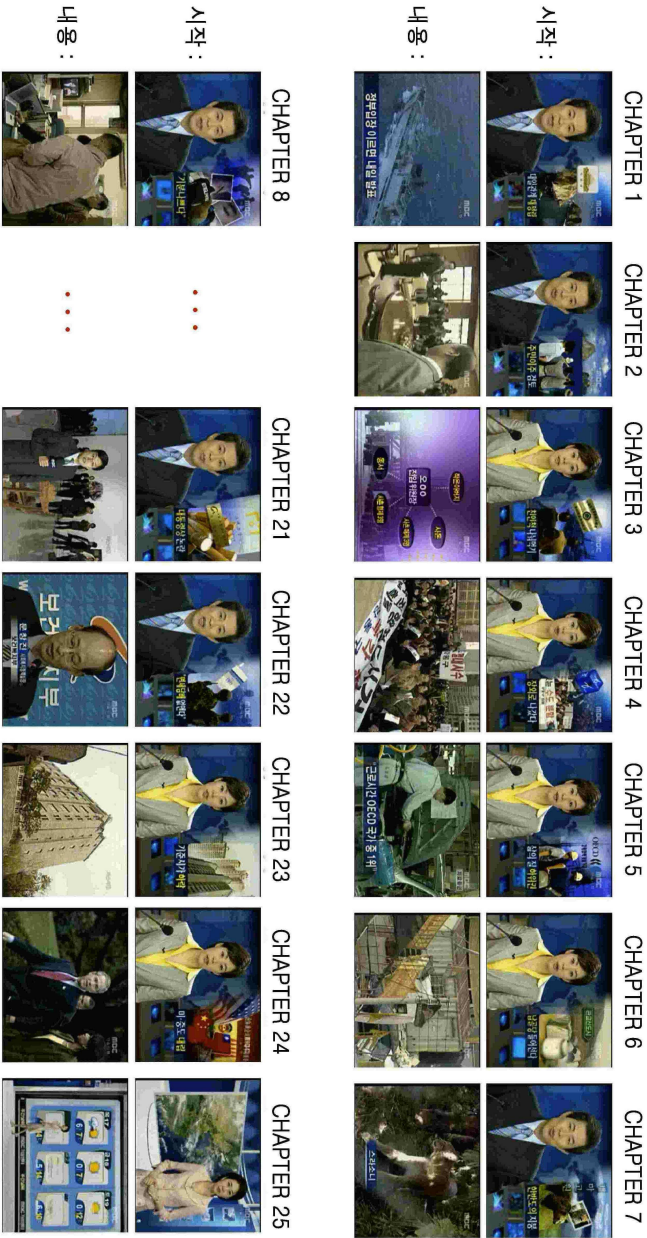
도 13은 뉴스 장르인 경우 본 발명에 따라 시작 샷으로 결정된 키 프레임의 예들을 보여주는 도면,

도 14는 교육 장르인 경우 본 발명에 따라 시작 샷으로 결정된 키 프레임의 예들을 보여주는 도면, 및

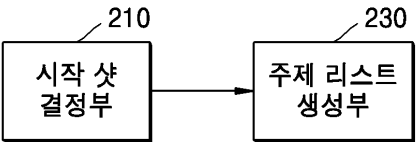
도 15는 정보/교양 장르인 경우 본 발명에 따라 시작 샷으로 결정된 키 프레임의 예들을 보여주는 도면이다.

도면

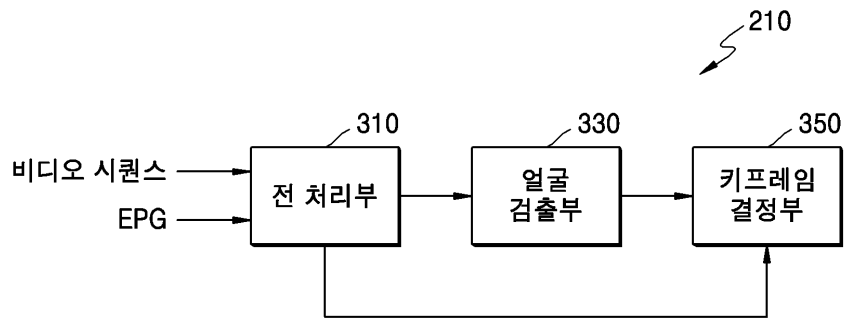
도면1



도면2



도면3



도면4a

HDD	
	Advertisement1.mph
	VJ 특공대.mpg ▷
	Drama1.mpg
	Drama2.mpg ▷
제목 : VJ 특공대 채널 : KBS2 방송 : 4/1 PM 9:55 -4/1 PM 11:05 장르 : 시사/다큐-사회 출연 : 황정민	

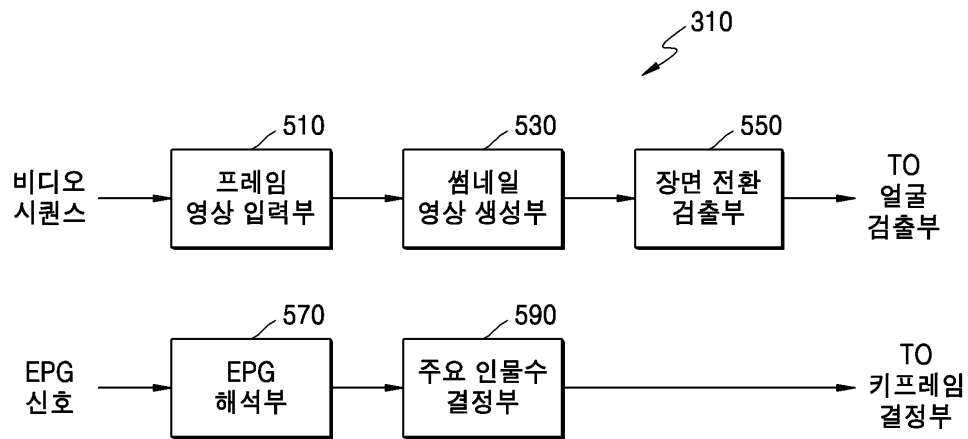
도면4b



도면4c



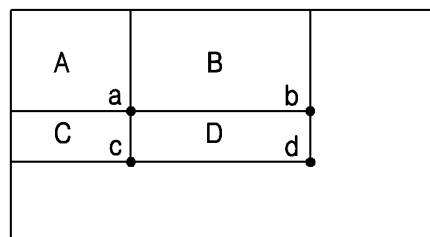
도면5



도면6a



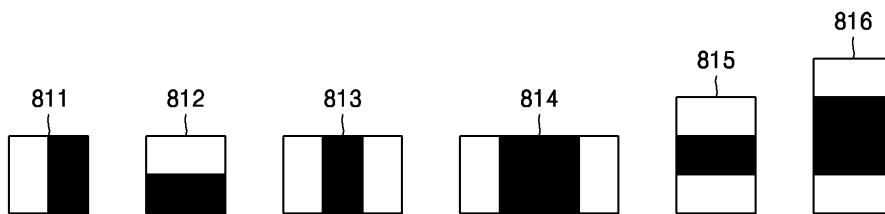
도면6b



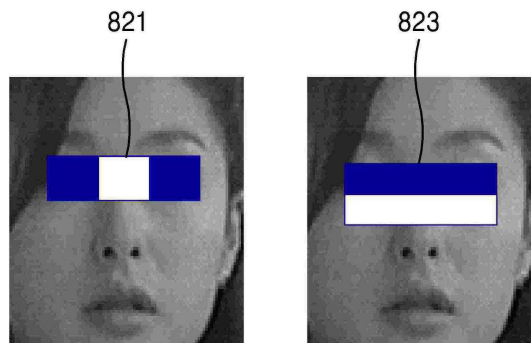
도면7



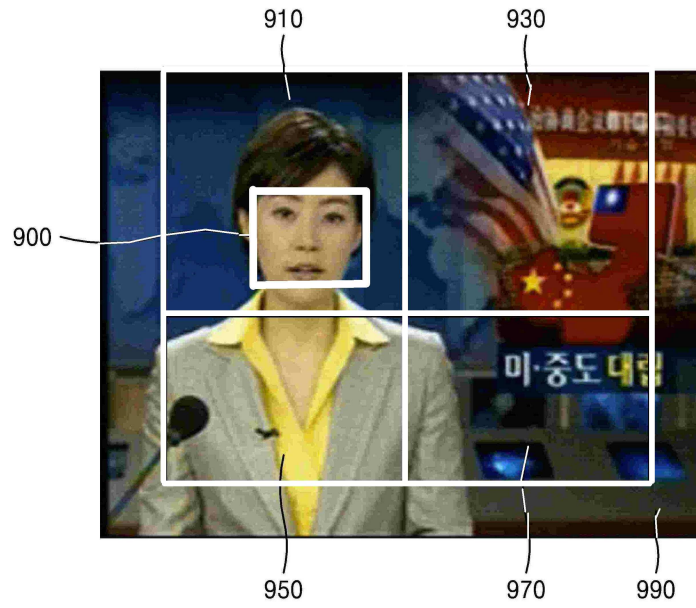
도면8a



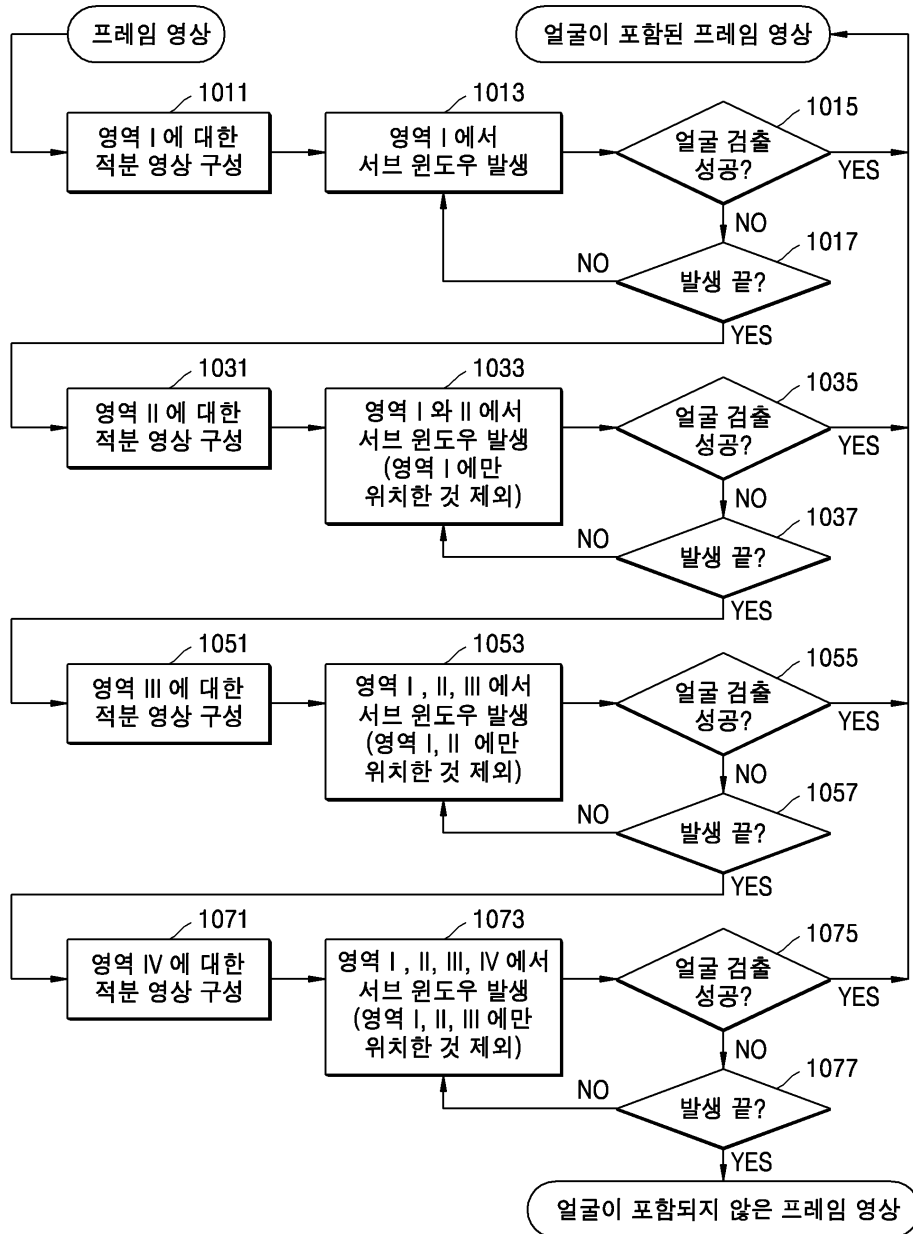
도면8b



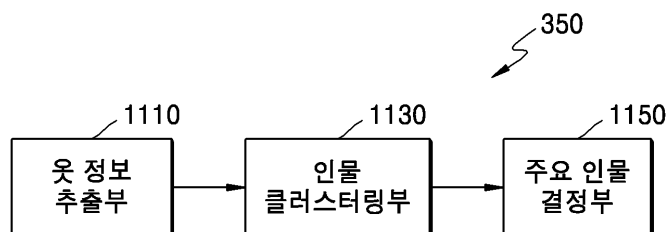
도면9



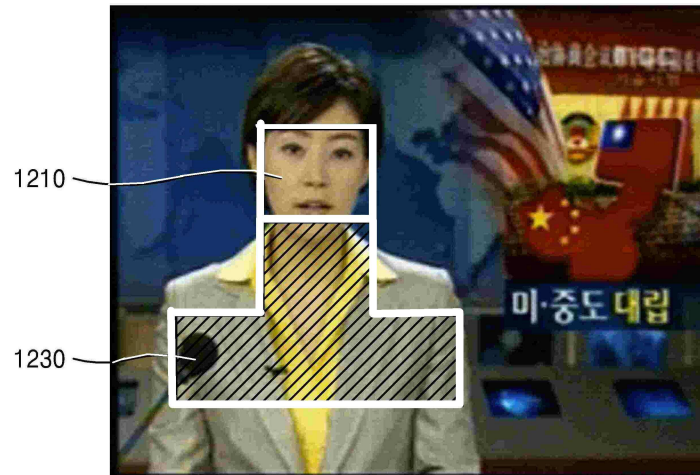
도면10



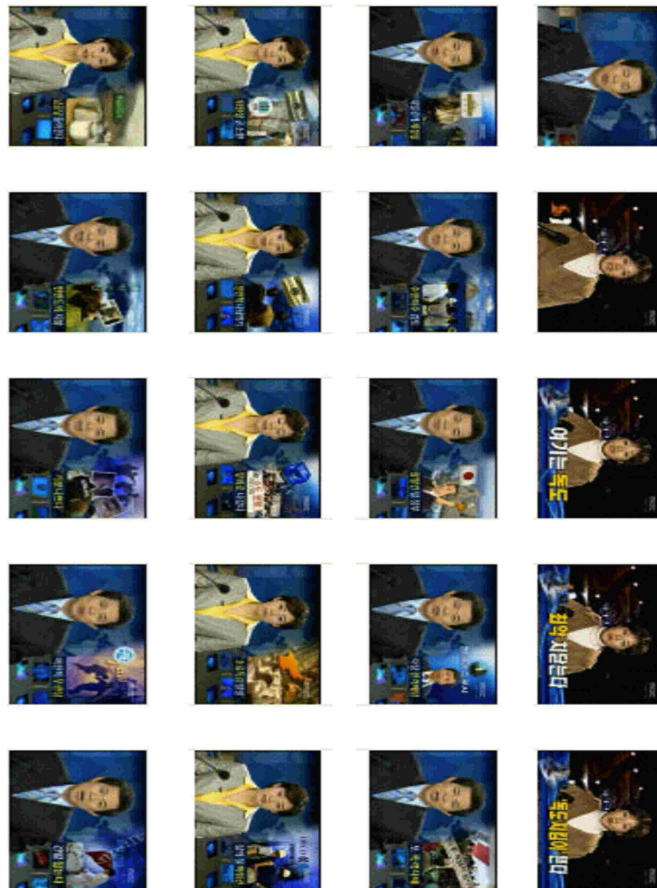
도면11



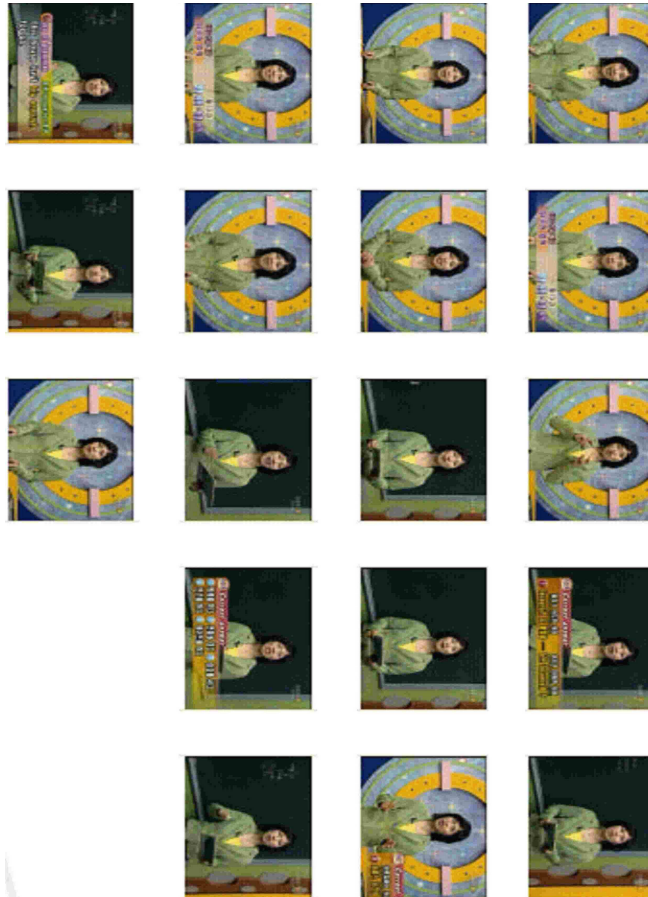
도면12



도면13



도면14



도면15

