

VIDEO STREAM SENDING METHOD, SYSTEM, ELECTRONIC DEVICE, AND STORAGE MEDIUM

The present application claims the priority to a Chinese patent application No. 202311634010.6 filed with the China National Intellectual Property Administration on December 01, 2023 and entitled “Video Stream Sending Method, System, Electronic Device, and Storage Medium”, which is incorporated herein by reference in its entirety.

Technical Field

The present disclosure relates to the technical field of video sending, and in particular, to a video stream sending method, system, electronic device, and storage medium.

Background

At present, when relaying a video, in order to meet the needs of users of different languages, speech in the video is often translated into subtitles in multiple languages, and different video streams are created for different languages. When playing on the user end, the user may select one of the video streams received in multiple different languages and play it.

However, even though the above method can meet the needs of users of different languages, the need to create different video streams for different languages results in sending multiple video streams to the user end, which occupies a large amount of broadband resources.

Summary of the Invention

An objective of embodiments of the present application is to provide a video stream sending method, system, electronic device, and storage medium so as to reduce the amount of resources used when sending a video stream. Specific technical solutions are as follows:

In a first aspect of the present application, first, a video stream sending method applied to a relay device is provided. The method comprises:

obtaining a first video stream, wherein, the first video stream comprises image data and audio data;

recognizing a speech in the audio data to obtain speech text; translating the speech text into subtitle text in at least one preset language;

creating a second video stream based on the first video stream;

adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text;

sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

In a possible implementation, adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text, comprises:

adding the subtitle text in at least one preset language to the second video stream, by means of Supplemental Enhancement Information SEI, to obtain the video stream containing the subtitle text.

In a possible implementation, recognizing a speech in the audio data to obtain speech text, comprises:

recognizing the speech in the audio data to obtain at least one line and a time stamp corresponding to each line;

wherein translating the speech text into subtitle text in at least one preset language comprises:

translating the at least one line to obtain subtitle text in at least one preset language corresponding to each line and a corresponding time stamp.

In a possible implementation, sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display, comprises:

sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render the subtitle text in the target preset language for display based on a time stamp corresponding to the subtitle text in the target preset language.

In a possible implementation, obtaining the first video stream comprises:

obtaining a live video stream, wherein, the live video stream is in a Real-Time Messaging Protocol RTMP format or a web video HLV format.

In a possible implementation, recognizing a speech in the audio data to obtain speech text, comprises:

recognizing the speech in the audio data, by means of Automatic Speech Recognition ASR, to obtain the speech text;

wherein translating the speech text into subtitle text in at least one preset language comprises:

translating the speech text into the subtitle text in at least one preset language by means of a preset machine translation tool.

In a possible implementation, creating the second video stream based on the first video stream, comprises:

receiving a subtitle switching request sent by the client;

in response to the subtitle switching request, creating the second video stream based on the first video stream.

In a possible implementation, creating the second video stream based on the first video stream comprises:

receiving a subtitle switching request sent by the client;

in response to the subtitle switching request, sending the video stream containing the subtitle text to the client to enable the client to obtain subtitle text in a target preset language from the subtitle text in the at least one preset language and render it for display.

In a second aspect of the present disclosure, a video stream sending system is provided. The system comprises a relay device and a client;

the relay device being configured to obtain a first video stream, wherein, the first video stream comprises image data and audio data; recognize a speech in the audio data to obtain speech text; translate the speech text into subtitle text in at least one preset language; create a second video stream based on the first video stream; add the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text; send the video stream containing the subtitle text to the client;

the client being configured to receive the video stream containing the subtitle text; parse the video stream containing the subtitle text to obtain the subtitle text in at least one preset language; identify subtitle text in a target preset language from the subtitle text in at least one preset language based on subtitle selection information; and render the subtitle text in the target preset language for display.

In a possible implementation, the relay device is specifically configured to add the subtitle text in at least one preset language to the second video stream, by means of Supplemental Enhancement Information SEI, to obtain the video stream containing the subtitle text.

In a possible implementation, the relay device is specifically configured to recognize a speech in the audio data to obtaining at least one line and a time stamp corresponding to each line; translate the at least one line to obtain subtitle text in at least one preset language corresponding to each line and a corresponding time stamp.

In a possible implementation, the relay device is specifically configured to obtain a live video stream, wherein, the live video stream is in a Real-Time Messaging Protocol RTMP format or a web video HLV format.

In a possible implementation, the relay device is specifically configured to recognize a speech in the audio data, by means of Automatic Speech Recognition ASR, to obtain the speech text; translate the speech text into subtitle text in at least one preset language by means of a preset machine translation tool.

In a possible implementation, the client is specifically configured to receive the subtitle selection information sent by the client; or, obtain geographical location information of the client; determine the subtitle selection information according to the obtained geographical location information.

In a possible implementation, the subtitle text in at least one preset language contains a time stamp corresponding to the subtitle text;

the client is specifically configured to render the subtitle text in the target preset language for display based on the time stamp corresponding to the subtitle text.

In a possible implementation, the relay device is specifically configured to receive a subtitle switching request sent by the client; and in response to the subtitle switching

request, create the second video stream based on the first video stream.

In a possible implementation, the relay device is specifically configured to receive a subtitle switching request sent by the client; and in response to the subtitle switching request, send the video stream containing the subtitle text to the client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

In yet another aspect of the present application, an electronic device is provided, the electronic device comprises a processor, a communication interface, a memory and a communication bus , wherein the processor, the communication interface and the memory communicate with each other through the communication bus;

the memory is configured to store a computer program;

the processor is configured to perform any of the video stream sending methods.

In yet another aspect of the present application, it provides a computer readable storage medium having a computer program stored therein, wherein the computer program, when executed by a processor, causes the processor to perform any of the video stream sending methods.

In yet another aspect of the present application, it provides a computer program product containing instructions, when running on a computer, causes the computer to perform any of the video stream sending methods.

Embodiments of the present application provide a video stream sending method, system, electronic device, and storage medium, wherein, by obtaining a first video stream, wherein, the first video stream comprises image data and audio data; recognizing a speech in the audio data to obtain speech text; translating the speech text into subtitle text in at least one preset language; creating a second video stream based on the first video stream; adding the subtitle text in at least one preset language to the second video stream to obtaining a video stream containing the subtitle text; sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display; after obtaining the first video stream, speech recognition and translation can be performed on the first video stream to obtain a subtitle in at least one preset language, and then a video

stream containing the subtitle text in at least one preset language is created and sent to a client, without the need to create a video stream for each subtitle, thus reducing resources used.

Brief Descriptions of the Drawings

In order to illustrate embodiments of the present disclosure or the technical solutions in the prior art more clearly, the drawings needed to describe the embodiments or the prior art will be briefly introduced below.

Fig. 1 is a schematic flow chart of a video stream sending method provided by an embodiment of the present application;

Fig. 2 is a schematic flow chart of obtaining subtitle text provided by an embodiment of the present application;

Fig. 3 is a schematic view of Supplemental Enhancement Information provided by an embodiment of the present application;

Fig. 4 is a schematic view of a structure of a video stream sending system provided by an embodiment of the present application;

Fig. 5 is a schematic view of a structure of an electronic device provided by an embodiment of the present application.

Detailed Description of the Embodiments

The technical solutions in embodiments of the present application will be described below with reference to the accompanying drawings in the embodiments of the present disclosure.

In order to solve the problem in the prior art that different video streams need to be created for different languages during the video sending process, resulting in multiple video streams being sent to the user end and occupying a large amount of broadband resources, the present application provides a video stream sending method, system, electronic device, and storage medium.

In a first aspect of the implementation of the present application, first, a video stream sending method applied to a relay device is provided. In reference to Fig. 1, which is a schematic flow chart of a video stream sending method provided by an embodiment of the

present application, the method comprises:

Step S11, obtaining a first video stream, wherein, the first video stream comprises image data and audio data;

Step S12, recognizing a speech in the audio data to obtain speech text; translating the speech text into subtitle text in at least one preset language;

Step S13, creating a second video stream based on the first video stream;

Step S14, adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text;

Step S15, sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

It can be seen that by means of the method of the embodiment of the present application, after obtaining a first video stream, speech recognition and translation can be performed on the first video stream to obtain a subtitle in at least one preset language. Then, a video stream containing the subtitle text in at least one preset language is created and sent to a client, it does not need to create a video stream for each subtitle, thus saving resources.

Concerning the above-described Step S11, wherein, the first video stream comprises image data and speech data, during actual use, the speech data can include speech when talking, such as a person's dialogue, commentary during a match, etc. The image data can be data of various kinds, for example, images of a landscape, images of a match, etc. They are not limited by the present application. In an example, the above-described first video stream can be a live video stream, the image data can be a picture of live streaming, and the speech data can be the dubbing or commentary during the live streaming process.

The method of the embodiment of the present application is applied to a smart terminal by which the above-described video stream sending method can be implemented. Specifically, the smart terminal can be a relay device in the course of sending a video. For example, a front-end device captures a video and sends it to the relay device. Then, the relay device forwards the video to a client. For example, during live streaming of a sports match, a front-end camera, after capturing images of the match and commentary of a host, sends the captured video and commentary to the relay device. Then, the relay device sends

the video and commentary to the client. A user can view the video on the client.

Concerning the above-described Step S12, in the embodiment of the present application, a speech recognition tool can be preset. Recognition of speech data in the first video stream is performed by means of the speech recognition tool to obtain a corresponding speech text. Then, the speech text is translated to obtain text in at least one preset language. During actual use, recognition of speech data in the first video stream can be directly performed by means of a network model obtained in advance through training to obtain subtitle text in at least one preset language. For example, sample speech data is inputted into a network model to be trained and processed by multiple network layers of the network model to obtain output subtitle text in at least one preset language. Then, the output subtitle text is matched with pre-created standard subtitle text to obtain a current loss of the network model to be trained. Then, structural parameters of the network model to be trained are adjusted based on the current loss. Then, sample speech data continue to be inputted into the adjusted network model, until the model's loss is smaller than a preset threshold or a preset number of iterations are reached, thus a trained network model is obtained. Speech data processing can be performed by means of the trained network model to obtain subtitle text in at least one preset language. Reference can be made to relevant prior art regarding model types and structures, which are not limited here by the present application. During actual use, the above-described at least one preset language can include the language corresponding to the speech data itself. For example, when the speech data comprises speech corresponding to “你好” in Chinese, it can be translated to obtain a subtitle, for example, the corresponding “Hello” in English or the corresponding “Bonjour” in French, or the corresponding “你好” in Chinese. During actual use, recognition of speech in the audio data can be performed manually by a human. In the case of manual recognition, the speech text can be manual stenography text.

Concerning the above-described Step S13, when creating a second video stream based on the first video stream, the second video stream can contain image data corresponding to the first video stream. During actual use, the second video stream and the first video stream can have the same video data. Because speech recognition and translation need to be performed based on the first video stream in the embodiment of the present application, after receiving the first video stream, sending can be done only after completing recognition

and translation. But recognition and translation take a certain amount of time, generally in the order of milliseconds. Therefore, after receiving the first video stream, it will be cached and wait for speech recognition and translation to be completed before sending the cached data as the second video stream. The specific cache duration may be set based on the time it takes for speech recognition and translation.

Concerning the above-described Step S14, in order to solve the problem in the prior art that a video stream needs to be created for each language and sent simultaneously to a user, resulting in the use of a large amount of resources, in the present application, subtitle text in at least one preset language and the second video stream are combined into one video stream, i.e., a video stream containing the subtitle text. Just send the video stream containing the subtitle text to the user to obtain a subtitle in at least one preset language, reducing resource usage. When adding the subtitle text in at least one preset language to the second video stream, various preset methods can be used to combine the video stream and the subtitle text. For example, the subtitle text is added to the second video stream by means of Supplemental Enhancement Information to obtain a corresponding video stream containing the subtitle text. Because when sending a video stream containing subtitle text to a user, the subtitle text therein will be rendered to the screen by the corresponding client for display, the video stream containing subtitle text can contain image data. During actual use, the video stream containing subtitle text can contain image data and speech data. For example, by caching the first video stream to create a second video stream, and adding subtitle text to the second video stream, a video stream containing the subtitle text is obtained.

Concerning the above-described Step S15, in the embodiment of the present application, the client can be a device where the user plays videos, such as a computer, a cell phone, etc. After the client receives a video stream containing subtitle text, the added subtitle text in at least one preset language can be recognized. Then, one of them can be selected as subtitle text in a target language, and when playing back image data in the video stream containing subtitle text, the subtitle text in the target language is rendered in the image data. For example, when the client receives a video stream containing subtitle text, it obtains subtitle texts corresponding to Chinese, English, and French. After the user selects Chinese, the subtitle text corresponding to Chinese can be rendered to the screen played by the client, thereby achieving the display of subtitle.

A client can determine a target preset language by means of various methods, including: receiving target preset language selection information sent by a user; or, obtaining geographic location information of the client; and determining the target preset language based on obtained geographic location information. Receiving target preset language selection information sent by a user can be selected through the display or setting interface of the client, and a user can select a target language by means of the options. For example, after recognizing and obtaining subtitle texts in multiple preset languages, selectable languages are displayed to a user for easy selection. When determining a target preset language based on obtained geographic location information, a language of the region can be determined as a target language based on obtained geographic location information, thus facilitating selection and improving user experience. For example, when the geographic location of a client in a certain region of China is obtained, as Chinese is mainly spoken in China, Chinese can be determined as a target language.

In a possible implementation, adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text, comprises: adding the subtitle text in at least one preset language to the second video stream, by means of Supplemental Enhancement Information (SEI), to obtain the video stream containing the subtitle text. In this embodiment, by means of Supplemental Enhancement Information (SEI), text information and audio and video information can be packaged together by a stream media channel and then pushed to a client, so as to achieve the synchronized sending of text data and audio and video data. In the embodiment of the present application, subtitle text in at least one preset language can be added to the second video stream to obtain the video stream containing subtitle text, by means of Supplemental Enhancement Information (SEI). The video stream containing the subtitle text is sent to the client. Then, the client performs subtitle text extraction, thus achieving the sending of at least one subtitle text by means of a single video stream.

In a possible implementation, in reference to Fig. 2, which is a schematic flow chart of obtaining subtitle text provided by an embodiment of the present application, wherein recognizing a speech in the audio data to obtain speech text, comprises:

Step S21, recognizing the speech in the audio data to obtain at least one line and a time stamp corresponding to each line;

wherein translating the speech text into subtitle text in at least one preset language comprises:

Step S22, translating the at least one line to obtain subtitle text in at least one preset language corresponding to each line and a corresponding time stamp.

In a possible implementation, sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display, comprises: sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render the subtitle text in the target preset language for display based on a time stamp corresponding to the subtitle text in the target preset language.

In the embodiment of the present application, when performing speech recognition on the speech data, a time stamp corresponding to each speech should be extracted as well. The time stamp can correspond to a certain time period in the first videos stream. When translating the speech text to obtain subtitle text, a corresponding time stamp can be determined for the subtitle text based on corresponding speech data. Specifically, speech recognition is performed on the speech data to obtain one or more lines and a time stamp corresponding to each line. Each line can correspond to a time stamp in the first video stream. Then, the one or more lines are translated to obtain a corresponding subtitle text. For each subtitle text, a time stamp for this subtitle text can be determined based on a time stamp corresponding to corresponding line. Thus, each line text can determine a display time period based on its own time stamp. Hence, when a client renders a subtitle, the rendering can be performed based on a time stamp corresponding to each subtitle text, preventing mismatch between a subtitle and an image or sound. For example, a sentence in the first video stream appears between 1 minute 10 seconds and 1 minute 12 seconds. Speech recognition is performed on this sentence to obtain a time stamp corresponding to the speech of 1 minute 10 seconds to 1 minute 12 seconds. Then, the time stamp of subtitle texts of multiple present languages obtained by translating the speech is in the range of 1 minute 10 seconds and 1 minute 12 seconds. This allows for rendering of the subtitle text on the client based on this timestamp, i.e. displaying the subtitle text in the screen from 1 minute 10 seconds to 1 minute 12 seconds. In the embodiment of the present application,

recognition and translation of the speech data can be performed by a preset translation tool or manually by a human. If manual translation is used, the cache time may be longer, possibly in seconds or minutes, when creating a second video stream based on the first video stream.

In a possible implementation, obtaining a first video stream comprises: obtaining a live video stream, wherein, the live video stream is in a Real-Time Messaging Protocol (RTMP) format or a web video (HLS) format. The method of the embodiment of the present application can be applied in the field of video live streaming. When obtaining a first video stream, the first video stream can be a live video stream, such as RTMP or HLS (web video format). During a live streaming, a relay device receives a video captured by a front-end device in real time and then sends it to a client. By means of the method of the embodiment of the present application, it is convenient for a user to observe the corresponding subtitle translation while watching a live streaming, and since only one video stream needs to be received, the occupied resources can be reduced. It can be understood that a relay device receives video segments one by one during a live streaming. By means of the method of the embodiment of the present application, each video segment can undergo speech recognition and translation and then be sent to be client.

In a possible implementation, recognizing a speech in the audio data to obtain speech text, comprises: recognizing the speech in the audio data, by means of Automatic Speech Recognition ASR, to obtain the speech text; wherein translating the speech text into subtitle text in at least one preset language comprises: translating the speech text into subtitle text in at least one preset language by means of a preset machine translation tool. By means of the method of the embodiment of the present application, a speech recognition and translation tool can be preset. Automatic recognition and translation can be performed by means of the speech recognition and translation tool, thus improving translation and recognition efficiency. In this embodiment, Automatic Speech Recognition (ASR) can convert word contents in a person's speech data to a computer-readable input, such as keys, binary coding, or character sequences, to obtain speech text. Then, subtitle texts of multiple preset languages are obtained by translating the speech text by means of a preset machine translation tool. By means of this method, unlike recognizing or locating a speaker, there is no need to recognize a speaker making a speech. Translation is directly performed by means of a machine-readable input, improving translation efficiency.

In a possible implementation, creating a second video stream based on the first video stream, comprises: receiving a subtitle switching request sent by the client; and in response to the subtitle switching request, creating the second video stream based on the first video stream.

In a possible implementation, creating the second video stream based on the first video stream comprises: receiving a subtitle switching request sent by the client; and in response to the subtitle switching request, sending the video stream containing the subtitle text to the client to enable the client to obtain subtitle text in a target preset language from the subtitle text in the at least one preset language and render it for display.

The main difference between the above-described two embodiments is that the timing of generating the video stream is different. In the former embodiment, after receiving a subtitle switching request sent by the client, in response to the subtitle switching request, a second video stream is created based on the first video stream. It can be understood that, only after the relay device receives a subtitle switching request sent by a client will it create the second video stream based on the switching request. It then creates a video stream containing subtitle text based on the second video stream and thus sends the created video stream containing the subtitle text to the client. The video stream containing subtitle text is parsed by the client to obtain subtitle text in at least one preset language. During actual use, speech recognition and translation can be performed only after receiving a switching request sent by a client. This solution is also within the scope of protection of the present application. In an example, the relay device in generally sends only image data and audio data captured by a front-end device to a client. Only after receiving a subtitle switching request sent by the client will it perform speech recognition and translation, create a video stream containing subtitle text based on a result of the recognition and translation, and send the created video stream containing subtitle text to a client. In the latter embodiment, the relay device receives a subtitle switching request sent by the client, and then in response to the subtitle switching request, sends the video stream containing subtitle text to the client to enable the client to obtain subtitle text in a target preset language from the subtitle text in the at least one preset language and render it for display. It can be understood that, before a relay device receives a subtitle switching request sent by a client, a video stream containing subtitle text has already been created. After a subtitle switching request sent by a client is received, the already created video stream containing subtitle text is sent to the client. In an

example, a relay device in generally sends only image data and audio data captured by a front-end device to a client, but also performs speech recognition and translation at the same time and creates a video stream containing subtitle text based on a result of the recognition and translation. Only after a subtitle switching request sent by a client is received, the already created video stream containing subtitle text will be sent to the client.

In order to describe the method of embodiments of the present application, description is made below in reference to specific embodiments. In reference to Fig. 3, which is a schematic view of Supplemental Enhancement Information provided by an embodiment of the present application, it comprises:

1. Providing two RTMP (or HLV) live streams, one being an original live stream and the other being a delayed stream, of which the time of delay can be very short (in the order of milliseconds), used for a subsequent subtitle carrier stream, wherein, during a live streaming, a relay device can be understood as receiving video segments one by one, and can perform speech recognition and translation for each video segment and then send it to a client. In reference to Fig. 3, for the original live stream, 4 video segments are contained. As to the delayed stream, it can entirely correspond to the original live stream and contain 4 video segments.

2. Parsing audio information in the original live stream, and translating text parsed from an audio by means of Automatic Speech Recognition (ASR) + machine translation technology, to obtain translated content, this content can be translated into different languages as needed.

3. Recording the translated content obtained in the previous step into Supplemental Enhancement Information (SEI) of the delayed stream and transmitting it to the client, wherein, by means of Supplemental Enhancement Information, text information and audio and video contents can be packaged together by a stream media channel and then pushed out from the relay device and received from a client, so as to achieve precisely synchronized sending of text data and audio and video contents. During actual use, both H.264 and H.265 video stream encoding formats support the use of SEI. In the case of both H.264 and H.265 encoding support the use of SEI, using H.264 as an example, H.264 source code stream is composed of NALU (NAL Units) one after another. The NALU is divided into an NAL header corresponding to video encoding and a Raw Byte

Sequence Payload (RBSP). NALU header is generally composed of 8 bits, the last 5 bits of which represent NAL Unit Type. Information such as Supplemental Enhancement Information can be added in those 5 bits. In reference to Fig. 3, in this figure, an original live stream is shown at the top, which can contain four video segments, and a delayed stream is shown at the bottom. A set of Supplemental Enhancement Information can be added in each video segment so as to achieve synchronized sending of subtitle text.

4. Playing back kernel in a client parses live stream data to obtain an audio and video stream and subtitle information, performing rendering based on needs, such as based on a subtitle language selected by a user, and drawing corresponding subtitle content. For example, in reference to Fig. 3, when a client receives a delayed stream, Supplemental Enhancement Information therein such as information added in the NALU Header of H.264 can be extracted. After the client has extracted subtitle text, the subtitle can be rendered by various ways, for example, hard subtitles, built-in subtitles, and external subtitles, etc. wherein, when rendering through hard subtitles, the subtitle image can be directly rendered on the video image during image compression on the client; when rendering through built-in subtitles, one or more subtitle streams can be encapsulated during encapsulation, and a subtitle can be rendered based on specific timeline and style information during client decoding; when rendering through external subtitles, subtitle text is treated as a separate subtitle file and not encapsulated in the video file. The client displays the subtitle file in layers.

It can be seen that through the method of the embodiment of the present application, after obtaining a first video stream, speech recognition and translation can be performed on the first video stream to obtain a subtitle in at least one preset language, and a video stream containing subtitle text in at least one preset language is created and sent to the client, without the need to create a video stream for each subtitle, thus reducing resources used.

In a second aspect of embodiments of the present disclosure, a video stream sending system is provided. In reference to Fig. 4, which is a schematic view of a structure of a video stream sending system provided by an embodiment of the present disclosure, the system comprises a relay device and a client;

the relay device 401 being configured to obtain a first video stream, wherein, the first video stream comprises image data and audio data; recognize a speech in the audio data to

obtain speech text; translate the speech text into subtitle text in at least one preset language; create a second video stream based on the first video stream; add the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text; send the video stream containing the subtitle text to the client;

the client 402 being configured to receive the video stream containing the subtitle text; parse the video stream containing the subtitle text to obtain the subtitle text in at least one preset language; identify subtitle text in a target preset language from the subtitle text in at least one preset language based on subtitle selection information; and render the subtitle text in the target preset language for display.

In a possible implementation, the relay device is specifically configured to add the subtitle text in at least one preset language to the second video stream, by means of Supplemental Enhancement Information SEI, to obtain the video stream containing the subtitle text.

In a possible implementation, the relay device is specifically configured to recognize a speech in the audio data to obtaining at least one line and a time stamp corresponding to each line; translate the at least one line to obtain subtitle text in at least one preset language corresponding to each line and a corresponding time stamp.

In a possible implementation, the relay device is specifically configured to obtain a live video stream, wherein, the live video stream is in a Real-Time Messaging Protocol (RTMP) format or a web video (HLS) format.

In a possible implementation, the relay device is specifically configured to recognize a speech in the audio data, by means of Automatic Speech Recognition ASR, to obtain the speech text; translate the speech text into subtitle text in at least one preset language by means of a preset machine translation tool.

In a possible implementation, the client is specifically configured to receive the subtitle selection information sent by the client; or, obtain geographical location information of the client; determine the subtitle selection information according to the obtained geographical location information.

In a possible implementation, the subtitle text in multiple preset languages contains a time stamp corresponding to the subtitle text;

the client is specifically configured to render the subtitle text in the target preset language for display based on the time stamp corresponding to the subtitle text.

In a possible implementation, the relay device is specifically configured to receive a subtitle switching request sent by the client; and in response to the subtitle switching request, create a second video stream based on the first video stream.

In a possible implementation, the relay device is specifically configured to receive a subtitle switching request sent by the client; and in response to the subtitle switching request, send the video stream containing the subtitle text to the client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

An embodiment of the present application further provides an electronic device, as shown in Fig. 5, comprising a processor 501, a communication interface 502, a memory 503 and a communication bus 504, wherein the processor 501, the communication interface 502 and the memory 503 communicate with each other through the communication bus 504;

the memory 503 is configured to store a computer program;

the processor 501 is configured to implement the following steps when executing the program stored in the memory 503:

obtaining a first video stream, wherein, the first video stream comprises image data and audio data;

recognizing a speech in the audio data to obtain speech text; translating the speech text into subtitle text in at least one preset language;

creating a second video stream based on the first video stream;

adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text;

sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

The communication bus mentioned in the above terminal may be a Peripheral

Component Interconnect (PCI) bus or an Extended Industry Standard Architecture (EISA) bus, or the like. The communication bus can be divided into address bus, data bus, control bus, or the like. For the convenience of presentation, only one thick line is used in the figures, but it does not mean that there is only one bus or one type of bus.

The communication interface is used for communication between the above terminal and other devices.

The memory may include a random access memory (RAM), and may also include a non-volatile memory (NVM), such as at least one disk memory. Optionally, the memory may also be at least one storage device located far away from the aforementioned processor.

The above-mentioned processor can be a general-purpose processor, including a central processing unit (CPU), a network processor (NP), or the like; it can also be a digital signal processor (DSP), application specific integrated circuit (Application Specific Integrated Circuit, ASIC), Field-Programmable Gate Array (FPGA) or other programmable logic devices, discrete gate or transistor logic devices, discrete hardware components.

In yet another embodiment provided by the present disclosure, a computer readable storage medium is further provided. A computer program is stored on the computer readable storage medium, and the computer program, when executed by a processor, performs any of the video stream sending methods in the above embodiments.

In yet another embodiment provided by the present disclosure, a computer program product containing instructions is further provided. The computer program product, when running on a computer, causes the computer to perform any of the video stream sending methods in the above embodiments.

In the above embodiments, all or part of them may be implemented by software, hardware, firmware or any combination thereof. When implemented using software, it may be implemented in whole or in part in the form of a computer program product. The computer program product includes one or more computer instructions. When the computer program instructions are loaded and executed on the computer, the processes or functions

according to the embodiments of the present application will be generated in whole or in part. The computer can be a general purpose computer, a special purpose computer, a computer network, or other programmable devices. The computer instructions may be stored in a computer-readable storage medium or transmitted from one computer-readable storage medium to another computer-readable storage medium, for example, the computer instructions may be transmitted from a website, computer, server or data center to another website site, computer, server, or data center by wired (e.g., coaxial cable, fiber optic, digital subscriber line (DSL)) or wireless (e.g., infrared, wireless, microwave, or the like) means. The computer-readable storage medium may be any available medium that can be accessed by a computer, or an integrated data storage device such as a server or a data center containing one or more available media storage. The available medium may be a magnetic medium (for example, a floppy disk, a hard disk, or a magnetic tape), an optical medium (for example, DVD), or a semiconductor medium (for example, a Solid State Disk (SSD)).

It should be noted that in the specification, relational terms such as first and second are only used to distinguish one entity or operation from another entity or operation, and do not necessarily require or imply that there is an actual relationship between these entities or operations. Furthermore, the term "comprises", "include" or any other variation thereof is intended to cover a non-exclusive inclusion such that a process, method, article, or apparatus comprising a set of elements includes not only those elements, but also include other elements not expressly listed or also include elements inherent in such a process, method, article, or device. Without further limitations, an element defined by the phrase "comprising a ..." does not exclude the presence of additional identical elements in the process, method, article or apparatus comprising said element.

Each embodiment in this specification is described in a related manner, the same and similar parts of each embodiment can be referred to each other, and each embodiment focuses on the differences from other embodiments. In particular, for the embodiment of the neural network model adjusting method applied to the first client, the embodiment of the neural network model adjusting method applied to the server, the embodiment of the

device, the embodiment of the computer-readable storage medium, and the embodiment of the computer program product, since they are basically similar to the embodiment of the neural network model adjusting system, the description is relatively simple, and for relevant parts, reference can be made to the part of the description of the system embodiment.

The above descriptions are only preferred embodiments of the present application, and are not intended to limit the protection scope of the present application. Any modifications, equivalent replacements, improvements, or the like made within the spirit and principles of the present application are included within the protection scope of this application.

CLAIMS

1. A video stream sending method, wherein, it is applied to a relay device and comprises:

obtaining a first video stream, wherein, the first video stream comprises image data and audio data;

recognizing a speech in the audio data to obtain speech text; translating the speech text into subtitle text in at least one preset language;

creating a second video stream based on the first video stream;

adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text;

sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

2. The method of claim 1, wherein, adding the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text, comprises:

adding the subtitle text in at least one preset language to the second video stream, by means of Supplemental Enhancement Information SEI, to obtain the video stream containing the subtitle text.

3. The method of claim 1, wherein, recognizing a speech in the audio data to obtain speech text, comprises:

recognizing the speech in the audio data to obtain at least one line and a time stamp corresponding to each line;

wherein translating the speech text into subtitle text in at least one preset language comprises:

translating the at least one line to obtain subtitle text in at least one preset language corresponding to each line and a corresponding time stamp.

4. The method of claim 3, wherein, sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display, comprises:

sending the video stream containing the subtitle text to a client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render the subtitle text in the target preset language for display based on a time stamp corresponding to the subtitle text in the target preset language.

5. The method of claim 1, wherein, obtaining the first video stream comprises:

obtaining a live video stream, wherein, the live video stream is in a Real-Time Messaging Protocol RTMP format or a web video HLV format.

6. The method of claim 1, wherein, recognizing a speech in the audio data to obtain speech text, comprises:

recognizing the speech in the audio data, by means of Automatic Speech Recognition ASR, to obtain the speech text;

wherein translating the speech text into subtitle text in at least one preset language comprises:

translating the speech text into the subtitle text in at least one preset language by means of a preset machine translation tool.

7. The method of claim 1, wherein, creating the second video stream based on the first video stream, comprises:

receiving a subtitle switching request sent by the client;

in response to the subtitle switching request, creating the second video stream based on the first video stream.

8. The method of claim 1, wherein, creating the second video stream based on the first video stream comprises:

receiving a subtitle switching request sent by the client;

in response to the subtitle switching request, sending the video stream containing the subtitle text to the client to enable the client to obtain subtitle text in a target preset

language from the subtitle text in the at least one preset language and render it for display.

9. A video stream sending system, wherein the system comprises a relay device and a client;

the relay device being configured to obtain a first video stream, wherein, the first video stream comprises image data and audio data; recognize a speech in the audio data to obtain speech text; translate the speech text into subtitle text in at least one preset language; create a second video stream based on the first video stream; add the subtitle text in at least one preset language to the second video stream to obtain a video stream containing the subtitle text; send the video stream containing the subtitle text to the client;

the client being configured to receive the video stream containing the subtitle text; parse the video stream containing the subtitle text to obtain the subtitle text in at least one preset language; identify subtitle text in a target preset language from the subtitle text in at least one preset language based on subtitle selection information; and render the subtitle text in the target preset language for display.

10. The system according to claim 9, wherein,

the relay device is specifically configured to add the subtitle text in at least one preset language to the second video stream, by means of Supplemental Enhancement Information SEI, to obtain the video stream containing the subtitle text.

11. The system according to claim 9, wherein,

the relay device is specifically configured to recognize a speech in the audio data to obtaining at least one line and a time stamp corresponding to each line; translate the at least one line to obtain subtitle text in at least one preset language corresponding to each line and a corresponding time stamp.

12. The system according to claim 9, wherein,

the relay device is specifically configured to obtain a live video stream, wherein, the live video stream is in a Real-Time Messaging Protocol RTMP format or a web video HLV format.

13. The system according to claim 9, wherein,

the relay device is specifically configured to recognize a speech in the audio data, by

means of Automatic Speech Recognition ASR, to obtain the speech text; translate the speech text into subtitle text in at least one preset language by means of a preset machine translation tool.

14. The system according to claim 9, wherein,

the client is specifically configured to receive the subtitle selection information sent by the client; or, obtain geographical location information of the client; determine the subtitle selection information according to the obtained geographical location information.

15. The system according to claim 9, wherein, the subtitle text in at least one preset language contains a time stamp corresponding to the subtitle text;

the client is specifically configured to render the subtitle text in the target preset language for display based on the time stamp corresponding to the subtitle text.

16. The system according to claim 9, wherein,

the relay device is specifically configured to receive a subtitle switching request sent by the client; and in response to the subtitle switching request, create the second video stream based on the first video stream.

17. The system according to claim 9, wherein,

the relay device is specifically configured to receive a subtitle switching request sent by the client; and in response to the subtitle switching request, send the video stream containing the subtitle text to the client to enable the client to obtain subtitle text in a target preset language from the subtitle text in at least one preset language and render it for display.

18. An electronic device comprising a processor, a communication interface, a memory and a communication bus, wherein the processor, the communication interface and the memory communicate with each other through the communication bus;

the memory is configured to store a computer program;

the processor is configured to perform the method according to any of claims 1-8.

19. A computer readable storage medium having a computer program stored therein, wherein the computer program, when executed by a processor, causes the processor to perform the method according to any of claims 1-8.