

SEGMENTATION METHOD, SEGMENTATION SYSTEM AND NON-TRANSITORY COMPUTER-READABLE MEDIUM

BACKGROUND

Technical Field

[0001] The present disclosure relates to a segmentation method, a segmentation system and non-transitory computer-readable medium. More particularly, the present application relates to a segmentation method and a segmentation system for video source and non-transitory computer-readable medium.

Description of Related Art

[0002] "Online Learning Platform" is a network service that stores a large amount of learning content data in a server, allowing users to connect to the server through the Internet to view the learning content data at any time. At present, the types of learning content data provided by various online learning platforms include videos, audio, presentation, documents or discussion forums, and the like.

[0003] Since the amount of learning content data stored in the online learning platform is very large, in order to be convenient for using, it is necessary to automatically segment the text of the learning content data. Therefore, a system for segmenting automatically according to the correlation between the audio content of the learning video and image content of the learning video is required.

SUMMARY

[0004] An aspect of the present disclosure is a segmentation method. The segmentation method includes operations of: receiving a video data record, wherein the video data record comprises an image data record and an audio data record; performing segmentation processing for the image data record to generate at least one image paragraph mark; performing segmentation processing for the audio data record to generate at least one audio paragraph mark; and calculating a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and generating at least one video content mark.

[0005] Another aspect of the present disclosure is a segmentation system. The segmentation system includes a first server, a storage unit and a processor. The storage unit is configured to store a video data record and at least one video content mark. The processor is electrically connected to the storage unit, and is configured to receive the video data record, wherein the video data record comprises an image data record and an audio data record. The processor further includes an image segmentation unit, an audio segmentation unit and a paragraph mark generating unit. The image segmentation unit is configured to perform segmentation processing for the image data record to generate at least one image paragraph mark. The audio segmentation unit is electrically connected to the image segmentation unit, and is configured to perform segmentation processing for the audio data

record to generate at least one audio paragraph mark. The paragraph mark generating unit is electrically connected to the image segmentation unit and the audio segmentation unit, and is configured to calculate a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and to generate the at least one video content mark result.

[0006] Another aspect of the disclosure is to provide a non-transitory computer-readable medium including one or more sequences of instructions to be executed by a processor for performing a segmentation method, wherein the segmentation method includes operations of: receiving a video data record, wherein the video data record comprises an image data record and an audio data record; performing segmentation processing for the image data record to generate at least one image paragraph mark; performing segmentation processing for the audio data record to generate at least one audio paragraph mark; and calculating a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and generating at least one video content mark.

[0007] Accordingly, since the segmentation method, the segmentation system, and non-transitory computer-readable medium for performing a paragraph marking for the image signal and the audio signal and generating the video content mark according to the paragraph mark of the image signal and the paragraph mark of the audio signal. Therefore, the segmentation system is able to have a function of automatic segmentation the video.

[0008] It is to be understood that both the foregoing general description and the following detailed description are by examples, and are intended to provide further explanation of the disclosure as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The present disclosure can be more fully understood by reading the following detailed description of the embodiment, with reference made to the accompanying drawings as follows:

[0010] FIG. 1 is a functional block diagram illustrating a segmentation system according to an embodiment of the disclosure.

[0011] FIG. 2 is a flow diagram illustrating a segmentation method according to an embodiment of this disclosure.

[0012] FIG. 3 is a flow diagram illustrating step S220 according to an embodiment of this disclosure.

[0013] FIG. 4 is a flow diagram illustrating step S222 according to an embodiment of this disclosure.

[0014] FIG. 5A is a flow diagram illustrating step S223 according to an embodiment of this disclosure.

[0015] FIG. 5B is a flow diagram illustrating step S223 according to an embodiment of this disclosure.

[0016] FIG. 6 is a flow diagram illustrating step S230 according to an embodiment of this disclosure.

DETAILED DESCRIPTION

[0017] It will be understood that, in the description herein and throughout the claims that follow, when an element is referred to as being "connected" or "coupled" to another element, it can be directly connected or coupled to the other element or intervening elements may be present. In contrast, when an element is referred to as being "directly connected" or "directly coupled" to another element, there are no intervening elements present. Moreover, "electrically connect" or "connect" can further refer to the interoperation or interaction between two or more elements.

[0018] It will be understood that, in the description herein and throughout the claims that follow, although the terms "first," "second," etc. may be used to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another. For example, a first element could be termed a second element, and, similarly, a second element could be termed a first element, without departing from the scope of the embodiments.

[0019] It will be understood that, in the description herein and throughout the claims that follow, the terms "comprise" or "comprising," "include" or "including," "have" or "having," "contain" or "containing" and the like used herein are to be understood to be open-ended, i.e., to mean including but not limited to.

[0020] It will be understood that, in the description herein and throughout the claims that follow, the phrase "and/or" includes any and all combinations of one or more of the associated listed items.

[0021] It will be understood that, in the description herein and throughout the claims that follow, words indicating direction used in the description of the following embodiments, such as "above," "below," "left," "right," "front" and "back," are directions as they relate to the accompanying drawings. Therefore, such words indicating direction are used for illustration and do not limit the present disclosure.

[0022] It will be understood that, in the description herein and throughout the claims that follow, unless otherwise defined, all terms (including technical and scientific terms) have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and will not be interpreted in an idealized or overly formal sense unless expressly so defined herein.

[0023] Any element in a claim that does not explicitly state "means for" performing a specified function, or "step for" performing a specific function, is not to be interpreted as a "means" or "step" clause as specified in 35 U.S.C. § 112(f). In particular, the use of "step of" in the claims herein is not intended to invoke the provisions of 35 U.S.C. § 112(f).

[0024] Reference is made to FIG. 1, which is a functional block diagram illustrating a segmentation system 100 according to an embodiment of the disclosure. As shown in FIG. 1, the segmentation system 100 includes a storage unit 110 and a processor 130. The storage unit 110 is electrically connected to the processor 130, and the storage unit 110 is configured to

store a video data record and at least one video content mark and a course database DB.

[0025] Afterwards, the processor 130 includes an image segmentation unit 131, an audio segmentation unit 132, and a paragraph mark generating unit 133. The audio segmentation unit 132 is electrically connected to the image segmentation unit 131 and the paragraph mark generating unit 133. In the embodiment, the storage unit 110 can be implemented by a memory, hardware, flash drive, memory card and etc. The processor 130 can be implemented by a micro controller, a microprocessor, a digital signal processor, an application specific integrated circuit (ASIC), a logical circuitry or any equivalent circuits of the aforesaid components.

[0026] Reference is made to FIG. 2, which is a flow diagram illustrating a segmentation method 200 according to an embodiment of this disclosure. In the embodiment, the segmentation method 200 shown in FIG. 2 can be applied to the segmentation system 100 of FIG. 1, and the processor 130 is configured to perform a paragraph marking for the image signal and the audio signal to generate the video content mark according to the steps described in the following segmentation method 200. As shown in FIG. 2, the segmentation method 200 firstly executes step S210 to receive a video data record. In the embodiment, the video data record includes an image data record and an audio data record. The processor 130 is configured to perform paragraph segmentation for the image data record and the audio data record, respectively. In the next step, firstly, the image data record will be calculated, and then the audio data record will be calculated. However,

the disclosure is not limited thereto. In another the embodiment, firstly, the audio data record will be calculated, and then the image data record will be calculated.

[0027] Afterwards, the segmentation method 200 executes step S220 to perform segmentation processing for the image data record to generate at least one image paragraph mark. In the embodiment, the step 220 further includes steps S221~S223, reference is made to FIG. 3, which is a flow diagram illustrating step S220 according to an embodiment of this disclosure. As the embodiment shown in FIG. 3, the segmentation method 200 executes step S221 to select M unit basis of the image data record, and grouping the selected image data record into a first image paragraph. In the embodiment, M unit basis can be realized as M seconds, wherein M is a positive integer. However, the disclosure is not limited thereto. In another embodiment, M unit basis can be realized as M frames. The M seconds can be assumed as 30 seconds, and thus, in the step, the current paragraph is composed by the 0th second of the image data record to the thirtieth seconds of the image data record. It is noticed that the M can be adjusted by the length of time of the video.

[0028] Afterwards, the segmentation method 200 executes step S222 to determine content of the current image paragraph to generate an image content result. Wherein the image content result includes a dynamic content and a static content. In the embodiment, the step 222 further includes steps S2221~S2223, reference is made to FIG. 4 which is a flow diagram illustrating step S222 according to an embodiment of this disclosure.

As the embodiment shown in FIG. 4, the segmentation method 200 executes step S2221 to select T unit basis of the first image paragraph, and to calculate similarity of the image data record in T unit basis to generate an image difference result. In the embodiment, T unit basis can be realized as T seconds, wherein T is a positive integer. However, the disclosure is not limited thereto. In another embodiment, T unit basis can be realized as T frames. For example, the T seconds can be assumed as 3 seconds, and there are 60 frames per second. The similarity calculation can be realized that the grayscale value of the image of the thirtieth frame of the first second is subtracted from the grayscale value of the image of the thirtieth frame of the 0th second to generate the image difference value of the first second. The image difference value of the first second can be realized as the image difference result and utilized to determine the content of the image. In another embodiment, the image difference value of the second and third seconds can be realized as the image difference result and utilized to determine content of the image. Thus, if more image difference values are utilized, the more accurate content of the image can be determined.

[0029] Afterwards, the segmentation method 200 executes step S2222 if the image difference result is larger than a first image threshold, determining content the current image paragraph as the dynamic content. Based on aforesaid embodiment, if the image difference result is larger than a first image threshold, it can be realized that the image content is change a lot within three seconds, and thus the content the current image paragraph may be the dynamic content. Afterwards, the segmentation method 200 executes step S2223 if the image difference result is less than the first image

threshold, determining content the current image paragraph as the static content. In the embodiment, if the image difference result is less than or equal to the first image threshold, it can be realized that the image content is similar within three seconds, and thus the content the current image paragraph may be the static content, and it means the image content would not be changed.

[0030] Afterwards, the segmentation method 200 executes step S223 to detect a change content for the image data record based on the image content result and to generate the at least one image paragraph mark according to time position of the change content. The change content is, for example but not limited to, between the dynamic content and static content, and it can be determined by the image segmentation unit 131 after calculating a similarity value between the image data record in different time period. In the embodiment, the step 223 further includes steps S2231~S2233a, reference is made to FIG. 5A, which is a flow diagram illustrating step S223 according to an embodiment of this disclosure. As the embodiment shown in FIG. 5A, the segmentation method 200 further executes step S2231a if the current image paragraph is the dynamic content, calculating similarity between an Mth unit basis of the image data record and an (M+1)th unit basis of the image data record to generate an image difference value. Based on aforesaid embodiment, the M seconds can be assumed as 30 seconds, and the current paragraph is composed by the 0th second of the image data record to the thirtieth seconds of the image data record. The Mth unit basis of the image data record is the thirtieth seconds of the image data record, and the (M+1)th unit basis of the image data record

is the thirty-first seconds of the image data record. In this case, the image difference value can be generated that the grayscale value of the image of the thirtieth frame of the thirty-first second is subtracted from the grayscale value of the image of the thirtieth frame of the thirtieth second. The image difference value also can be calculated by the image of other frames, however, the disclosure is not limited thereto.

[0031] Afterwards, the segmentation method 200 further executes step S2232a if the image difference value is larger than a second image threshold, merging the (M+1)th unit basis of the image data record into the current image paragraph. Based on aforesaid embodiment, if aforesaid image difference value is larger than the second image threshold, it means that the image of the next second of the current image paragraph is still belonged to dynamic content, and thus the image data record of the thirty-first second can be merged into the current image paragraph. Afterwards, the segmentation method 200 further executes step S2233a if the image difference value is less than the second image threshold, generating the at least one image paragraph mark in time position of the (M+1)th unit basis of the image data record, and selecting M unit basis of the image data record, and grouping the selected image data record into the next image paragraph. From the above, aforesaid image difference value is less than or equal to the second image threshold, it means that the image of the next second of the current image paragraph may be belonged to static content, and thus the image paragraph mark will be generated in time position of the thirty-first unit basis of the image data record. Then, the

current image paragraph will be composed by the thirty-first second of the image data record to the sixtieth seconds of the image data record.

[0032] Afterwards, the step 223 further includes steps S2231b~S2233b, reference is made to FIG. 5B, which is a flow diagram illustrating step S223 according to an embodiment of this disclosure. As the embodiment shown in FIG. 5B, the segmentation method 200 executes step S2231b if the current image paragraph is the static content, calculating similarity between an Mth unit basis of the image data record and an (M+1)th unit basis of the image data record to generate an image difference value. The operation of the step S2231b is similar with the operation of the step S2231a, and thus, for the sake of brevity, those descriptions will not be repeated here.

[0033] Afterwards, the segmentation method 200 further executes step S2232b if the image difference value is less than the second image threshold, merging the (M+1)th unit basis of the image data record into the current image paragraph. Based on aforesaid embodiment, if aforesaid image difference value is less than or equal to the second image threshold, it means that the image of the next second of the current image paragraph is still belonged to stasuc content, and thus the image data record of the thirty-first second can be merged into the current image paragraph. Afterwards, the segmentation method 200 further executes step S2233b if the image difference value is larger than the second image threshold, generating the at least one image paragraph mark in time position of the (M+1)th unit basis of the image data record, and selecting M unit basis of the image data record, and grouping the selected image data record into the next

image paragraph. From the above, aforesaid image difference value is larger than the second image threshold, it means that the image of the next second of the current image paragraph may be belonged to dynamic content, and thus the image paragraph mark will be generated in time position of the thirty-first unit basis of the image data record. Then, the current image paragraph will be composed by the thirty-first second of the image data record to the sixtieth seconds of the image data record.

[0034] In another embodiment, the similarity of image can be calculated by the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), the texture or color of image, the specific pattern, or other method. However, the disclosure is not limited thereto.

[0035] Afterwards, the segmentation method 200 executes step S230 to perform segmentation processing for the audio data record to generate at least one audio paragraph mark. Wherein the step S230 further includes steps S231~S233, reference is made to FIG. 6, which is a flow diagram illustrating step S230 according to an embodiment of this disclosure. As shown in FIG. 6, the segmentation method 200 further executes step S231 to transform the audio data record into an audio time domain signal and an audio frequency domain signal, respectively. In the embodiment, it can be utilized the Fourier transform to convert the audio signal into the frequency signal. However, the disclosure is not limited thereto. The signal transformed by the Fourier transform is a discontinuous signal, and it can utilized to detect the difference of the tone and timbre to determine the at least one audio paragraph mark.

[0036] Afterwards, the segmentation method 200 executes step S232 to select a time domain section from the audio time domain signal, and to determine whether amplitude of the time domain section is less than a first threshold, if amplitude of the time domain section is less than the first threshold, generating the at least one audio paragraph mark. In the embodiment, it can be utilized the window to select the time domain section from the audio time domain signal. For example, the window can be assumed as five seconds, and thus the time domain section is an audio time domain signal with five seconds. Then, it is determined whether amplitude of the time domain section with five seconds (time domain section) is less than the first threshold, if amplitude of the time domain section is less than the first threshold, it can be realized that the time domain section with five seconds selected by the window can be a silent section, and this situation can be represented that the audio time domain signal happens an interruption. Therefore, when the amplitude of the time domain section is less than the first threshold, the audio paragraph mark can be generated.

[0037] Afterwards, the segmentation method 200 executes step S233 to select a first frequency domain section and a second frequency domain section from the audio frequency domain signal, and to determine whether a difference value of spectral intensity (phase difference) between the first frequency domain section and the second frequency domain section is greater than a second threshold, if the difference value of spectral intensity between the first frequency domain section and the second frequency domain section is larger than the second threshold, generating the at least one audio paragraph mark. In the embodiment, it can be utilized the

window to select the time domain section from the frequency time domain signal. For example, the window can be assumed as five seconds, and thus the first frequency domain section and the second frequency domain section is an audio frequency domain signal with five seconds (the first frequency domain section and the second frequency domain section are selected different sections). The window size (length) used by the audio time domain signal and the audio frequency domain signal can be different. In another embodiment, the window size (length) used by the audio time domain signal and the audio frequency domain signal can be the same. However, the disclosure is not limited thereto. Afterwards, it is determined whether difference value of spectral intensity between the first frequency domain section and the second frequency domain section is greater than the second threshold. If the difference value of spectral intensity is larger than the second threshold, it can be realized that the audio frequency domain signal selected by the window may have different timbre or tone, and this situation can be represented that the audio frequency domain signal includes different voice. Therefore, when the difference value of spectral intensity is larger than the second threshold, the audio paragraph mark can be generated.

[0038] Afterwards, the segmentation method 200 executes step S240 to calculate a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and to generate at least one video content mark. In the embodiment, the step S220 and the step S230 are performed segmentation processing for the image data record and the audio data record, respectively. For example, the image data record is divided into five

paragraphs, and the image paragraph marks includes the first paragraph mark (00:45), the second paragraph mark (01:56), the third paragraph mark (03:25), the fourth paragraph mark (05:10) and the fifth paragraph mark (05:55). The audio data record is divided into four paragraphs, and the audio paragraph marks includes the first paragraph mark (02:02), the second paragraph mark (03:12), the third paragraph mark (04:30), and the fourth paragraph mark (05:00). The threshold can be assumed as 15 seconds, and then the difference between the second paragraph mark (01:56) of the image data record and the first paragraph mark (02:02) of the audio data record is less than the threshold. The video content mark can be calculated by the time average of the second paragraph mark (01:56) of the image data record and the first paragraph mark (02:02) of the audio data record, and therefore the video content marks of the video source includes the first paragraph mark (01:59). Afterwards, the difference between the third paragraph mark (03:25) of the image data record and the second paragraph mark (03:12) of the audio data record is less than the threshold, and the difference between the fourth paragraph mark (05:10) of the image data record and the fourth paragraph mark (05:00) of the audio data record is less than the threshold. Therefore, the video content mark of the second paragraph mark (03:18) and the third paragraph mark (05:05) are generated. Based on aforesaid example, the difference between the first paragraph mark (00:45) of the image data record, the fifth paragraph mark (05:55) of the image data record, and third paragraph mark (04:30) of the audio data record and other paragraph marks are larger than the threshold, and thus the paragraph marks are belonged to the ignored paragraph marks. Finally, the

video content marks are stored in the course database DB of the storage unit 110.

[0039] Based on aforesaid embodiments, the segmentation method, the segmentation system, and non-transitory computer-readable medium are capable of performing a paragraph marking for the image data record and the audio data record and then generating the video content mark according to the paragraph mark of the image data record and the paragraph mark of the audio data record. In some embodiments, the segmentation system is able to have a function of automatic segmentation the learning video.

[0040] The foregoing outlines features of several embodiments so that those skilled in the art may better understand the aspects of the present disclosure. Those skilled in the art should appreciate that they may readily use the present disclosure as a basis for designing or modifying other processes and structures for carrying out the same purposes and/or achieving the same advantages of the embodiments introduced herein. Those skilled in the art should also realize that such equivalent constructions do not depart from the spirit and scope of the present disclosure, and that they may make various changes, substitutions, and alterations herein without departing from the spirit and scope of the present disclosure.

WHAT IS CLAIMED IS

1. A segmentation method, comprising:

receiving a video data record, wherein the video data record comprises an image data record and an audio data record;

performing segmentation processing for the image data record to generate at least one image paragraph mark;

performing segmentation processing for the audio data record to generate at least one audio paragraph mark; and

calculating a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and generating at least one video content mark.

2. The segmentation method of claim 1, wherein performing segmentation processing for the image data record to generate the at least one image paragraph mark, further comprising:

selecting M unit basis of the image data record, and grouping the selected image data record into a first image paragraph;

determining content of the first image paragraph to generate an image content result, wherein the image content result comprises a dynamic content and a static content; and

detecting a change content for the image data record based on the image content result and generating the at least one image paragraph mark according to time position of the change content.

3. The segmentation method of claim 2, wherein determining content of the first image paragraph to generate the image content result, further comprising:

selecting T unit basis of the first image paragraph, and calculating similarity of the image data record in T unit basis to generate an image difference result;

if the image difference result is larger than a first image threshold, determining content the first image paragraph as the dynamic content; and

if the image difference result is less than the first image threshold, determining content the first image paragraph as the static content.

4. The segmentation method of claim 2, wherein detecting the change content for the image data record based on the image content result and generating the at least one image paragraph mark according to time position of the change content, further comprising:

if the first image paragraph is the dynamic content, calculating similarity between an Mth unit basis of the image data record and an (M+1)th unit basis of the image data record to generate an image difference value;

if the image difference value is larger than a second image threshold, merging the (M+1)th unit basis of the image data record into the first image paragraph; and

if the image difference value is less than the second image threshold, generating the at least one image paragraph mark in time position of the (M+1)th unit basis of the image data record, and selecting M unit basis of the

image data record, and grouping the selected image data record into a second image paragraph.

5. The segmentation method of claim 2, wherein detecting the change content for the image data record based on the image content result and generating the at least one image paragraph mark according to time position of the change content, further comprising:

if the first image paragraph is the static content, calculating similarity between an Mth unit basis of the image data record and an (M+1)th unit basis of the image data record to generate an image difference value;

if the image difference value is less than a second image threshold, merging the (M+1)th unit basis of the image data record into the first image paragraph; and

if the image difference value is larger than the second image threshold, generating the at least one image paragraph mark in time position of the (M+1)th unit basis of the image data record, and selecting M unit basis of the image data record, and grouping the selected image data record into a second image paragraph.

6. The segmentation method of claim 1, wherein performing segmentation processing for the audio data record to generate the at least one audio paragraph mark, further comprising:

transforming the audio data record into an audio time domain signal and an audio frequency domain signal, respectively;

selecting a time domain section from the audio time domain signal, and determining whether amplitude of the time domain section is less than a first threshold, if amplitude of the time domain section is less than the first threshold, generating the at least one audio paragraph mark; and

selecting a first frequency domain section and a second frequency domain section from the audio frequency domain signal, and determining whether a difference value of spectral intensity between the first frequency domain section and the second frequency domain section is greater than a second threshold, if the difference value of spectral intensity between the first frequency domain section and the second frequency domain section is larger than the second threshold, generating the at least one audio paragraph mark.

7. A segmentation system, comprising:

a storage unit, configured to store a video data record and at least one video content mark; and

a processor, electrically connected to the storage unit, and configured to receive the video data record, wherein the video data record comprises an image data record and an audio data record, the processor further comprises:

an image segmentation unit, configured to perform segmentation processing for the image data record to generate at least one image paragraph mark;

an audio segmentation unit, electrically connected to the image segmentation unit, and configured to perform segmentation processing for the audio data record to generate at least one audio paragraph mark; and

a paragraph mark generating unit, electrically connected to the image segmentation unit and the audio segmentation unit, and configured to calculate a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and to generate the at least one video content mark.

8. The segmentation system of claim 7, wherein wherein the image segmentation unit is further configured for executing following steps:

selecting M unit basis of the image data record, and grouping the selected image data record into a first image paragraph;

determining content of the first image paragraph to generate an image content result, wherein the image content result comprises a dynamic content and a static content; and

detecting a change content for the image data record based on the image content result and generating the at least one image paragraph mark according to time position of the change content.

9. The segmentation system of claim 8, wherein wherein the image segmentation unit is further configured for executing following steps:

selecting T unit basis of the first image paragraph, and calculating similarity of the image data record in T unit basis to generate an image difference result;

if the image difference result is larger than a first image threshold, determining content the first image paragraph as the dynamic content; and

if the image difference result is less than the first image threshold, determining content the first image paragraph as the static content.

10. The segmentation system of claim 8, wherein the image segmentation unit is further configured for executing following steps:

if the first image paragraph is the dynamic content, calculating similarity between an Mth unit basis of the image data record and an (M+1)th unit basis of the image data record to generate an image difference value;

if the image difference value is larger than a second image threshold, merging the (M+1)th unit basis of the image data record into the first image paragraph; and

if the image difference value is less than the second image threshold, generating the at least one image paragraph mark in time position of the (M+1)th unit basis of the image data record, and selecting M unit basis of the image data record, and grouping the selected image data record into a second image paragraph.

11. The segmentation system of claim 8, wherein the image segmentation unit is further configured for executing following steps:

if the first image paragraph is the static content, calculating similarity between an Mth unit basis of the image data record and an (M+1)th unit basis of the image data record to generate an image difference value;

if the image difference value is less than a second image threshold, merging the (M+1)th unit basis of the image data record into the first image paragraph; and

if the image difference value is larger than the second image threshold, generating the at least one image paragraph mark in time position of the (M+1)th unit basis of the image data record, and selecting M unit basis of the image data record, and grouping the selected image data record into a second image paragraph.

12. The segmentation system of claim 7, wherein the audio segmentation unit is further configured for executing following steps:

transforming the audio data record into an audio time domain signal and an audio frequency domain signal, respectively;

selecting a time domain section from the audio time domain signal, and determining whether amplitude of the time domain section is less than a first threshold, if amplitude of the time domain section is less than the first threshold, generating the at least one audio paragraph mark; and

selecting a first frequency domain section and a second frequency domain section from the audio frequency domain signal, and determining whether a difference value of spectral intensity between the first frequency domain section and the second frequency domain section is greater than a second threshold, if the difference value of spectral intensity between the first frequency domain section and the second frequency domain section is larger than the second threshold, generating the at least one audio paragraph mark.

13. A non-transitory computer-readable medium including one or more sequences of instructions to be executed by a processor for performing a segmentation method, wherein the segmentation method comprises:

receiving a video data record, wherein the video data record comprises an image data record and an audio data record;

performing segmentation processing for the image data record to generate at least one image paragraph mark;

performing segmentation processing for the audio data record to generate at least one audio paragraph mark; and

calculating a difference value between an image marking time of the at least one image paragraph mark and an audio marking time of the at least one audio paragraph mark and generating at least one video content mark.