Summary

[0005] An objective of the disclosure is to provide a new audio retrieval and recognition method and device.

[0006] The objective of the disclosure is achieved by adopting the technical solution as follows. An audio retrieval and recognition method provided by the disclosure includes following steps: acquiring an audio fingerprint of a query audio; and recognizing the query audio according to the audio fingerprint to obtain a recognition result. The audio fingerprint comprises a first part configured for indicating a content feature of the query audio and a second part configured for indicating credibility of the first part~~.~~; wherein the acquiring an audio fingerprint of a query audio comprises: converting the query audio to a spectrogram; determining a feature point in the spectrogram; in the spectrogram, determining one or more masks for the feature point; wherein each of the one or more masks comprises a plurality of spectrum areas; determining mean energy of each of the plurality of spectrum areas; determining an audio fingerprint bit according to the mean energy of the plurality of spectrum areas in each of the one or more masks, further comprising determining the audio fingerprint bit according to a difference value of the mean energy of the plurality of spectrum regions comprised in the one or more masks; judging credibility of the audio fingerprint bit to determine a weight bit, further comprising judging if an absolute value of the difference value attains or exceeds a pre-set weight bit threshold or not, if it attains or exceeds the pre-set weight bit threshold, determining the audio fingerprint bit to be a strong bit, or else determining the audio fingerprint bit to be a weak bit; determining the weight bit according to the audio fingerprint bit is the strong bit or the weak bit; and determining

the audio fingerprint of the query audio according to the audio fingerprint bit and the weight bit.

[0007] The objective of the disclosure can further be achieved by adopting the technical solution as follows.

[0008] In the foregoing audio retrieval and recognition method, the process of acquiring an audio fingerprint of a query audio includes: converting the query audio to a spectrogram; determining a feature point in the spectrogram; in the spectrogram, determining one or more masks for the feature point; determining mean energy of each of the plurality of spectrum areas; determining an audio fingerprint bit according to mean energy of the plurality of spectrum areas in the one or more masks; judging credibility of the audio fingerprint bit to determine a weight bit; and determining the audio fingerprint of the query audio according to the audio fingerprint bit and the weight bit. Each of the one or more masks includes a plurality of spectrum areas.

[0009] In the foregoing audio retrieval and recognition method, the process of converting the query audio to a spectrogram includes: converting the query audio to a two-dimensional spectrogram of time-frequency by a fast Fourier transformation; a value of each point in

a retrieval and recognition system, configured for recognizing the query audio according to the audio fingerprint to obtain a recognition result. The audio fingerprint includes a first part configured for indicating a content feature of the query audio and a second part configured for indicating credibility of the first part. Acquiring an audio fingerprint of a query audio comprises: converting the query audio to a spectrogram; determining a feature point in the spectrogram; in the spectrogram, determining one or more masks for the feature point; wherein each of the one or more masks comprises a plurality of spectrum areas; determining mean energy of each of the plurality of spectrum areas; determining an audio fingerprint bit according to the mean energy of the plurality of spectrum areas in each of the one or more masks, further comprising determining the audio fingerprint bit according to a difference value of the mean energy of the plurality of spectrum regions comprised in the one or more masks; judging credibility of the audio fingerprint bit to determine a weight bit, further comprising judging if an absolute value of the difference value attains or exceeds a pre-set weight bit threshold or not, if it attains or exceeds the pre-set weight bit threshold, determining the audio fingerprint bit to be a strong bit, or else determining the audio fingerprint bit to be a weak bit; determining the weight bit according to the audio fingerprint bit is the strong bit or the weak bit; and determining the audio fingerprint of the query audio according to the audio fingerprint bit and the weight bit.

[0038] The objective of the disclosure is achieved by adopting the technical solution as follows.

[0039] The foregoing audio retrieval and recognition device further includes a module executing steps in any audio retrieval and recognition

method described above.

[0040] The objective of the disclosure is achieved by adopting the technical solution as follows. An audio retrieval and recognition hardware device provided by the disclosure includes a memory configured for storing an executable instruction; and a processor configured for executing the computer readable instruction for the computer to implement any audio retrieval and recognition method above.

[0041] The objective of the disclosure is achieved by adopting the technical solution as follows. A computer readable storage medium provided by the disclosure configured for storing a non-transitory computer readable instruction; when the non-transitory computer readable instruction is executed by a computer, the computer executes any audio retrieval and recognition method above.

[0042] The objective of the disclosure is achieved by adopting the technical solution as follows. A terminal device provided by the disclosure includes any audio retrieval and recognition device above.

[0043] The illustration above only is the overview of the technical solution of the disclosure. In order to clearly describe the technological solutions of the disclosure and follow the specification, as well as better

Claims

1. An audio retrieval and recognition method, wherein the audio retrieval and recognition method comprises:

acquiring an audio fingerprint of a query audio; wherein the audio fingerprint comprises a first part configured for indicating a content feature of the query audio and a second part configured for indicating credibility of the first part; and

recognizing the query audio according to the audio fingerprint to obtain a recognition result.;

2. The audio retrieval and recognition method according to claim 1, wherein the acquiring an audio fingerprint of a query audio comprises:

converting the query audio to a spectrogram;

determining a feature point in the spectrogram;

in the spectrogram, determining one or more masks for the feature point; wherein each of the one or more masks comprises a plurality of spectrum areas;

determining mean energy of each of the plurality of spectrum areas;

determining an audio fingerprint bit according to the mean energy of the plurality of spectrum areas in each of the one or more masks, further comprising determining the audio fingerprint bit according to a difference value of the mean energy of the plurality of spectrum regions comprised in the one or more masks;

judging credibility of the audio fingerprint bit to determine a weight bit, further comprising judging if an absolute value of the difference value attains or exceeds a pre-set weight bit threshold or not, if it attains or exceeds the pre-set weight bit threshold, determining the audio fingerprint

bit to be a strong bit, or else determining the audio fingerprint bit to be a weak bit; determining the weight bit according to the audio fingerprint bit is the strong bit or the weak bit; and

determining the audio fingerprint of the query audio according to the audio fingerprint bit and the weight bit.

~~3~~2. The audio retrieval and recognition method according to claim 1~~2~~, wherein the converting the query audio to a spectrogram comprises: converting the query audio to a two-dimensional spectrogram of time-frequency by a fast Fourier transformation; wherein a value of each point in the spectrogram indicates energy of the query audio.

~~4~~3. The audio retrieval and recognition method according to claim ~~3~~2, wherein the converting the query audio to a spectrogram further comprises: processing the spectrogram by MEL transformation.

~~5~~4. The audio retrieval and recognition method according to claim ~~3~~2, wherein the feature point is a fixed point in the spectrogram.

~~6~~5. The audio retrieval and recognition method according to claim ~~5~~4, wherein the feature point is a point whose frequency is equal to a plurality of preset frequency set values.

6~~7~~. The audio retrieval and recognition method according to claim 2~~3~~, wherein the feature point is a maximal energy point in the spectrogram, or the feature point is a minimal energy point in the spectrogram.

7~~8~~. The audio retrieval and recognition method according to claim 1~~2~~, wherein the plurality of spectrum regions comprised in the one or more mask are symmetrical distributed.

~~9~~8. The audio retrieval and recognition method according to claim 7~~8~~, wherein the plurality of spectrum regions comprised in the one or more mask have an identical frequency range, and/or an identical time

range, and/or are centrosymmetric with the feature point as a center.

~~10~~9. The audio retrieval and recognition method according to claim ~~2~~1, wherein the mean energy of the plurality of spectrum regions is a mean value of energy values of all points comprised in the plurality of spectrum regions.

~~11. The audio retrieval and recognition method according to claim 2, wherein the determining an audio fingerprint bit according to the mean energy of the plurality of spectrum areas in the one or more masks comprises:~~

~~determining the audio fingerprint bit according to a difference value of the mean energy of the plurality of spectrum regions comprised in the one or more masks.~~

~~12. The audio retrieval and recognition method according to claim 11, wherein the judging credibility of the audio fingerprint bit to determine a weight bit comprises:~~

~~judging if an absolute value of the difference value attains or exceeds a pre-set weight bit threshold or not, if it attains or exceeds the pre-set weight bit threshold, determining the audio fingerprint bit to be a strong bit, or else determining the audio fingerprint bit to be a weak bit; determining the weight bit according to the audio fingerprint bit is the strong bit or the weak bit.~~

~~13~~10. The audio retrieval and recognition method according to claim ~~2~~1, wherein the audio retrieval and recognition method further comprises:

dividing the query audio into a plurality of sub-audio according to time;

extracting the audio fingerprint of each of the plurality of sub-audio; and

combining the audio fingerprints of the plurality of sub-audio to

obtain the audio fingerprint of the query audio.

~~14~~11. The audio retrieval and recognition method according to claim 1~~2~~, wherein the audio fingerprint of the query audio is defined as a first audio fingerprint; the first audio fingerprint comprises a plurality of first audio fingerprint units and first weight units corresponding to the plurality of first audio fingerprint units; the first audio fingerprint units comprise the plurality of audio fingerprint bits of the query audio; the first weight units comprise the plurality of weight bits corresponding to the plurality of audio fingerprint bits.

~~15~~12. The audio retrieval and recognition method according to claim 1~~4~~, wherein the recognizing the query audio according to the audio fingerprint comprises:

carrying out a first ranking on a plurality of known audio according to each of the plurality of first audio fingerprint units, according to a result of the first rank, selecting first k individuals of the plurality of known audio as a first candidate audio collection, where k is a positive integer; and

carrying out a second ranking on the first candidate audio collection according to the first audio fingerprint units sequentially arranged, according to a result of the second rank, selecting first n individuals of the first candidate audio collection as the recognition result, where n is a positive integer.

~~16~~13. The audio retrieval and recognition method according to claim ~~15~~12, wherein the audio retrieval and recognition method further comprises:

pre-acquiring audio fingerprint of the plurality of known audio as second audio fingerprints; wherein the second audio fingerprints comprise a plurality of second audio fingerprint units and second weight

units corresponding to the plurality of second audio fingerprint units; and

pre-acquiring fingerprint indexes of the plurality of known audio by indexing the second audio fingerprints.

~~17~~14. The audio retrieval and recognition method according to claim ~~16~~13, wherein during carrying out the first ranking or the second ranking, according to the first weight units and/or the second weight units, the plurality of first audio fingerprint units and/or the plurality of second audio fingerprint units are weighted.

~~18~~15. The audio retrieval and recognition method according to claim ~~16~~13, wherein the carrying out a first ranking on the plurality of known audio according to each of the plurality of first audio fingerprint units comprises: carrying out term frequency – inverse document frequency ranking on the plurality of known audio according to each of the plurality of first audio fingerprint units.

~~19~~16. The audio retrieval and recognition method according to claim ~~18~~15, wherein the carrying out term frequency – inverse document frequency ranking on the plurality of known audio according to each of the plurality of first audio fingerprint units comprises:

matching the fingerprint indexes of the plurality of known audio and the plurality of first audio fingerprint units to carry out the term frequency – inverse document frequency ranking on the plurality of known audio.

~~20~~17. The audio retrieval and recognition method according to claim ~~19~~16, wherein the pre-acquiring fingerprint indexes of the plurality of known audio comprises:

according to the plurality of second weight units, pre-acquiring a forward index and/or an inverted index of the plurality of known audio.

~~21~~18. The audio retrieval and recognition method according to claim ~~19~~16, wherein the matching the fingerprint indexes of the plurality of

known audio and the plurality of first audio fingerprint units comprises:

according to the plurality of first weight units, ~~absolutely~~ matching the fingerprint indexes of the audio and the first audio fingerprint units.

~~22~~19. The audio retrieval and recognition method according to claim ~~16~~13, wherein the carrying out a second ranking on the first candidate audio collection according to the first audio fingerprint units sequentially arranged comprises:

acquiring a similarity matrix of audio in the first candidate audio collection according to the fingerprint indexes of the plurality of known audio and the first audio fingerprint, and ranking the audio in the first candidate audio collection according to the similarity matrix.

~~23~~20. The audio retrieval and recognition method according to claim ~~22~~19, wherein the acquiring a similarity matrix of audio in the first candidate audio collection according to the fingerprint indexes of the plurality of known audio and the first audio fingerprint, and ranking the audio in the first candidate audio collection according to the similarity matrix comprises:

weighting by utilizing the plurality of first weight units and/or the plurality of second weight units to obtain a weighted similarity matrix, ranking the audio in the first candidate audio collection according to the weighted similarity matrix.

~~24~~21. The audio retrieval and recognition method according to claim ~~22~~19, wherein the ranking the audio in the first candidate audio collection according to the similarity matrix comprises:

ranking the audio in the first candidate audio collection according to a straight line in the similarity matrix.

~~25~~22. The audio retrieval and recognition method according to claim ~~22~~19, wherein the acquiring an audio fingerprint of a query audio further

comprises:

acquiring various types of first audio fingerprints of the query audio;

wherein the pre-acquiring audio fingerprint of the plurality of known audio as second audio fingerprints comprises: acquiring various types of second audio fingerprints of audio in the first candidate audio collection;

wherein the acquiring a similarity matrix of audio in the first candidate audio collection according to the fingerprint indexes of the plurality of known audio and the first audio fingerprint comprises: determining the similarity matrix according to the various types of first audio fingerprints and the various types of second audio fingerprints.

~~26~~23. The audio retrieval and recognition method according to claim ~~25~~22, wherein each of the various types of first audio fingerprints comprises a plurality of first audio fingerprint units; each of the various types of second audio fingerprints comprises a plurality of second audio fingerprint units;

wherein the determining the similarity matrix according to the various types of first audio fingerprints and the various types of second audio fingerprints comprises:

respectively determining unit similarities between the second audio fingerprint units of an identical type and the plurality of first audio fingerprint units to obtain the unit similarities of various types; and determining the similarity matrix according to a mean value or a minimal value of the various types of unit similarities.

~~27~~24. The audio retrieval and recognition method according to claim ~~16~~13, wherein the audio retrieval and recognition method further comprises:

pre-slicing the query audio and the plurality of known audio according to a preset time span to obtain a plurality of query sub-audio

and a plurality of known sub-audio, respectively extracting audio fingerprints from the plurality of query sub-audio and the plurality of known sub-audio to obtain a plurality of first sub-audio fingerprints and a plurality of second sub-audio fingerprints whose lengths are identical.

~~28~~25. The audio retrieval and recognition method according to claim ~~16~~13, wherein the audio retrieval and recognition method further comprises:

before carrying out the first rank, slicing the first audio fingerprints of the obtained query audio and the second audio fingerprints of the plurality of known audio according to a preset length to obtain a plurality of first sub-audio fingerprints and a plurality of second sub-audio fingerprints whose lengths are identical.

~~29~~26. The audio retrieval and recognition method according to claim ~~22~~19, wherein the plurality of first audio fingerprint units are arranged in the first audio fingerprint according to a time sequence; the plurality of second audio fingerprint units are arranged in the second audio fingerprint according to the time sequence.

~~30~~27. The audio retrieval and recognition method according to claim ~~29~~26, wherein the audio retrieval and recognition method further comprises:

determining a repetition fragment of audio in query audio and the recognition result according to the similarity matrix.

~~31~~28. An audio retrieval and recognition device, wherein the audio retrieval and recognition device comprises:

an audio fingerprint acquirement system, configured for acquiring an audio fingerprint of query audio; wherein the audio fingerprint comprises a first part configured for indicating a content feature of the query audio and a second part configured for indicating credibility of the first part;

and

a retrieval and recognition system, configured for recognizing the query audio according to the audio fingerprint to obtain a recognition result;

wherein the acquiring an audio fingerprint of a query audio comprises:

converting the query audio to a spectrogram;

determining a feature point in the spectrogram;

in the spectrogram, determining one or more masks for the feature point; wherein each of the one or more masks comprises a plurality of spectrum areas;

determining mean energy of each of the plurality of spectrum areas;

determining an audio fingerprint bit according to the mean energy of the plurality of spectrum areas in each of the one or more masks, further comprising determining the audio fingerprint bit according to a difference value of the mean energy of the plurality of spectrum regions comprised in the one or more masks;

judging credibility of the audio fingerprint bit to determine a weight bit, further comprising judging if an absolute value of the difference value attains or exceeds a pre-set weight bit threshold or not, if it attains or exceeds the pre-set weight bit threshold, determining the audio fingerprint bit to be a strong bit, or else determining the audio fingerprint bit to be a weak bit; determining the weight bit according to the audio fingerprint bit is the strong bit or the weak bit; and

determining the audio fingerprint of the query audio according to the audio fingerprint bit and the weight bit.

~~32~~29. The audio retrieval and recognition device according to claim ~~31~~28, wherein the audio retrieval and recognition device further

comprises a module executing steps in any of claims 2 to ~~30~~27.

~~33~~30. An audio retrieval and recognition hardware device, comprising:

a memory, configured for storing an executable instruction; and

a processor, configured for implementing the audio retrieval and recognition method of any one of claims 1 to ~~30~~ 27 when the executable instruction is performed by the processor.

~~34~~31. A computer readable storage medium, configured for storing a non-transitory computer readable instruction; wherein when the non-transitory computer readable instruction is executed by a computer, the computer executes the audio retrieval and recognition method according to any of claims 1 to ~~30~~27.

~~35~~32. A terminal device, comprising the audio retrieval and recognition device according to claim ~~31~~ 28 or claim ~~32~~29.