

CLAIMS

1. An identity authentication method, wherein the method comprises:
 - obtaining an audio and video stream, wherein the audio and video stream is generated by a target object to be authenticated (501);
 - determining whether lip reading and voice in the audio and video stream are consistent (502);
 - in response to determining that the lip reading and the voice are consistent, using voice content obtained by performing voice recognition (503) on an audio stream in the audio and video stream as an object identifier of the target object (504);
 - determining whether pre-stored object registration information comprises the object identifier;
 - in response to determining that the pre-stored object registration information comprises the object identifier, obtaining a model physiological feature corresponding to the object identifier from the pre-stored object registration information;
 - performing physiological recognition on the audio and video stream to obtain a physiological feature of the target object (505, 506);
 - comparing the physiological feature of the target object with the model physiological feature to obtain a comparison result (507, 508);
 - determining whether the comparison result satisfies an authentication condition; and
 - in response to determining that the comparison result satisfies an authentication condition, determining that the target object has been authenticated.
2. The method according to claim 1, wherein determining whether lip reading and voice in the audio and video stream are consistent, using voice content obtained by performing voice recognition on an audio stream in the audio and video stream as an object identifier of the target object, and obtaining a model physiological feature corresponding to the object identifier from object registration information are executed in parallel.
3. The method according to claim 1 or claim 2, wherein the physiological feature comprises a voiceprint feature and a facial feature, and the model physiological feature comprises a model facial feature and a model voiceprint feature.

4. The method according to claim 3, wherein performing physiological recognition on the audio and video stream to obtain a physiological feature of the target object comprises:

performing voiceprint recognition on the audio and video stream to obtain a voiceprint feature of the target object; and

performing facial recognition on the audio and video stream to obtain a facial feature of the target object.

5. The method according to claim 4, wherein comparing the physiological feature of the target object with the model physiological feature to obtain a comparison result, and if the comparison result satisfies an authentication condition, determining that the target object has been authenticated, comprises:

comparing the voiceprint feature of the target object with the model voiceprint feature to obtain a voiceprint comparison score;

comparing the facial feature of the target object with the model facial feature to obtain a facial comparison score;

determining whether the voiceprint comparison score and the facial comparison score satisfy the authentication condition; and

in response to determining that the voiceprint comparison score and the facial comparison score satisfy the authentication condition, determining that the target object has been authenticated.

6. The method according to any one of claims 3 to 5, wherein it is determined that the target object has been authenticated if the voiceprint comparison score and the facial comparison score satisfy at least one of the following:

the voiceprint comparison score is greater than a voiceprint score threshold, and the facial comparison score is greater than a facial score threshold; or

a product of the voiceprint comparison score and the facial comparison score is greater than a corresponding product threshold; or

a weighted sum of the voiceprint comparison score and the facial comparison score is greater than a corresponding weighted threshold.

7. The method according to claim 1 or claim 2, wherein determining whether lip reading and voice in the audio and video stream are consistent, and if the lip reading and the voice are consistent, using voice content obtained by performing voice recognition on an audio stream in the audio and video stream as an object identifier of the target object comprises:

identifying a voice syllable and a corresponding time point in the audio stream in the audio and video stream;

identifying a lip reading syllable and a corresponding time point in a video stream in the audio and video stream; and

determining that the lip reading and the voice are consistent if both the voice syllable and the lip reading syllable are consistent with the corresponding time point.

8. The method according to claim 1 or claim 2, wherein before obtaining a collected audio and video stream, the method further comprises:

obtaining an audio and video stream to be registered of the target object;

in response to determining that the lip reading and the voice in the audio and video stream to be registered are consistent, using voice content obtained by performing voice recognition on an audio stream in the audio and video stream as the object identifier of the target object;

performing physiological recognition on the audio and video stream to be registered to obtain the model physiological feature of the audio and video stream to be registered; and

storing the object identifier of the target object and the corresponding model physiological feature in the object registration information.

9. The method according to claim 1 or claim 2, wherein the target object to be authenticated comprises a user of a user device, the audio stream comprises a reading voice of the user, the video stream comprises a reading video image of the user, and the collected audio and video stream comprises an audio and video stream generated when the user registers with an application.

10. An identity authentication apparatus, wherein the apparatus comprises multiple modules configured to perform operations according to the method of any one of claims 1 to 9.