

CLAIMS

1. A method for anti-spoofing detection, comprising:

obtaining at least one image subsequence from an image sequence, wherein the image sequence is acquired by an image acquisition apparatus after a user is prompted to read a specified content, and the image subsequence comprises at least one image in the image sequence;

performing lipreading on the at least one image subsequence to obtain a lipreading result of the at least one image subsequence; and

determining an anti-spoofing detection result based on the lipreading result of the at least one image subsequence.

2. The method according to claim 1, wherein obtaining the at least one image subsequence from the image sequence comprises:

obtaining the at least one image subsequence from the image sequence according to a segmentation result of audio corresponding to the image sequence.

3. The method according to claim 2, wherein the segmentation result of the audio comprises: an audio segment corresponding to each of at least one character included in the specified content; and

obtaining the at least one image subsequence from the image sequence according to the segmentation result of the audio corresponding to the image sequence comprises:

obtaining from the image sequence the image subsequence corresponding to each character according to time information of the audio segment corresponding to each character in the specified content.

4. The method according to claim 2 or 3, further comprising:

obtaining the audio corresponding to the image sequence; and

segmenting the audio to obtain at least one audio segment, wherein each of the

at least one audio segment corresponds to one character in the specified content.

5. The method according to any one of claims 1 to 4, wherein performing lipreading on the at least one image subsequence to obtain the lipreading result of the at least one image subsequence comprises at least one of:

- obtaining lip region images from at least two target images included in the image subsequence; and

- obtaining the lipreading result of the image subsequence based on the lip region images of the at least two target images,

- or,

- obtaining lip morphology information of the at least two target images included in the image subsequence; and

- obtaining the lipreading result of the image subsequence based on the lip morphology information of the at least two target images.

6. The method according to claim 5, wherein obtaining the lip region images from the at least two target images included in the image subsequence comprises:

- performing key point detection on the at least two target images to obtain information of face key points, wherein the information of the face key points comprises position information of lip key points; and

- obtaining the lip region images from the at least two target images based on the position information of the lip key points.

7. The method according to claim 5 or 6, wherein obtaining the lipreading result of the image subsequence based on the lip region images of the at least two target images comprises:

- performing recognition processing on the input lip region images of the at least two target images by using a first neural network to output the lipreading result of the image subsequence.

8 The method according to claim 5, wherein obtaining the lip morphology information of the at least two target images included in the image subsequence comprises:

performing feature extraction processing on a lip region image obtained from each of the at least two target images to obtain a lip morphology feature of each target image, wherein the lip morphology information of the target image comprises the lip morphology feature.

9. The method according to any one of claims 5 to 8, further comprising:

selecting a first image that satisfies a predetermined quality standard from the image subsequence; and

determining the first image and at least one second image adjacent to the first image as the at least two target images.

10. The method according to claim 9, wherein the at least one second image comprises at least one image that is before the first image and adjacent to the first image, and comprises at least one image that is after the first image and adjacent to the first image.

11. The method according to any one of claims 1 to 10, wherein each of the at least one image subsequence corresponds to one character in the specified content.

12. The method according to any one of claims 1 to 11, wherein determining the anti-spoofing detection result based on the lipreading result of the at least one image subsequence comprises:

fusing the lipreading result of the at least one image subsequence to obtain a fusion recognition result;

determining whether the fusion recognition result matches a voice recognition result of the audio corresponding to the image sequence; and

determining the anti-spoofing detection result based on a matching result

between the fusion recognition result and the voice recognition result of the audio.

13. The method according to claim 12, wherein fusing the lipreading result of the at least one image subsequence to obtain the fusion recognition result comprises:

fusing, based on the voice recognition result of the audio corresponding to the image sequence, the lipreading result of the at least one image subsequence to obtain the fusion recognition result.

14. The method according to claim 13, wherein fusing, based on the voice recognition result of the audio corresponding to the image sequence, the lipreading result of the at least one image subsequence to obtain the fusion recognition result comprises:

sorting probabilities, that each image subsequence of the at least one image subsequence is classified as each of multiple predetermined characters corresponding to the specified content, to obtain a feature vector corresponding to the each image subsequence; and

concatenating feature vectors of the at least one image subsequence based on the voice recognition result of the audio corresponding to the image sequence to obtain a concatenating result, wherein the fusion recognition result comprises the concatenating result.

15. The method according to any one of claims 12 to 14, wherein determining whether the fusion recognition result matches the voice recognition result of the audio corresponding to the image sequence comprises:

obtaining a matching probability between the lipreading result and the voice recognition result based on the fusion recognition result and the voice recognition result by using a second neural network model; and

determining whether the lipreading result matches the voice recognition result based on the matching probability between the lipreading result and the voice recognition result.

16. The method according to any one of claims 12 to 15, further comprising:

performing voice recognition processing on the audio corresponding to the image sequence to obtain the voice recognition result; and

determining whether the voice recognition result is consistent with the specified content;

wherein determining the anti-spoofing detection result based on the matching result between the fusion recognition result and the voice recognition result of the audio comprises:

determining, in response to the voice recognition result of the audio corresponding to the image sequence being consistent with the specified content and the lipreading result of the image sequence matching the voice recognition result of the audio, that the anti-spoofing detection result is that the anti-spoofing detection is passed.

17. The method according to any one of claims 1 to 16, wherein the lipreading result of the image subsequence comprises: probabilities that the image subsequence is classified as each of multiple predetermined characters corresponding to the specified content.

18. An apparatus for anti-spoofing detection, comprising:

a first obtaining module, configured to obtain at least one image subsequence from an image sequence, wherein the image sequence is acquired by an image acquisition apparatus after a user is prompted to read a specified content, and the image subsequence comprises at least one image in the image sequence;

a lipreading module, configured to perform lipreading on the at least one image subsequence to obtain a lipreading result of the at least one image subsequence; and

a first determination module, configured to determine an anti-spoofing detection result based on the lipreading result of the at least one image subsequence.

19. An electronic device, comprising:

a memory, configured to store computer programs; and

a processor, configured to execute the compute programs stored in the memory, wherein the computer programs, when being executed, implement the method according to any one of claims 1 to 17.

20. A computer-readable storage medium having stored thereon computer programs that, when being executed by a processor, cause the processor to implement the method according to any one of claims 1 to 17.