CLAIMS

1. A video cutting method, comprising:

extracting to-be-identified video data from video stream data, and extracting image data and audio data from the to-be-identified video data;

inputting the image data into a preset marking behavior recognition model to obtain a marking behavior recognition result, and inputting the audio data into a preset marking speech recognition model to obtain a marking speech recognition result;

obtaining a marking recognition result based on the marking behavior recognition result, the marking speech recognition result, and a preset marking triggering rule;

adding a cutting point identifier to the to-be-identified video data when a type of the marking recognition result is a marking operation; and

performing cutting processing on the video stream data based on the cutting point identifier to obtain video segment data;

extracting audio segment data from the video segment data;

querying a preset speech recognition model;

inputting the audio segment data into the speech recognition model to obtain transcription data of the video segment data; and

determining a business type corresponding to the video segment data based on the transcription data, and storing the video segment data in a storage location corresponding to the business type.

2. The method according to claim 1, wherein the extracting to-be-identified video data from video stream data comprises:

obtaining the video stream data;

determining a video stream identification length; and

extracting the to-be-identified video data from the video stream data based on the video stream identification length.

3. The method according to claim 1, wherein the inputting the image data into a preset marking behavior recognition model to obtain a marking behavior recognition result, and

inputting the audio data into a preset marking speech recognition model to obtain a marking speech recognition result comprises:

determining identification information of a business person corresponding to the to-be-identified video data;

querying a preset marking behavior recognition model and a preset marking speech recognition model separately corresponding to the identification information;

extracting image feature data from the image data, and extracting audio feature data from the audio data; and

inputting the image feature data into the marking behavior recognition model to obtain a marking behavior recognition result, and inputting the audio feature data into the marking speech recognition model to obtain a marking speech recognition result.

4. The method according to claim 3, wherein before the querying a preset marking behavior recognition model and a preset marking speech recognition model separately corresponding to the identification information, the method further comprises:

obtaining historical behavior image data and historical marking speech data from a business system;

classifying the historical behavior image data and the historical marking speech data by business person to obtain historical behavior image data corresponding to each business person and historical marking speech data corresponding to each business person;

training the historical behavior image data corresponding to each business person to obtain the marking behavior recognition model; and

training the historical marking speech data corresponding to each business person to obtain the marking speech recognition model.

5. The method according to claim 1, wherein the obtaining a marking recognition result based on the marking behavior recognition result, the marking speech recognition result, and a preset marking triggering rule comprises:

querying a preset marking triggering rule, wherein the marking triggering rule comprises a behavior triggering rule and a speech triggering rule;

comparing the marking behavior recognition result with the behavior triggering rule to obtain a behavior triggering result;

comparing the marking speech recognition result with the speech triggering rule to obtain a speech triggering result; and

obtaining the marking recognition result based on the behavior triggering result and the speech triggering result.

6. The method according to claim 5, wherein the obtaining the marking recognition result based on the behavior triggering result and the speech triggering result comprises:

performing an OR operation on the behavior triggering result and the speech triggering result to obtain the marking recognition result.

7. The method according to claim 1, wherein the adding a cutting point identifier to the to-be-identified video data when a type of the marking recognition result is a marking operation comprises:

determining the type of the marking recognition result;

querying a preset label adding rule when the type of the marking recognition result is a marking operation; and

determining a key frame from the to-be-identified video data based on the label adding rule, and adding a cutting label to the key frame, wherein the cutting point identifier comprises the cutting label.

8. The method according to any one of claims 1 to 7, further comprising:

when receiving a marking-based cutting instruction, determining a cutting moment value of the marking-based cutting instruction;

determining a to-be-cut video frame corresponding to the cutting moment value from the to-be-identified video data;

adding a cutting point identifier to the to-be-cut video frame; and

returning to the step of performing cutting processing on the video stream data based on the cutting point identifier to obtain video segment data.

9. The method according to claim 1, wherein the determining a business type corresponding to the video segment data based on the transcription data, and storing the video segment data in a storage location corresponding to the business type comprises:

extracting a business keyword from the transcription data;

determining the business type corresponding to the video segment data based on the business keyword;

querying a preset storage location corresponding to the business type; and

storing the video segment data in the storage location.

10. A video cutting apparatus, comprising:

an identified-data extraction module, configured to extract to-be-identified video data from video stream data, and extract image data and audio data from the to-be-identified video data;

a marking recognition processing module, configured to input the image data into a preset marking behavior recognition model to obtain a marking behavior recognition result, and input the audio data into a preset marking speech recognition model to obtain a marking speech recognition result;

a marking result obtaining module, configured to obtain a marking recognition result based on the marking behavior recognition result, the marking speech recognition result, and a preset marking triggering rule;

a cutting identifier adding module, configured to add a cutting point identifier to the to-be-identified video data when a type of the marking recognition result is a marking operation; and

a video cutting module, configured to perform cutting processing on the video stream data based on the cutting point identifier to obtain video segment data;

wherein the apparatus further is configured to extract audio segment data from the video segment data, query a preset speech recognition model, input the audio segment data into the speech recognition model to obtain transcription data of the video segment data, determine a business type corresponding to the video segment data based on the transcription data, and store the video segment data in a storage location corresponding to the business type.

11. The apparatus according to claim 10, wherein the identified-data extraction module comprises:

a video stream obtaining unit, configured to obtain the video stream data;

an identification length determining unit, configured to determine a video stream identification length; and

an identified-data extraction unit, configured to extract the to-be-identified video data from the video stream data based on the video stream identification length.

12. The apparatus according to claim 10, wherein the marking recognition processing module comprises:

an identification determining unit, configured to determine identification information of a business person corresponding to the to-be-identified video data;

a recognition model query unit, configured to query a preset marking behavior recognition model and a preset marking speech recognition model separately corresponding to the identification information;

a feature data extraction unit, configured to extract image feature data from the image data, and

extract audio feature data from the audio data; and

a marking recognition unit, configured to input the image feature data into the marking behavior recognition model to obtain a marking behavior recognition result, and input the audio feature data into the marking speech recognition model to obtain a marking speech recognition result.

13. The apparatus according to claim 12, wherein the apparatus further comprises:

a historical data obtaining module, configured to obtain historical behavior image data and historical marking speech data from a business system;

a historical data classification module, configured to classify the historical behavior image data and the historical marking speech data by business person to obtain historical behavior image data corresponding to each business person and historical marking speech data corresponding to each business person;

a behavior recognition model training module, configured to train the historical behavior image data corresponding to each business person to obtain the marking behavior recognition model; and

a speech recognition model training module, configured to train the historical marking speech data corresponding to each business person to obtain the marking speech recognition model.

14. A computer device, comprising a memory and one or more processors, wherein the memory stores a computer-readable instruction, and when the computer-readable instruction is executed by the one or more processors, the one or more processors are enabled to perform the following steps:

extracting to-be-identified video data from video stream data, and extracting image data and audio data from the to-be-identified video data;

inputting the image data into a preset marking behavior recognition model to obtain a marking behavior recognition result, and inputting the audio data into a preset marking speech recognition model to obtain a marking speech recognition result;

obtaining a marking recognition result based on the marking behavior recognition result, the marking speech recognition result, and a preset marking triggering rule;

adding a cutting point identifier to the to-be-identified video data when a type of the marking recognition result is a marking operation; and

performing cutting processing on the video stream data based on the cutting point identifier to obtain video segment data;

extracting audio segment data from the video segment data;

querying a preset speech recognition model;

inputting the audio segment data into the speech recognition model to obtain transcription data of the video segment data; and

determining a business type corresponding to the video segment data based on the transcription data, and storing the video segment data in a storage location corresponding to the business type.

15. The computer device according to claim 14, wherein when executing the computer-readable instruction, the processor further performs the following steps:

obtaining the video stream data;

determining a video stream identification length; and

extracting the to-be-identified video data from the video stream data based on the video stream identification length.

16. The computer device according to claim 14, wherein when executing the computer-readable instruction, the processor further performs the following steps:

determining identification information of a business person corresponding to the to-be-identified video data;

querying a preset marking behavior recognition model and a preset marking speech recognition model separately corresponding to the identification information;

extracting image feature data from the image data, and extracting audio feature data from the audio data; and

inputting the image feature data into the marking behavior recognition model to obtain a marking behavior recognition result, and inputting the audio feature data into the marking speech recognition model to obtain a marking speech recognition result.

17. One or more non-volatile computer-readable storage media that store a computer-readable instruction, wherein when the computer-readable instruction is executed by one or more processors, the one or more processors are enabled to perform the following steps:

extracting to-be-identified video data from video stream data, and extracting image data and audio data from the to-be-identified video data;

inputting the image data into a preset marking behavior recognition model to obtain a marking behavior recognition result, and inputting the audio data into a preset marking speech recognition model to obtain a marking speech recognition result;

obtaining a marking recognition result based on the marking behavior recognition result, the marking speech recognition result, and a preset marking triggering rule;

adding a cutting point identifier to the to-be-identified video data when a type of the marking recognition result is a marking operation; and

performing cutting processing on the video stream data based on the cutting point identifier to obtain video segment data;

extracting audio segment data from the video segment data;

querying a preset speech recognition model;

inputting the audio segment data into the speech recognition model to obtain transcription data of the video segment data; and

determining a business type corresponding to the video segment data based on the transcription data, and storing the video segment data in a storage location corresponding to the business type.

18. The storage medium according to claim 17, wherein when executing the computer-readable instruction, the processor further performs the following steps:

obtaining the video stream data;

determining a video stream identification length; and

extracting the to-be-identified video data from the video stream data based on the video stream identification length.

19. The storage medium according to claim 17, wherein when executing the computer-readable instruction, the processor further performs the following steps:

determining identification information of a business person corresponding to the to-be-identified video data;

querying a preset marking behavior recognition model and a preset marking speech recognition model separately corresponding to the identification information;

extracting image feature data from the image data, and extracting audio feature data from the audio data; and

inputting the image feature data into the marking behavior recognition model to obtain a marking behavior recognition result, and inputting the audio feature data into the marking speech recognition model to obtain a marking speech recognition result.