

MULTIMEDIA INTERACTION METHOD, DEVICE, AND EQUIPMENT, AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is based on, and claims benefit of priority to, Chinese Application No. 202010352029.1 filed on April 28, 2020. Disclosure of the Chinese Application is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

The disclosure relates to the field of AI, and more particularly, to a multimedia interaction method, device, and equipment, and a storage medium.

BACKGROUND

In recent years, with the popularization of Artificial Intelligence (AI), a network teaching platform is becoming increasingly popular. Network teaching generally needs interaction between a teacher and a student, or between student to improve a teaching effect and experience.

However, in order to implement interaction at a teaching platform, additional dedicated equipment is generally needed, which causes problems such as increased cost, inconvenient operation, etc.

SUMMARY

The present disclosure provides a multimedia interaction method, device, and equipment, and a storage medium.

A first aspect of embodiments of the present disclosure provides a multimedia interaction method, including: calling a multimedia interaction component of a teaching platform; acquiring a control instruction using the multimedia interaction component; extracting information based on the control instruction using the multimedia interaction

component; and displaying or playing the information through the multimedia interaction component, implementing interaction at the teaching platform without using external equipment, facilitating information acquisition, improving a network teaching effect.

The control instruction may include at least one of an image capture control instruction, a video recording control instruction, or an audio recording control instruction, implementing interaction at the teaching platform without using external equipment, allowing to acquire image, video or audio information conveniently and fast, improving a network teaching effect.

Extracting the information based on the control instruction using the multimedia interaction component may include at least one of: in response to the control instruction being the image capture control instruction, triggering capture of an image by a camera device according to the image capture control instruction; in response to the control instruction being the video recording control instruction, triggering capture of multiple frames of images by the camera device according to the video recording control instruction; or in response to the control instruction being the audio recording control instruction, triggering audio recording by a recording device according to the audio recording control instruction, implementing interaction at the teaching platform, allowing to perform video recording, image capture, and audio recording directly using the multimedia interaction component without using external equipment, facilitating information acquisition, improving a network teaching effect.

The multimedia interaction method may further include: after triggering capture of the multiple frames of images by the camera device according to the video recording control instruction in response to the control instruction being the video recording control instruction, acquiring a number of Frames Per Second (FPS) of the multiple frames of images; and selectively playing the multiple frames of images according to a correspondence between a frame number and time through the number of FPS, implementing interaction at the teaching platform, allowing playback according to a number of transmitted frames, rendering smoother video presentation, as well as a more flexible interaction effect.

Displaying or playing the information through the multimedia interaction component may include: in response to the control instruction being the video recording control instruction or the image capture control instruction, sequentially playing the multiple frames

of images according to timing of acquiring the multiple frames of images; or combining the multiple frames of images into a video file, and playing the video file, so that the multiple images are played smoothly while saving the time and process for saving the images into a video file. Alternatively, the multiple frames of images may be combined into a video file, and then the video file may be played, allowing the video to be played normally.

Extracting the information based on the control instruction using the multimedia interaction component may include: extracting the information from pre-stored preset information based on the control instruction using the multimedia interaction component. The multimedia interaction component may call a local preset voice to imitate a recording or call a preset image to imitate an image, saving the process of uploading a multimedia file such as a video and a photo of a student to the server for processing. Therefore, the teaching interaction does not rely on a large bandwidth, and has good real-time performance.

Extracting the information based on the control instruction using the multimedia interaction component may further include: setting a timing period of a timer according to the control instruction; and in response to expiration of the timing period of the timer, controlling extraction of the information by the multimedia interaction component, implementing more flexible interaction at the teaching platform.

The multimedia interaction method may further include: after displaying or playing the information through the multimedia interaction component, controlling a window for displaying or playing the information to perform any one of displacement, window scaling, or window hiding, implementing more flexible interaction at the teaching platform.

Displaying or playing the information through the multimedia interaction component may include: acquiring a preset code through the multimedia interaction component, and preprocessing the information according to the preset code; and displaying or playing the preprocessed information through the multimedia interaction component. In this way, the user may independently acquire the desired information through the multimedia interaction component, preprocess the information according to the preset code to acquire a processing result, better matching a teaching requirement, improving efficiency in programming learning. In addition, information processing may be done through local equipment, without having to transmit the information to a server for processing, reducing dependence on a bandwidth,

reducing network interaction, improving an information processing rate.

The preprocessing may include at least one of: performing an image processing operation on one image acquired or multiple frames of images acquired; or performing at least one of voice noise reduction, voice-to-text conversion, or voice synthesis on acquired audio.

The multimedia interaction method may further include: in response to the control instruction being the video recording control instruction or the image capture control instruction, after displaying or playing the information through the multimedia interaction component, acquiring an image from the multiple frames of images being played or the image being played, allowing to clip an image from the image or the video being played for subsequent processing, further facilitating information acquisition, improving flexibility in interaction as well as a network teaching effect.

A second aspect of embodiments of the present disclosure provides a multimedia interaction device, including a calling module, an inputting module, an information extracting module, and an outputting module. The calling module is configured to call a multimedia interaction component of a teaching platform. The inputting module is configured to acquire a control instruction using the multimedia interaction component. The information extracting module is configured to extract information based on the control instruction using the multimedia interaction component. The outputting module is configured to display or play the information through the multimedia interaction component.

A third aspect of embodiments of the present disclosure provides multimedia interaction equipment, including a memory and a processor. The memory stores program instructions which, when called by the processor, implement any multimedia interaction method herein.

A fourth aspect of embodiments of the present disclosure provides a computer-readable storage medium, having stored thereon a program file executable to implement any multimedia interaction method herein.

A fifth aspect of embodiments of the present disclosure provides a computer program product, including a computer-readable code which, when executed in electronic equipment, allows a processor in the electronic equipment to implement the multimedia interaction

method of the first aspect.

With the present disclosure, a multimedia interaction component of a teaching platform is called. A control instruction is acquired through the multimedia interaction component. Information is extracted based on the control instruction. The information is displayed or played through the multimedia interaction component, implementing multimedia interaction based on the multimedia interaction component, rendering multimedia interaction more flexible and convenient, improving a teaching effect.

BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWINGS

Drawings herein are introduced briefly for clearer illustration of a technical solution of the present disclosure. Note that the drawings described below refer merely to some embodiments of the present disclosure. For a person having ordinary skill in the art, other drawings may be acquired according to the drawings herein without creative effort.

FIG. 1 is a flowchart of a multimedia interaction method according to an embodiment of the present disclosure.

FIG. 2 is a diagram of a structure of connection between a web front end and a local computer in a multimedia interaction method of the present disclosure.

FIG. 3 is a flowchart of a multimedia interaction method according to an embodiment of the present disclosure.

FIG. 4A is a flowchart of a multimedia interaction method according to an embodiment of the present disclosure.

FIG. 4B is a diagram of an overall framework of a multimedia interaction method of the present disclosure.

FIG. 5 is a diagram of a structure of a multimedia interaction device of the present disclosure.

FIG. 6 is a diagram of a structure of multimedia interaction equipment of the present disclosure.

FIG. 7 is a diagram of a structure of a computer-readable storage medium of the present disclosure.

DETAILED DESCRIPTION

Clear complete description to a technical solution of the present disclosure is given below with reference to the drawings and embodiments of the present disclosure. Clearly, embodiments illustrated herein are but some, instead of all, embodiments according to the present disclosure. Based on embodiments of the present disclosure, a person having ordinary skill in the art may acquire another embodiment without creative effort. Any such embodiment falls within the scope of the present disclosure.

The present disclosure will be elaborated below with reference to the drawings and embodiments.

Multimedia interaction performed on a teaching platform mainly refers to man-machine interaction by way of image capturing, (audio) recording, video recording, screenshot, etc., during teaching. However, computers of a student and a teacher may differ in configuration, making multimedia interaction difficult. Therefore, embodiments of the present disclosure provide a multimedia interaction method. Having acquired a control instruction input by a user using a multimedia interaction component, a teaching platform may extract information directly based on the control instruction via the multimedia interaction component, preprocess the information using a preset code, and then display or play the processed information. Therefore, the information may be acquired locally and processed locally, and multimedia interaction may be performed without using external equipment.

FIG. 1 is a flowchart of a multimedia interaction method according to an embodiment of the present disclosure. The method may include a step as follows.

At S11, a multimedia interaction component of a teaching platform is called.

The teaching platform is a network teaching system to which one may log in through a local browser, such as a platform that teaches programming, a platform that teaches AI, etc. In an embodiment, the multimedia interaction component may be a component in the teaching platform that performs a processing operation such as multimedia information acquisition. For example, the multimedia interaction component may be a program code in

the teaching platform that has a function such as calling a local camera to acquire an image and transmitting the image to a local server, i.e., a server run on a local computer, etc. Call of the multimedia interaction component may be triggered according to an operation by a user on the teaching platform. A local server may be a developed preset software (equivalent to a local engine). Local equipment (the local computer) may log in to the teaching platform and download the preset software to the local computer and run the preset software on the local computer. The local server may be configured to implement a function of the multimedia interaction component.

When the multimedia interaction component is to be started, such as when a user clicks on a multimedia interaction component button on the teaching platform or inputs a user code for calling the multimedia interaction component, etc., the teaching platform may call the multimedia interaction component. A web front end of the teaching platform may be connected to the local server. In one embodiment, the web front end is connected to the local server via a communication interface. For example, the web front end 201 is connected to the local server 202 via a socket input/output 203 (socket IO), as shown in FIG. 2.

In an embodiment, the web front end may be a browser. The browser may be a general-purpose browser on a computer, such as a 360 browser, a Baidu browser, a Google browser, a QQ browser, a Sogou browser, etc. The browser may be another type of browser, which is not limited herein. In another embodiment, the web front end may also be application (APP) software, such as a third-party application of smart equipment. The web front end may be an interface that teaches programming for application software or a browser.

At S12, a control instruction is acquired using the multimedia interaction component.

The control instruction may be a control instruction input by a user, such as a code instruction input by the user, or a control instruction input remotely through other equipment. Of course, the control instruction may be a voice control instruction, or a manually triggered control instruction, such as a control instruction triggered by pressing a key. In another embodiment, the control instruction may also be an automatically triggered control instruction. For example, a timer may be provided in the multimedia interaction component. A control instruction may be generated upon expiration of the timing period of the timer.

Alternatively, when a user enters an experimental course requiring multimedia interaction, the teaching platform may automatically generate a control instruction corresponding to the experimental course.

In a possible implementation, when a user inputs a control instruction through the web front end, the teaching platform may acquire the control instruction using the multimedia interaction component.

The control instruction may include at least one of an image capture control instruction, a video recording control instruction, or an audio recording control instruction. That is, after acquiring the control instruction, the teaching platform may perform an operation such as image capture, audio recording, video recording, screenshot acquisition, etc., via the multimedia interaction component.

At S13, Information is extracted based on the control instruction using the multimedia interaction component.

After receiving the control instruction through the multimedia interaction component, the teaching platform may extract the information according to the control instruction using the multimedia interaction component. In one embodiment, if the control instruction is an image capture control instruction, a camera device may be triggered to acquire an image according to the image capture control instruction. For example, the multimedia interaction component may be triggered to control a local camera device to capture an image. The camera device may be a camera included in the local computer equipment logged on to the teaching platform, or may be an external camera etc. In an embodiment, when capturing an image according to the image capture control instruction, the timing period of a timer may be set. When the timing period of the timer expires, the multimedia interaction component may be controlled to extract information. For example, if the timing period of the timer is set to 5 seconds, the camera device may be automatically started to capture an image upon the fifth second after receiving the control instruction. 5 seconds after the image has been captured, the camera device may be started again to capture an image. The image capturing process may end when a next control instruction is received. Alternatively, an image capture number may be set while setting the timing period, and image capture may stop automatically when reaching the image capture number.

If the control instruction is a video recording control instruction, the teaching platform may trigger, according to the video recording control instruction using the multimedia interaction component, the camera device, such as an external or local camera logged on to the teaching platform, to acquire multiple frames of images. The multiple frames of images may be linked together to form a video. In an embodiment, when video recording is performed according to a video recording control instruction, a timing period of a timer may also be set. Upon expiration of the timing period of the timer, the multimedia interaction component may be controlled to extract the information. For example, if the timing period of the timer is set to 5 seconds, the camera device may be automatically started to record a video 5 seconds after the control instruction has been received. 5 seconds after complete of the video recording, the camera device may be started again to record a video. The video recording may end upon receipt of the next control instruction. Alternatively, an image capture number may be set while setting the timing period, and image capture may stop automatically when reaching the image capture number. In an embodiment, during video recording, audio may be blocked out, and no audio recording is performed, i.e., the recorded video may include no sound information. In other embodiments, while video recording is performed, audio recording may be performed at the same time, i.e., the recorded video may further include sound information.

If the control instruction is an audio recording control instruction, the teaching platform may trigger, according to the audio recording control instruction using the multimedia interaction component, a recording device such as a recorder, a microphone, etc., to perform audio recording. Likewise, when audio recording is performed, a timing period of a timer may also be set. Upon expiration of the timing period of the timer, the multimedia interaction component may be controlled to extract the information. For example, if the timing period of the timer is set to 5 seconds, the recording device may be automatically started to record an audio 5 seconds after the control instruction has been received. 5 seconds after complete of the audio recording, the recording device may be started again to record an audio. The audio recording may end upon receipt of the next control instruction. Alternatively, a recording number may be set while setting the timing period, and audio recording may stop automatically when reaching the recording number. It should be noted that, in an embodiment,

during audio recording, the camera device may also be started to record a video, and the acquired audio may further include image information.

Further refer to FIG. 1. At S14, the information is displayed or played through the multimedia interaction component.

If the information is information on an image, i.e., photo information, after an image has been acquired, the image may be displayed through the multimedia interaction component. In one embodiment, multiple frames of images may also be acquired continuously within a preset period of time, and displayed sequentially.

The multiple frames of images may be displayed according to a set time frequency, or played and displayed according to an image frame transmission rate. A number of Frames Per Second (FPS) of the multiple frames of images may be acquired. The multiple frames of images may be selectively played according to a correspondence between a frame number and time through the number of FPS. For example, a total of 1,000 frames of images are acquired, with 200 FPS. To play the video of the third second, playback may start directly from the 401st frame. Understandably, to play the 401st frame, one may directly drag the video frame to the 3rd second.

In another embodiment, the timing period of the timer may be set. At a timing period, one of the multiple frames of images may be acquired and displayed, and iteration may be performed in sequence. For example, if the timing period is 1 second, one frame of image may be played per second, thereby slowing down the speed of playing the video, allowing clear view and improved memory during interaction. In another embodiment, during video playback, a screenshot may be taken from a video being played. For example, an image may be acquired from the multiple frames of images being played or the image being played, that is, a screenshot may be taken. After a present image has been acquired as a screenshot, the image may be simultaneously displayed through a display window of the multimedia interaction component, or an image processing operation such as target recognition may be performed on the screenshot image.

If the information is audio information, the audio information may be played through the display window of the multimedia interaction component.

In an embodiment, during playback through the multimedia interaction component,

the recorded audio may also be converted into text. That is, while the audio is being played, the audio converted text result corresponding to the audio may be displayed, facilitating understanding the meaning of the audio. Of course, in another embodiment, just the recorded audio file is converted into text, with no audio playback, so that audio content may still be acquired in an environment in which it is not suitable to play the audio.

In one of the application scenes, in case the equipment used by the user cannot be networked or the user does not have access to an external network, the information may be extracted from pre-stored preset information based on the control instruction using the multimedia interaction component. In one embodiment, teaching interaction may be performed according to a preset voice imitation recording. In another embodiment, teaching interaction may be performed according to a preset image imitation image acquisition process. For example, the teaching platform may play voice by calling prerecorded voice content. For example, the user may record preset voice content in advance and preprocesses the voice content to imitate a process of audio recording and processing in actual teaching to achieve a teaching goal.

The multimedia interaction component may call a local preset voice to imitate a recording or call a preset image to imitate an image, saving the process of uploading a multimedia file such as a video and a photo of a student to the server for processing. Therefore, the teaching interaction does not rely on a large bandwidth, and has good real-time performance.

An example of the teaching platform is used to clearly illustrate the execution process. If the user has access to an external network, the teaching platform may control to pop up a question window, displaying a question. The user may input an audio recording control instruction through the web front end, i.e., the question window. The teaching platform may call, according to the audio recording control instruction, the multimedia interaction component to directly record voice, and recognize the recorded voice to detect whether it is correct. For example, the teaching platform may pop up a window showing “Who is the author of *The Historical Records*?”. The user may enter an answer “Sima Qian” via the multimedia interaction component. The teaching platform may pre-process the user’s answer using the multimedia interaction component, and identify and verify the user's answer, to

detect whether the answer is correct, enriching teaching interaction modes. In another scene, if the user does not have access to an external network, the teaching platform may pop up and simultaneously displays a plurality of selectable answers according to the preset voice content. For example, the teaching platform may pop up a window “Who is the author of *The Historical Records*?”, and call a plurality of answers recorded in advance in the multimedia interaction component. For example, the teaching platform may pop up a plurality of answers windows such as “Sima Qian, Luo Guanzhong, Shi Nai'an”, etc. The user may directly click on a selected answer. For example, the user clicks on the voice answer “Sima Qian”. In this case, the teaching platform may recognize and verify the answer of the user to detect whether the answer is correct, enriching teaching interaction modes.

In yet another scene, if the user does not have access to an external network, the teaching platform may pop up a voice recording window and call preset voice content, and may also simulate and display the voice recording process, and then call the voice or text information acquired by processing the preset voice content (such as voice noise reduction, voice-to-text conversion, etc.) for playback and/or display, so as to simulate and display the voice recording and processing process without network connection, implementing teaching effect.

With the multimedia interaction method provided in embodiments of the present disclosure, a control instruction is acquired using the multimedia interaction component. Information is extracted based on the control instruction using the multimedia interaction component. The information is displayed or played through the multimedia interaction component. Therefore, multimedia interaction may be completed using a component of a locally run teaching platform, and it is not necessary to extract information by calling external dedicated equipment and upload the extracted information to the local computer, simplifying the operation, making multimedia interaction of more fun, improving teaching effect.

With embodiments of the present disclosure, the local server is loaded on the computer and connected to the browser, implementing multimedia interaction by calling the multimedia interaction component of the teaching platform. The teaching platform in embodiments of the present disclosure may relate to a computer visual scene, such as an

algorithm for face recognition, image recognition, object tracking, etc. When running, the local server can call a camera, a microphone, etc., of the local computer to perform multimedia interaction. The user may independently capture an image and perform recording to acquire a desired video, photo, or audio. Therefore, it is not necessary to call external equipment for information extraction and upload the extracted information to the local computer, simplifying the operation, making multimedia interaction of more fun, improving teaching effect.

In addition, the multimedia interaction component in embodiments of the present disclosure may implement local voice or image processing, and it is not necessary to upload voice or a multimedia file implementing a video, an image, etc., to a server for processing, without relying on large bandwidth, with good real-time performance.

FIG. 3 is a flowchart of a multimedia interaction method according to an embodiment of the present disclosure. S31, S32, and S33 are the same as S11, S12, and S13 in the first embodiment. Refer to FIG. 1 and related text description, which is not repeated here. The embodiment differs from the embodiment corresponding to FIG. 1 in that the embodiment further comprises an option after S33, as follows.

At S34, a preset code may be acquired through the multimedia interaction component. The information may be preprocessed according to the preset code.

After the information has been extracted, the teaching platform may acquire a preset code through the multimedia interaction component, and then preprocess the extracted information according to the preset code. In an embodiment of the present disclosure, the preset code is a code written into the multimedia interaction component. After the multimedia interaction component has extracted information according to the control instruction, the teaching platform may preprocess the information using the preset code written into the multimedia interaction component.

In one embodiment, the preprocessing may include at least one operation as follows. An image processing operation may be performed on one image acquired or multiple frames of images acquired. Alternatively, at least one of voice noise reduction, voice-to-text conversion, or voice synthesis may be performed on acquired audio. For example, if the acquired information is one image or multiple frames of images, a target in the one image or

the multiple frames of images may be recognized or tracked according to a preset code. In an embodiment, the preset code may be a preset model integrated in the multimedia interaction component. The preset model may be a model integrated with a neural network algorithm capable of target recognition or target tracking. Of course the preset model may be integrated with another algorithm model capable of target recognition or target tracking. When the acquired information is audio information, processing such as voice noise reduction, voice-to-text conversion, voice synthesis, etc., may be performed on the audio information according to a preset code. After the audio information has been acquired, the multimedia interaction component may perform noise reduction processing on the audio file through a preset code in order to smooth subsequent audio playback and reduce noise. In an embodiment, if the audio information acquired within a fixed period of time is multiple audios, the multimedia interaction component may perform voice synthesis processing on the audio files through a preset code to facilitate subsequent playback. In another embodiment, the multimedia interaction component may convert the acquired audio information into text, and then display the text. Voice-to-text conversion may be performed during audio playback. Of course, voice playback may be performed after the voice-to-text conversion processing and the text display, which is not limited herein.

In an embodiment, to better match information interaction and information extraction in the multimedia interaction component to a requirement of an AI algorithm, when extracting the information, the channel number and the bit rate of the extracted image, video, audio, etc., have to match a channel number and a bit rate supported by the multimedia interaction component. A parameter such as a bit rate and a channel number may be set according to a user code or a preset processing algorithm, to reduce the probability of occurrence of a recognition error in preprocessing the extracted information.

In another embodiment of the present disclosure, the preset code may be a code input by a user. For example, a user may input a user code in the teaching platform according to an information processing operation (such as a target detection operation) to be performed. The code may indicate a type of a preprocessing operation to be performed. After acquiring the user code, the multimedia interaction component may analyze the user code to acquire a preprocessing operation to be performed on the information, and then call an algorithm

module corresponding to the preprocessing operation to perform information preprocessing. The preset code may be a code instruction acquired when the control instruction is acquired. When acquiring the control instruction, the multimedia interaction component may perform analysis to acquire both the control instruction and the preset code.

At S35, the preprocessed information may be displayed or played through the multimedia interaction component.

In an embodiment, after video recording, audio extraction, image capture, and preprocessing have been completed, the acquired audio, video, and image may also be saved in a designated folder. For example, when video recording is performed, after recording is completed, the recorded video is stored in a designated folder. In playback, the video in the folder is automatically opened and played. Alternatively, in an embodiment, after video recording, audio extraction, image capture, and preprocessing have been completed, the acquired audio, video, and image may be directly displayed or played without being stored.

In an embodiment, if the extracted information is multiple frames of images, i.e., video capture information, multiple frames of images are acquired, and then played and displayed through the multimedia interaction component. In an embodiment, after acquiring the multiple frames of images constituting a video through the multimedia interaction component, in particular, after performing preprocessing on the acquired multiple frames of images, i.e., image processing (such as target detection, target recognition, etc.), the processed multiple frames of images may form an image set. An image in the image set per se may include no time information. However, each image is acquired at a time point. That is, each image has an acquisition time corresponding to the image. Therefore, in order to allow smooth playback of the video, when an image is acquired, the time of acquiring each image may be recorded. In playback, the acquired multiple frames of images may be directly played and displayed in time sequence of image acquisition, so that the multiple processed images are played smoothly while saving the time and process for saving the images into a video file. In other embodiments, the acquired multiple frames of images may be combined into a video file, and then the video file may be played and displayed.

With a multimedia interaction method according to embodiments of the present disclosure, a control instruction is acquired through a locally run multimedia interaction

component. Information is extracted based on the control instruction using the multimedia interaction component. Further, the acquired information is preprocessed using a preset code via the multimedia interaction component. For example, an image processing operation such as target recognition, target tracking, etc., is performed on one or more images acquired. Alternatively, any one or more operations of voice noise reduction, voice-to-text conversion, voice synthesis, etc., may be performed on the acquired audio. After the preprocessing, display or playback is performed by the multimedia interaction component. Therefore, it is not necessary to call external equipment for information extraction and upload the extracted information to the local computer, simplifying the operation, making multimedia interaction of more fun, improving teaching effect.

FIG. 4A is a flowchart of a multimedia interaction method according to an embodiment of the present disclosure. S41, S42, S43, S44, and S45 are the same as S31, S32, S33, S34, and S35 described in FIG. 3A, with the different in that the embodiment may further include a step as follows.

At S46, a window for displaying or playing the information may be controlled to perform any one of displacement, window scaling, or window hiding.

When the multimedia interaction component performs display or playback, a display or playback window may be dragged to change the location of the window. Alternatively, the display or playback window may be scaled to change the size of the window. Alternatively, the display or playback window may be hidden to allow more flexible multimedia interaction. For example, if the display or playback window blocks the view of a display interface performing teaching, live streaming, chatting, etc., the interface with blocked view may be displayed by an operation such as changing the location of the window, hiding the window, scaling the window, etc. As another example, if the display or playback window is too small and the video and image playback interface is not clear, the window may be scaled to enlarge the window.

In a possible implementation, a series of multimedia interactive interfaces based on a PyQt platform are implemented. The interfaces may mainly involve multimedia interactive interfaces such as those for image capture, video recording, audio recording, screenshot acquisition, audio playback, video playback, etc. The multimedia interaction component may

be directly packaged into an installation package. After the local computer has logged on to the teaching platform, and downloaded and installed the component, multimedia interaction such as image capture, video recording, audio recording, etc., may be performed directly without depending on another decoder, etc. The operation is simple, clear and easy, facilitating implementation of multimedia interaction in online education. For example, with the instructional platform design, research, and development, there is a local engine installed on the local computer. The local engine may be started as a bearer of a service for running a user written code, and communicate with a browser front end through a socket input/output (I/O). The teaching platform may also be designed to encapsulate a multimedia interactive code library. Algorithm teaching involving a computer visual scene such as face recognition, image recognition, object tracking, etc., may call the multimedia interactive code library via a local engine. In the process that a student writes a user code, it is supported to call an encapsulated algorithm module. Interaction via a pop-up window may be performed when a code is running. The student may independently capture an image and perform recording to acquire the video and photo resources the student desires. If an object is to be selected, a screenshot function may be called by dragging the mouse to acquire a screenshot. Then, an algorithm of a course may be called to perform processing. Finally, an interface for playing a video or displaying an image may be called to display the final result of the algorithm.

During network teaching, one may log on to the teaching platform on a computer, and input a first code such as a computer programming language (Python) code in the teaching platform. By running the code by a local server run by the local computer, the multimedia interaction component may be retrieved, and a multimedia interactive interface (such as a main interface of the PyQt) may be displayed. Based on the multimedia interaction component, the student may independently perform an operation of image capture, video recording, audio recording, etc.

In a possible implementation, the user inputs an audio recording control instruction. The multimedia interaction component may record voice based on the control instruction. After the recording is completed, an operation such as voice synthesis, voice analysis, voice-to-text conversion, etc., may further be performed on the audio using a second code. When playing the recorded audio, a play window may also be displayed. The play window

may include a play speed, a play progress bar, etc. The user may also control to perform an operation such as moving, scaling, hiding, etc., on the play window. When the recorded audio is played, a segment of the audio may also be clipped. Then the clipped audio may be processed. In one embodiment, the audio may or may not include a video image. For example, in audio recording, an audio processing module (pyAudio) may be used to acquire the audio. Then, a ripple animation (wave) component may convert the acquired audio into a standard file format in recording, which supports setting of a parameter such as a bit rate, a channel number, etc., supporting a demand of an AI algorithm more flexibly. In audio playback, audio playback may be performed using a new media (Qmedia) component using a standard audio playback format. Main functions of playback, pause, drag, and time display may be implemented on the main interface.

In a possible implementation, the user inputs an image capture control instruction. The multimedia interaction component may capture an image based on the control instruction. After the image capture is completed, an operation such as target recognition, target tracking, etc., may further be performed on the image using a second code. When displaying the captured image, a screenshot of the displayed image may be clipped to acquire another image. Then, the clipped image may be processed. The user may also control to perform an operation such as moving, scaling, hiding, etc., on the image display window. For example, when capturing an image, a cross-platform computer vision library such as OpenCV may be triggered regularly to acquire an image by a camera based on a timer such as QTimer. Then, the image may be displayed on a multimedia interactive interface such as the main interface of the PyQt platform. In image capture, a present frame triggering image capture may be stored.

In a possible implementation, the user inputs a video recording control instruction. The multimedia interaction component may record a video based on the control instruction to acquire multiple frames of images. After the video recording is completed, an operation such as target recognition, target tracking, voice synthesis, voice analysis, voice-to-text conversion, etc., may further be performed on the video using a second code. When displaying and playing the captured video, a screenshot of the displayed and played video may be clipped to acquire another image. Then, the clipped image may be processed. The user may also control

to perform an operation such as moving, scaling, hiding, etc., on the video playback window. For example, when recording a video, a cross-platform computer vision library may be triggered regularly to acquire an image by a camera based on a timer. Then, the image may be displayed on the main interface of the PyQt platform. When recording a video, each frame may be written to a local video file through the cross-platform computer vision library. In image capture, a present frame triggering image capture may be stored. When playing a video, the video file may be opened based on the cross-platform computer vision library. A timer may perform regular trigger to acquire an image according to Frames Per Second (FPS) information of the video. The image may be displayed on the main interface. In addition, a function such as playback, pause, etc., may be supported. Based on a special teaching scene, the multimedia interaction component may acquire an image set after processing the video through an algorithm. The multimedia interaction component may further support an image set as an input parameter, which may be played without being saved in an addition as a video file. That is, the multimedia interaction component may acquire and display a frame of the image set as regularly triggered by a timer, with sequential iteration. Moreover, an effect similar to pause, progress bar dragging, etc., in a video player may be implement, so as to implement more flexible multimedia display.

With the multimedia interaction method of the embodiment shown in FIG. 1 to the embodiment shown in FIG. 3 according to the present disclosure, information extraction (such as video recording, image capture, audio recording, etc.) may be performed by a multimedia interaction component. Multimedia interaction may be performed using a local resource, which does not need purchase of additional teaching and research equipment. In addition, the extracted information may be pre-processed by the multimedia interaction component, and does not have to be uploaded to a cloud, occupying no additional broadband, saving a computing resource and a network resource. With the present disclosure, video recording, audio recording, image capture, screenshot, etc., may be implemented based on a local computer, needing no third party decoder and program. The multimedia interaction component may be called provided that a web page is installed on the local computer such that the web page runs a first code through a local server, rendering a low requirement on the host computer as well as a green operating environment. The multimedia interaction

component may perform multiple types of multimedia interaction, better meeting a teaching demand, reducing additional conversion steps compared to multimedia interaction performed with the help of external equipment.

The multimedia interaction method is described below with reference to an example. However, it should be noted that the example is merely for better illustrating embodiments of the present disclosure, and does not constitute an improper limitation to embodiments of the present disclosure.

In recent years, with the popularization of AI, machine learning and AI programming education have gradually become an important part of online education. For programming teaching aimed at students, especially young students, some fun and interaction may be added in general to attract students' attention to better achieve a teaching goal. Multimedia interaction is an essential part.

Multimedia interaction mainly refers to a scene such as image capture, audio recording, video recording, screenshot, etc. In teaching some AI algorithms of computer vision, if a student may independently capture an images for image processing, such as face recognition, object recognition, interaction using video and screenshot, performing object tracking, teaching may be made of a lot more fun. In teaching an algorithm relating to natural language processing, education by demonstrating voice recognition may be performed using recording. Language control may be experienced with reference to voice-to-text conversion. In addition, voice generation may be experienced using the play function.

AI teaching inevitably involves some teaching content on computer vision, natural voice processing. In order to add interactivity and fun, it may be an essential function to support for a student or a teacher to record an audio, capture an image, record a video, etc., to perform algorithm presentation. With some existing methods, complete embedded equipment is provide for video recording, audio recording, etc., requiring additional purchase of teaching and research equipment. At the same time, after local equipment collects a video, an image, etc., the video, image, etc., may have to be uploaded to the cloud for algorithm processing, before a result is returned. Additionally, for some no local green operating environment is provided, and normal operation needs additional installation dependency.

With the example, a series of multimedia interactive interfaces based on a PyQt

platform are implemented. The interfaces may mainly involve multimedia interactive interfaces such as those for image capture, video recording, audio recording, screenshot acquisition, audio playback, video playback, etc., implementing interaction using a local resource, requiring additional purchase of teaching and research equipment. At the same time, after the multimedia interactive interface has been directly packaged into an installation package, and installed in the local equipment, multimedia interaction such as image capture, video recording, audio recording, etc., may be performed directly without depending on another decoder, etc. The operation is simple, clear and easy, which is an excellent implementation of multimedia interaction and online education. Finally, with the example, deep customization is performed with reference to an education scene.

The multimedia interactive interface provided in the example may be implemented based on a PyQt platform. As shown in FIG. 4B, a web front end 410 (equivalent to the web front end) is connected to a local engine 420 through a communication interface. The web front end 410 initiates scheduling. That is, the local engine 420 is triggered to run on the local computer according to an operation of a user on the web front end of the teaching platform. The local engine 420 may call a multimedia Application Programming Interface (API) to pop up a display window 430. The display window 430 is a multimedia interactive interface configured to play or display the extracted information, such as to display a captured image. In this case, the user may also control to move, scale, hide the display window, etc. The multimedia interaction component may perform multiple types of multimedia interaction, better meeting a teaching demand, reducing additional conversion steps compared to multimedia interaction performed with the help of external equipment.

The web front end 410 may be an interface that teaches programming for application software or a browser.

The local engine 420, that is, a local server, may be a developed preset software. Local equipment, i.e., the local computer, may log in to the teaching platform and download the local engine 420 to the local computer and run the local engine on the local computer.

The display window 430 may also be a question window, displaying a question. For example, the user may input an audio recording control instruction through the question window. According to the audio recording control instruction, the local engine 420 may

directly record voice using an encapsulated computer programming language function (python) function (equivalent to the multimedia interaction component), and recognize the recorded voice to detect whether it is correct.

As a result, interaction is performed using a local resource of the local computer. Computer vision computation is performed on the local computer, and the multimedia interaction of video recording and audio recording is also performed on the local computer, occupying no additional bandwidth, saving a server computing resource and a network bandwidth resource.

Next, the basic implementation of video interaction is described taking video recording and playback of a teaching scene as an example. That is, the extracted information includes just images and includes no audio information.

In video recording and image capture, a cross-platform computer vision library may be triggered periodically to acquire an image of a camera based on a timer. Then, the image may be displayed on a main interface of a multimedia interactive interface. For video recording, each frame may be written to a local video file through the cross-platform computer vision library. For image capture, a present frame triggering image capture may be saved.

In video playback, a video file may be opened based on the cross-platform computer vision library. A timer may perform regular trigger to acquire an image according to Frames Per Second (FPS) information of the video. The image may be displayed on the main interface. In addition, a function such as playback, pause, etc., may be supported. In this way, video recording, playback, and screenshot acquisition may be implemented based on the cross-platform computer vision library, instead of any third-party decoder and program, rendering the operating environment green and controllable. That is, by being green, usage is allowed upon installation, and the installation is convenient, with a low requirement on the configuration of the host computer.

Based on a special teaching scene, an image set may be acquired after processing the video through an algorithm. The interface may further support an image set as an input parameter, which may be played without being saved in an addition as a video file. The main principle lies in that a frame of the image set may be acquired and displayed according to

regular trigger from a timer, with sequential iteration. Moreover, an effect similar to pause, progress bar dragging, etc., in a video player may be implemented, so as to implement more flexible multimedia display. In this way, deep customization is performed with reference to an education scene, better matching a teaching need, reducing additional conversion steps.

Next, the basic implementation of audio interaction is described taking audio recording and playback of a teaching scene as an example.

In audio recording, an audio processing module may be used to acquire the audio. Then, a ripple animation component may convert the audio into a standard file format, which supports setting of a parameter such as a bit rate, a channel number, etc., supporting a demand of an AI algorithm more flexibly.

In playback, audio playback may be performed using a new media component. Main functions of playback, pause, drag, and time display may be implemented on the main interface.

In recognition, a voice-to-text function may be supported after audio recording. A main function of the voice-to-text function may be implemented based on an open Application Programming Interface (API) at a cloud platform. When the user has no access to an external network, the interaction from audio recording to voice recognition may be imitated according to the preset voice content to achieve a goal of education. In this way, a computer vision-related algorithm may perform computation processing directly on the local computer without any upload to the cloud, occupying no additional bandwidth, saving a computing resource.

With multimedia interactive implementation of a teaching platform, a student may capture an image, record a video, record an audio independently to learn a relevant algorithm, instead of just presetting a lot of media content by the teaching platform, making it more flexible and of more fun, solving a problem such as inconsistent configurations and inconsistent environments of computers of students, etc. Multimedia interaction is implemented without the need to install an additional decoder or another dependent installation. There is no need to upload a multimedia file such as a student's video and photo to a server for processing, relying on no large bandwidth, with good real-time performance.

With embodiments of the present disclosure, local (audio) recording equipment as

well as local video recording equipment may be called, needing no additional dedicated hardware equipment, and interaction may be performed with the local equipment to achieve a teaching goal. With embodiments of the present disclosure, a user does not have to upload a multimedia file acquired by image capture, video recording, audio recording, etc., to a network. Everything may be taught and demonstrated locally. At the same time, optimized customization may be performed according to a scene. For example, video playback supports both a video and an image set. This is because in an AI teaching scene, each frame of a raw video may be analyzed, producing an image set after the processing is done. A video playback interface may support an image set as an input parameter to be played directly, supporting pause, progress drag and drop, etc., implementing more flexible interaction.

Embodiments of the present disclosure may be applied to a computer vision scene such as algorithm teaching such as face recognition, image recognition, object tracking, etc. While a student is writing a user code, an encapsulated windows program developing module (equivalent to the multimedia interaction component) may be called. Pop-up interaction may be performed when the code runs. The student may independently capture an image and perform recording to acquire a desired video and photo resource. If an object is to be selected, a screenshot function may be called by dragging the mouse to acquire a screenshot. Then, an algorithm of a course may be called to perform processing. Finally, an interface for playing a video or displaying an image may be called to display the final result of the algorithm. Embodiments of the present disclosure may also be applied to a natural voice processing scene such as a voice command, voice synthesis, etc.

FIG. 5 is a diagram of a structure of a multimedia interaction device of the present disclosure. The multimedia interaction device may include a calling module 51, an inputting module 52, an information extracting module 53, and an outputting module 54.

The calling module 51 is configured to call a multimedia interaction component of a teaching platform.

The teaching platform is a network teaching system to which one may log in through a local browser, such as a platform that teaches programming, a platform that teaches AI, etc. In an embodiment, the multimedia interaction component may be a component in the teaching platform that performs a processing operation such as multimedia information

acquisition. For example, the multimedia interaction component may be a program code in the teaching platform that has a function such as calling a local camera to acquire an image and transmitting the image to a local server, i.e., a server run on a local computer, etc. Call of the multimedia interaction component may be triggered according to an operation by a user on the teaching platform. A local server may be a developed preset software (equivalent to a local engine). Local equipment (the local computer) may log in to the teaching platform and download the preset software to the local computer and run the preset software on the local computer. The local server may be configured to implement a function of the multimedia interaction component.

In an embodiment, the web front end may be a browser. The browser may be a general-purpose browser on a computer, such as a 360 browser, a Baidu browser, a Google browser, a QQ browser, a Sogou browser, etc. The browser may be another type of browser, which is not limited herein. In another embodiment, the web front end may also be application (APP) software, such as a third-party application of smart equipment. The web front end may be an interface that teaches programming for application software or a browser.

The calling module 51 may be further configured to call the multimedia interaction component to preprocess the extracted information according to the preset code. The preprocessing may include at least one operation as follows. An image processing operation may be performed on one image acquired or multiple frames of images acquired. Alternatively, at least one of voice noise reduction, voice-to-text conversion, or voice synthesis may be performed on acquired audio. For example, if the acquired information is one image or multiple frames of images, a target in the one image or the multiple frames of images may be recognized or tracked according to a preset code. In an embodiment, the preset code may be a preset model integrated in the multimedia interaction component. The preset model may be a model integrated with a neural network algorithm capable of target recognition or target tracking. Of course the preset model may be integrated with another algorithm model capable of target recognition or target tracking. When the acquired information is audio information, processing such as voice noise reduction, voice-to-text conversion, voice synthesis, etc., may be performed on the audio information according to a

preset code. After the audio information has been acquired, the multimedia interaction component may perform noise reduction processing on the audio file through a preset code in order to smooth subsequent audio playback and reduce noise. In an embodiment, if the audio information acquired within a fixed period of time is multiple audios, the multimedia interaction component may perform voice synthesis processing on the audio files through a preset code to facilitate subsequent playback. In another embodiment, the multimedia interaction component may convert the acquired audio information into text, and then display the text. Voice-to-text conversion may be performed during audio playback. Of course, voice playback may be performed after the voice-to-text conversion processing and the text display, which is not limited herein.

The inputting module 52 is configured to acquire a control instruction using the multimedia interaction component. The inputting module 52 may be configured to acquire at least one of an image capture control instruction, a video recording control instruction, or an audio recording control instruction. The control instruction may control the multimedia interaction component to perform an operation such as image capture, audio recording, video recording, screenshot acquisition, etc.

The inputting module 52 may be further configured to acquire a preset code through the multimedia interaction component. After the information has been extracted, the teaching platform may acquire a preset code through the multimedia interaction component, and then preprocess the extracted information according to the preset code. In an embodiment of the present disclosure, the preset code is a code written into the multimedia interaction component. After the multimedia interaction component has extracted information according to the control instruction, the teaching platform may preprocess the information using the preset code written into the multimedia interaction component.

The information extracting module 53 is configured to extract information based on the control instruction using the multimedia interaction component. The information extracting module may be configured to implement at least one of: in response to the control instruction being the image capture control instruction, triggering capture of an image by a camera device according to the image capture control instruction; in response to the control instruction being the video recording control instruction, triggering capture of multiple

frames of images by the camera device according to the video recording control instruction; or in response to the control instruction being the audio recording control instruction, triggering audio recording by a recording device according to the audio recording control instruction. After receiving the control instruction through the multimedia interaction component, the teaching platform may extract the information according to the control instruction using the multimedia interaction component. In one embodiment, if the control instruction is an image capture control instruction, a camera device may be triggered to acquire an image according to the image capture control instruction. If the control instruction is a video recording control instruction, the teaching platform may trigger, according to the video recording control instruction using the multimedia interaction component, the camera device, such as an external or local camera logged on to the teaching platform, to acquire multiple frames of images. The multiple frames of images may be linked together to form a video. If the control instruction is an audio recording control instruction, the teaching platform may trigger, according to the audio recording control instruction using the multimedia interaction component, a recording device such as a recorder, a microphone, etc., to perform audio recording.

The information extracting module 53 may be further configured to acquire an image from the multiple frames of images being played or the image being played. When playing or displaying a video or an image, the multimedia interaction component may acquire an image from the multiple frames of images being played or the image being played, to acquire a screenshot. The information extracting module 53 may be further configured to set a timing period of a timer according to the control instruction; and in response to expiration of the timing period of the timer, control extraction of the information by the multimedia interaction component. When the timing period of the timer expires, the multimedia interaction component may be controlled to extract information. For example, if the timing period of the timer is set to 5 seconds, information extraction such as image capture, video recording, audio recording, etc., may be started automatically upon the fifth second after receiving the control instruction. 5 seconds after completing information extraction, information extraction may be started again. The information extraction process may end when a next control instruction is received. Alternatively, an information extraction number may be set while

setting the timing period, and information extraction may stop automatically when reaching the information extraction number.

The outputting module 54 is configured to display or play the information through the multimedia interaction component. The outputting module 54 may display multiple frames of images according to a set time frequency, or play and display the multiple frames of images according to an image frame transmission rate.

In one embodiment, after acquiring multiple frames of images constituting a video through the multimedia interaction component, the teaching platform further acquires a number of FPS of the multiple frames of images; and selectively plays the multiple frames of images according to a correspondence between a frame number and time through the number of FPS. For example, a total of 1,000 frames of images are acquired, with 200 FPS. To play the video of the third second, playback may start directly from the 401st frame. Understandably, to play the 401st frame, one may directly drag the video frame to the 3rd second.

In another implementation, the outputting module 54 is further configured to display or play the preprocessed information through the multimedia interaction component. The preprocessing may include at least one of: performing an image processing operation on one image acquired or multiple frames of images acquired; or performing at least one of voice noise reduction, voice-to-text conversion, or voice synthesis on acquired audio. For example, if the acquired information is one image or multiple frames of images, an image processing operation such as target recognition or target tracking may be performed on a target in the one image or the multiple frames of images according to a preset code. If the extracted information is multiple frames of images, i.e., video capture information, multiple frames of images are acquired, and then played and displayed through the multimedia interaction component. In an embodiment, after acquiring the multiple frames of images constituting a video through the multimedia interaction component, in particular, after performing preprocessing on the acquired multiple frames of images, i.e., image processing (such as target detection, target recognition, etc.), the multiple frames of images may form an image set. The image set per se may include no time information. However, each image is acquired at a time point. That is, each image has an acquisition time corresponding to the image.

Therefore, in order to allow smooth playback of the video, when an image is acquired, the time of acquiring each image may be recorded. In playback, the acquired multiple frames of images may be directly played and displayed in time sequence of image acquisition, so that the multiple images are played smoothly while saving the time and process for saving the images into a video file. In other embodiments, the acquired multiple frames of images may be combined into a video file, and then the video file may be played and displayed.

In an embodiment, the preset code may be a preset model. The preset model may be a model integrated with a neural network algorithm capable of target recognition or target tracking. Of course the preset model may be integrated with another algorithm model capable of target recognition or target tracking. In another embodiment, if the acquired information is audio information, processing such as voice noise reduction, voice-to-text conversion, voice synthesis, etc., may be performed on the audio information according to a preset code. After the audio information has been acquired, the multimedia interaction component may perform noise reduction processing on the audio file through a preset code in order to smooth subsequent audio playback and reduce noise. In another embodiment, if the audio information acquired within a fixed period of time is multiple audios, the multimedia interaction component may perform voice synthesis processing on the audio files through a preset code to facilitate subsequent playback. In another embodiment, the multimedia interaction component may convert the acquired audio information into text, and then display the text. Voice-to-text conversion may be performed during audio playback. Of course, voice playback may be performed after the voice-to-text conversion processing and the text display, which is not limited herein. The outputting module 54 may be further configured to control a window for displaying or playing the information to perform any one of displacement, window scaling, or window hiding, thereby rendering multimedia interaction more flexible.

With the multimedia interaction device provided by the present disclosure, information extraction (such as video recording, image capture, audio recording, etc.) may be performed by a multimedia interaction component. Multimedia interaction may be performed using a local resource, which does not need purchase of additional teaching and research equipment. In addition, the extracted information may be pre-processed by the multimedia interaction component, and does not have to be uploaded to a cloud, occupying no additional

broadband, saving a computing resource and a network resource. With the present disclosure, video recording, audio recording, image capture, screenshot, etc., may be implemented based on a local computer, needing no third party decoder and program. The multimedia interaction component may be called provided that a web page is installed on the local computer such that the web page runs a first code through a local server, rendering a low requirement on the host computer as well as a green operating environment. The multimedia interaction component may perform multiple types of multimedia interaction, better meeting a teaching demand, reducing additional conversion steps compared to multimedia interaction performed with the help of external equipment.

Embodiments of the present disclosure further provide a computer-readable storage medium, having stored thereon at least an instruction or at least a program which, when loaded and executed by a processor, implements the method herein. The computer-readable storage medium may be a non-volatile computer-readable storage medium.

FIG. 6 is a diagram of a structure of multimedia interaction equipment of the present disclosure. The multimedia interactive equipment includes a memory 62 and a processor 61 connected to each other.

The memory 62 is configured to store program instructions for implementing any multimedia interaction method herein.

The processor 61 is configured to execute the program instructions stored in the memory 62.

The processor 61 may also be referred to as a Central Processing Unit (CPU). The processor 61 may be an integrated circuit chip having signal processing capability. The processor may also be a general purpose processor, a DSP, an Application Specific Integrated Circuit (ASIC), a Field Programmable Gate Array (FPGA) or another programmable logic device, a discrete gate or transistor logic device, a discrete hardware component. The general purpose processor may be a microprocessor or the processor 61 may be any conventional processor, etc.

The memory 62 may be a memory stick, a Trans-Flash (TF) card, etc., and may store all information in multimedia interactive equipment, including input raw data, a computer program, an intermediate execution result, and a final execution result. It stores and retrieves

information according to a location designated by a controller. The memory allows the multimedia interactive equipment to have a memory function to ensure normal operation. The memory of the multimedia interactive equipment may be divided, according to usage, into a main memory (memory) and an auxiliary memory (external memory), or divided into an external memory and (an internal) memory. An external memory is generally a magnetic medium, an optical disc, etc., capable of storing information for a long time. Internal memory refers to a storage unit on a motherboard for storing a program and data being executed, yet just temporarily. The data may be lost upon power-off.

In a number of embodiments provided in the present disclosure, it should be understood that disclosed method and device may be implemented in other ways. For example, described device implementation is merely illustrative. For example, module or unit division is merely logic function division and there may be another division in actual implementation. For example, units or components can be combined, or integrated into another system, or some features / characteristics may be omitted or skipped. Furthermore, the coupling, or direct coupling or communicational connection illustrated or discussed herein may be implemented through indirect coupling or communicational connection among some interfaces, devices, or units, and may be electrical, mechanical, or of another form.

Units described as separate components may or may not be physically separated. Components shown as units may be or may not be physical units. They may be located in one place, or distributed on multiple network units. Some or all of the units may be selected to achieve the purpose of a solution of the present embodiments as needed.

In addition, functional units in embodiments of the present disclosure may all be integrated in one processing unit, or physically exist as separate units respectively; or two or more such units may be integrated in one unit. The integrated unit may be implemented in form of hardware, or hardware plus software functional unit(s).

When implemented in form of a software functional unit and sold or used as an independent product, an integrated unit may be stored in a computer-readable storage medium. Based on such an understanding, the essential part of the technical solution of the present disclosure or a part contributing to prior art or all or part of the technical solution may appear in form of a software product, which software product is stored in storage media, and

includes a number of instructions for allowing computer equipment (such as a personal computer, a system server, network equipment, and/or the like) or a processor to execute all or part of the steps of the methods of the present disclosure.

FIG. 7 is a diagram of a structure of a computer-readable storage medium of the present disclosure. The storage medium of the present disclosure stores a program file 71 capable of implementing any multimedia interaction method herein. The program file 71 may be stored in the storage medium in form of a software product, and includes a number of instructions for allowing computer equipment (such as a personal computer, a server, network equipment, and/or the like) or a processor to execute all or part of the steps of the methods of implementations of the present disclosure. The storage medium includes various media that may store program codes, such as a U disk, a mobile hard disk, Read-Only Memory (ROM), Random Access Memory (RAM), a magnetic disk, a CD, etc., or terminal equipment such as a computer, a server, a mobile phone, a tablet, etc.

What described are merely implementations of the present disclosure, and are not intended to limit the scope of the present disclosure. Any equivalent structure or equivalent flow variation made using content of the specification and/or accompanying drawings of the present disclosure, or direct or indirect application in another related technical field, likewise falls within the scope of the present disclosure.

Industrial applicability

In embodiments of the present disclosure, a multimedia interaction component of a teaching platform is called. A control instruction is acquired using the multimedia interaction component. Information is extracted based on the control instruction using the multimedia interaction component. The information is displayed or played through the multimedia interaction component, implementing multimedia interaction without using external equipment, improving a network teaching effect.

WHAT IS CLAIMED IS:

1. A multimedia interaction method, comprising:

calling a multimedia interaction component of a teaching platform;
acquiring a control instruction using the multimedia interaction component;
extracting information based on the control instruction using the multimedia interaction component; and
displaying or playing the information through the multimedia interaction component.

2. The multimedia interaction method of claim 1, wherein the control instruction comprises at least one of an image capture control instruction, a video recording control instruction, or an audio recording control instruction.

3. The multimedia interaction method of claim 2, wherein extracting the information based on the control instruction using the multimedia interaction component comprises at least one of:

in response to the control instruction being the image capture control instruction, triggering capture of an image by a camera device according to the image capture control instruction;

in response to the control instruction being the video recording control instruction, triggering capture of multiple frames of images by the camera device according to the video recording control instruction; or

in response to the control instruction being the audio recording control instruction, triggering audio recording by a recording device according to the audio recording control instruction.

4. The multimedia interaction method of claim 3, further comprises: after triggering capture of the multiple frames of images by the camera device according to the video recording

control instruction in response to the control instruction being the video recording control instruction,

acquiring a number of Frames Per Second (FPS) of the multiple frames of images;
and

selectively playing the multiple frames of images according to a correspondence between a frame number and time through the number of FPS.

5. The multimedia interaction method of claim 3, wherein displaying or playing the information through the multimedia interaction component comprises: in response to the control instruction being the video recording control instruction or the image capture control instruction,

sequentially playing the multiple frames of images according to timing of acquiring the multiple frames of images; or

combining the multiple frames of images into a video file, and playing the video file.

6. The multimedia interaction method of any one of claims 1 to 5, wherein extracting the information based on the control instruction using the multimedia interaction component comprises:

extracting the information from pre-stored preset information based on the control instruction using the multimedia interaction component.

7. The multimedia interaction method of any one of claims 1 to 5, wherein extracting the information based on the control instruction using the multimedia interaction component comprises:

setting a timing period of a timer according to the control instruction; and

in response to expiration of the timing period of the timer, controlling extraction of the information by the multimedia interaction component.

8. The multimedia interaction method of any one of claims 1 to 7, further comprising: after displaying or playing the information through the multimedia interaction component,

controlling a window for displaying or playing the information to perform any one of displacement, window scaling, or window hiding.

9. The multimedia interaction method of any one of claims 1 to 8, wherein displaying or playing the information through the multimedia interaction component comprises:

acquiring a preset code through the multimedia interaction component, and preprocessing the information according to the preset code; and

displaying or playing the preprocessed information through the multimedia interaction component.

10. The multimedia interaction method of claim 9, wherein the preprocessing comprises at least one of:

performing an image processing operation on one image acquired or multiple frames of images acquired; or

performing at least one of voice noise reduction, voice-to-text conversion, or voice synthesis on acquired audio.

11. The multimedia interaction method of any one of claims 3 to 10, further comprising: in response to the control instruction being the video recording control instruction or the image capture control instruction, after displaying or playing the information through the multimedia interaction component,

acquiring an image from the multiple frames of images being played or the image being played.

12. A multimedia interaction device, comprising a calling module, an inputting module, an information extracting module, and an outputting module,

wherein the calling module is configured to call a multimedia interaction component of a teaching platform,

wherein the inputting module is configured to acquire a control instruction using the multimedia interaction component,

wherein the information extracting module is configured to extract information based on the control instruction using the multimedia interaction component,

wherein the outputting module is configured to display or play the information through the multimedia interaction component.

13. The multimedia interaction device of claim 12, wherein the control instruction comprises at least one of an image capture control instruction, a video recording control instruction, or an audio recording control instruction.

14. The multimedia interaction device of claim 13, wherein the information extracting module is further configured to implement at least one of:

in response to the control instruction being the image capture control instruction, triggering capture of an image by a camera device according to the image capture control instruction;

in response to the control instruction being the video recording control instruction, triggering capture of multiple frames of images by the camera device according to the video recording control instruction; or

in response to the control instruction being the audio recording control instruction, triggering audio recording by a recording device according to the audio recording control instruction.

15. The multimedia interaction device of claim 14, wherein the information extracting module is further configured to implement:

acquiring a number of Frames Per Second (FPS) of the multiple frames of images;
and

selectively playing the multiple frames of images according to a correspondence between a frame number and time through the number of FPS.

16. The multimedia interaction device of claim 14, wherein the information extracting module is further configured to implement, in response to the control instruction being the

video recording control instruction or the image capture control instruction,
sequentially playing the multiple frames of images according to timing of acquiring the multiple frames of images; or
combining the multiple frames of images into a video file, and playing the video file.

17. The multimedia interaction device of any one of claims 12 to 16, wherein the information extracting module is further configured to extract the information from pre-stored preset information based on the control instruction using the multimedia interaction component.

18. The multimedia interaction device of any one of claims 12 to 17, wherein the information extracting module is further configured to implement:

setting a timing period of a timer according to the control instruction; and
in response to expiration of the timing period of the timer, controlling extraction of the information by the multimedia interaction component.

19. The multimedia interaction device of any one of claims 12 to 18, wherein the outputting module is further configured to control a window for displaying or playing the information to perform any one of displacement, window scaling, or window hiding.

20. The multimedia interaction device of any one of claims 12 to 19, wherein the inputting module is further configured to acquire a preset code through the multimedia interaction component, wherein the calling module is further configured to call the multimedia interaction component to preprocess the information according to the preset code, wherein the outputting module is further configured to display or play the preprocessed information through the multimedia interaction component.

21. The multimedia interaction device of claim 20, wherein the preprocessing comprises at least one of:

performing an image processing operation on one image acquired or multiple frames of images acquired; or

performing at least one of voice noise reduction, voice-to-text conversion, or voice synthesis on acquired audio.

22. The multimedia interaction device of any one of claims 14 to 21, wherein the outputting module is further configured to, in response to the control instruction being the video recording control instruction or the image capture control instruction, acquire an image from the multiple frames of images being played or the image being played.

23. Multimedia interaction equipment, comprising a memory and a processor, wherein the memory stores program instructions which, when called by the processor, implement the multimedia interaction method of any one of claims 1 to 11.

24. A computer-readable storage medium, having stored thereon a program file executable to implement the multimedia interaction method of any one of claims 1 to 11.

25. A computer program product, comprising a computer-readable code which, when executed in electronic equipment, allows a processor in the electronic equipment to implement the multimedia interaction method of any one of claims 1 to 11.