CLAIMS

1. A method for video processing, comprising:

determining preselected videos associated with a query text paragraph in a plurality of videos according to paragraph information of the query text paragraph and video information of the plurality of videos in a video library; and

determining a target video in the preselected videos according to video frame information of the preselected videos and sentence information of the query text paragraph,

wherein the paragraph information comprises second feature information of the query text paragraph, and the video information comprises third feature information of the videos, the sentence information comprises first feature information of one or more sentences of the query text paragraph, and the video frame information comprises fourth feature information of a plurality of video frames of the preselected videos.

2. The method according to claim 1, wherein

determining the preselected videos associated with the query text paragraph in the plurality of videos according to the paragraph information of the query text paragraph and the video information of the plurality of videos in the video library comprises:

determining the preselected videos associated with the query text paragraph in the plurality of videos according to the second feature information and the third feature information of the plurality of videos in the video library.

3. The method according to claim 2, wherein determining the preselected videos associated with the query text paragraph in the plurality of videos according to the second feature information and the third feature information of the plurality of videos in the video library comprises:

determining first correlation scores between the query text paragraph and the plurality of videos respectively according to the second feature information and the third feature information of the plurality of videos in the video library; and

determining the preselected videos in the plurality of videos according to the first correlation scores.

4. The method according to claim 3, wherein determining the first correlation scores between the query text paragraph and the plurality of videos respectively according to the second feature information and the third feature information of the plurality of videos in the video library comprises:

mapping third feature information of a first video and the second feature information to a vector space of a same dimension, to obtain a third feature vector of the first video and a second feature vector of the query text paragraph, wherein the first video is any one of the plurality of videos; and

determining a cosine similarity between the second feature vector and the third feature vector as a first correlation score between the query text paragraph and the first video.

5. The method according to any one of claims 1 to 4, wherein

determining the target video in the preselected videos according to the video frame information of the preselected videos and the sentence information of the query text paragraph comprises:

determining the target video in the preselected videos according to the first feature information of the one or more sentences and the fourth feature information of the plurality of video frames of the preselected videos.

6. The method according to claim 5, wherein determining the target video in the preselected videos according to the first feature information of the one or more sentences and the fourth feature information of the plurality of video frames of the preselected videos comprises:

determining second correlation scores between the query text paragraph and the preselected videos according to the first feature information of the one or more sentences and the fourth feature information of the plurality of video frames of the preselected videos; and

determining the target video in the preselected videos according to the first correlation scores and the

second correlation scores.

7. The method according to claim 6, wherein determining the second correlation scores between the query text paragraph and the preselected videos according to the first feature information of the one or more sentences and the fourth feature information of the plurality of video frames of the preselected videos comprises:

mapping fourth feature information of a plurality of video frames of a target preselected video and the first feature information of the one or more sentences to a vector space of a same dimension, to obtain fourth feature vectors of the plurality of video frames of the target preselected video and first feature vectors of the one or more sentences, respectively, wherein the target preselected video is any one of the preselected videos;

determining, from the fourth feature vectors, target feature vectors whose cosine similarities to a first feature vector of a target sentence are greater than or equal to a similarity threshold, wherein the target sentence is any one of the one or more sentences;

aggregating video frames corresponding to the target feature vectors into a video clip corresponding to the target sentence;

determining a fifth feature vector of the video clip corresponding to the target sentence according to the target feature vectors; and

determining a second correlation score between the query text paragraph and the target preselected video according to fifth feature vectors of video clips respectively corresponding to the one or more sentences and the first feature vectors of the one or more sentences.

8. The method according to claim 6, wherein determining the target video in the preselected videos according to the first correlation scores and the second correlation scores comprises:

determining products of the first correlation scores and the second correlation scores as third correlation scores; and

determining the target video in the preselected videos according to the third correlation scores.

9. The method according to any one of claims 1 to 4, further comprising:

performing feature extraction processing on a plurality of video frames of a second video respectively to obtain fourth feature information of the plurality of video frames of the second video, wherein the second video is any one of the plurality of videos; and

determining third feature information of the second video according to the fourth feature information of the plurality of video frames of the second video.

10. The method according to any one of claims 1 to 4, further comprising:

performing feature extraction processing on the one or more sentences of the query text paragraph respectively to obtain the first feature information of the one or more sentences; and

determining the second feature information of the query text paragraph according to the first feature information of the one or more sentences of the query text paragraph.

11. A device for video processing, comprising:

a preselected video determining module, configured to determine preselected videos associated with a query text paragraph in a plurality of videos according to paragraph information of the query text paragraph and video information of the plurality of videos in a video library; and

a target video determining module, configured to determine a target video in the preselected videos according to video frame information of the preselected videos and sentence information of the query text paragraph,

wherein the paragraph information comprises second feature information of the query text paragraph, and the video information comprises third feature information of the videos, the sentence information comprises first feature information of one or more sentences of the query text paragraph, and the video frame information comprises fourth feature information of a plurality of video frames of the preselected videos.

12. The device according to claim 11, wherein

the preselected video determining module is further configured to:

determine the preselected videos associated with the query text paragraph in the plurality of videos according to the second feature information and the third feature information of the plurality of videos in the video library.

13. The device according to claim 12, wherein the preselected video determining module is further configured to:

determine first correlation scores between the query text paragraph and the plurality of videos respectively according to the second feature information and the third feature information of the plurality of videos in the video library; and

determine the preselected videos in the plurality of videos according to the first correlation scores.

14. The device according to claim 13, wherein the preselected video determining module is further configured to:

map third feature information of a first video and the second feature information to a vector space of the same dimension, to obtain a third feature vector of the first video and a second feature vector of the query text paragraph, wherein the first video is any one of the plurality of videos; and

determine a cosine similarity between the second feature vector and the third feature vector as a first correlation score between the query text paragraph and the first video.

15. The device according to any one of claims 11 to 14, wherein

the target video determining module is further configured to:

determine the target video in the preselected videos according to the first feature information of the one or more sentences and the fourth feature information of the plurality of video frames of the preselected videos.

16. The device according to claim 15, wherein the target video determining module is further configured to:

determine second correlation scores between the query text paragraph and the

preselected

videos according to the first feature information of the one or more sentences and the fourth feature information of the plurality of video frames of the preselected videos; and

determine the target video in the preselected videos according to the first correlation scores and the second correlation scores.

17. The device according to claim 16, wherein the target video determining module is further configured to:

map fourth feature information of a plurality of video frames of a target preselected video and the first feature information of the one or more sentences to a vector space of the same dimension, to obtain fourth feature vectors of the plurality of video frames of the target preselected video and first feature vectors of the one or more sentences, respectively, wherein the target preselected video is any one of the preselected videos;

determine, from the fourth feature vectors, target feature vectors whose cosine similarities to a first feature vector of a target sentence are greater than or equal to a similarity threshold, wherein the target sentence is any one of the one or more sentences;

aggregate video frames corresponding to the target feature vectors into a video clip corresponding to the target sentence;

determine a fifth feature vector of the video clip corresponding to the target sentence according to the target feature vector; and

determine a second correlation score between the query text paragraph and the target preselected video according to fifth feature vectors of video clips respectively corresponding to the one or more sentences and the first feature vectors of the one or more sentences.

18. The device according to claim 16, wherein the target video determining module is further configured to:

determine products of the first correlation scores and the second correlation scores

as third correlation scores; and

determine the target video in the preselected videos according to the third correlation scores.

19. The device according to any one of claims 11 to 14, further comprising:

a video feature extracting module, configured to perform feature extraction processing on a plurality of video frames of a second video respectively to obtain fourth feature information of the plurality of video frames of the second video, wherein the second video is any one of the plurality of videos; and

a determining module, configured to determine third feature information of the second video according to the fourth feature information of the plurality of video frames of the second video.

20. The device according to any one of claims 11 to 14, further comprising:

a sentence feature extracting module, configured to perform feature extraction processing on the one or more sentences of the query text paragraph respectively to obtain the first feature information of the one or more sentences; and

a second determining module, configured to determine the second feature information of the query text paragraph according to the first feature information of the one or more sentences of the query text paragraph.

21. An electronic device, comprising:

a processor; and

a memory configured to store processor-executable instructions,

wherein the processor is configured to: execute the method according to any one of claims 1 to 10.

22. A computer-readable storage medium, having computer program instructions stored thereon, wherein when the computer program instructions are executed by a processor, the method according to any one of claims 1 to 10 is implemented.