

USER INTERACTION METHOD AND SYSTEM

FIELD OF INVENTION

- 5 The present invention relates to a user interaction method and system, specifically applied to virtual objects.

BACKGROUND

- 10 With the rapid development of technology, especially the continuous innovation of artificial intelligence technology, computer-generated virtual objects such as digital humans and virtual humans are becoming increasingly common in many application scenarios. These virtual objects interact with users to meet their information needs in fields such as, for example, gaming, virtual reality, and online education. Virtual objects
15 are typically used for role-playing, teaching assistance, or information display.

- In related technologies, a common technique to enhance the realism of virtual object is to synchronize moving images and audio, ensuring that the lip movements of virtual object in the moving images match voice content in the audio. The implementation of
20 such technology enables the lip movements of the virtual object to realistically simulate the speaking manner of human beings while speaking, thereby improving the realism of virtual object.

- However, even though present synchronization techniques can simulate more natural
25 lip movements, virtual objects still appear stiff and rigid in the overall interaction process. This mainly manifests in the lack of flexibility and diversity in the body movements, facial expressions, and eye expressions of virtual objects compared to real humans. Therefore, despite having natural lip movements, the realism of virtual objects is typically still relatively low, which significantly reduces the user's interactive
30 experience.

Therefore, in order to enhance user interaction experience and meet the growing demand for highly realistic interactions with virtual objects, it is necessary to develop a technical solution that can further improve the overall realism of virtual objects.

SUMMARY

In a first aspect, there is provided a user interaction method applied to virtual object, characterized in that the method is applied to an interaction system comprising an image acquisition device and a screen. It is preferable that the screen is configured to display virtual object in the interaction interface, the image acquisition device is configured to collect real-time dynamic data within the first preset area in real time, and the method comprises:

obtaining real-time dynamic data comprising user object collected by the image acquisition device, wherein the user object is a user who interacts with the virtual object within the first preset area;

determining the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data;

determining the deviation position information of the user object relative to the virtual object based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object;

determining the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation position information; and

adjusting the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

In a second aspect, there is provided an interactive system that preferably comprises a processor, an image acquisition device connected to the processor, and a screen:

the screen being configured to display virtual object in an interactive interface;

the image acquisition device being configured to collect real-time dynamic data within the first preset area in real time;

the processor being configured to obtain real-time dynamic data comprising user object collected in real-time by the image acquisition device; the user object is a user who interacts with the virtual object within the first preset area; determine the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data; determine the deviation position information

of the user object relative to the virtual object based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object; determine the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation
5 position information; adjust the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

There is also provided an electronic device, preferably comprising a processor, a memory, and a bus; whereby the memory is configured to store machine-readable
10 instructions executable by the processor; when the electronic device is operating, the processor communicates with the memory via the bus; when the machine-readable instructions are executed by the processor, the user interaction method applied to virtual object as described in the preceding paragraphs.

15 There is also provided a computer-readable storage medium, whereby the computer-readable storage medium stores a computer program, and when the computer program is run by a processor, it executes the user interaction method applied to virtual object as described in the preceding paragraphs.

20 Finally, there is provided a computer program product comprising a computer program, and when the computer program is run by a processor, it executes the user interaction method as described in the preceding paragraphs.

It will be appreciated that the broad forms of the invention and their respective features
25 can be used in conjunction, interchangeably and/or independently, and reference to separate broad forms is not intended to be limiting.

DESCRIPTION OF FIGURES

30 By reading the detailed description of exemplary embodiments in the following text, those skilled in the art will understand the advantages and benefits described herein, as well as other advantages and benefits. The accompanying drawings are only for the purpose of demonstrating exemplary embodiments and are not considered a

limitation on the application. And throughout all drawings, the same components are represented by the same numbers. In the attached drawings:

FIG 1 is a schematic diagram of a computing object displaying a virtual object based interaction scenario provided in the embodiment of the application;

FIG 2 is a method flowchart of a user interaction method applied to a virtual object provided in the embodiment of the application;

FIG 3 is a schematic diagram of facial recognition detection for real-time dynamic data provided in the embodiment of the application;

FIG 4 is a schematic diagram of target pose adjustment for virtual object provided in the embodiment of the application;

FIG 5 is another schematic diagram of target pose adjustment for virtual object provided in the embodiment of the application; and

FIG 6 is a schematic diagram of an interactive system provided in the embodiment of the application.

DETAILED DESCRIPTION

Exemplary embodiments of the present application will be described in more detail below with reference to all accompanying drawings. Although the exemplary embodiments of the present application are shown in the accompanying drawings, it should be understood that the present application can be implemented in various forms and should not be limited by the embodiments described here. On the contrary, these embodiments are provided to enable a more thorough understanding of the present application and to fully convey the scope of the present application to those skilled in the art.

In the description of the embodiments of the application, it should be understood that terms such as "including" or "having" are intended to indicate the presence of disclosed features, numbers, steps, actions, components, parts, or combinations thereof in this specification, and do not exclude the possibility of the existence of one or more other features, numbers, steps, actions, components, parts, or combinations thereof.

Unless otherwise specified, "/" represents the meaning of "or". For example, A/B can represent A or B; "and/or" in this document is just a way to describe the relationship between associated objects, indicating that there can be three types of relationships. For example, A and/or B can represent three situations: the existence of A alone, the simultaneous existence of A and B, and the existence of B alone.

Terms such as "first", "second", and the like are used to distinguish between identical or similar technical features for descriptive convenience only and should not be interpreted as indicating or implying the relative importance or quantity of these technical features. Thus, features defined by "first", "second", etc., can explicitly or implicitly include one or more of these features. In the description of the embodiments of the application, unless otherwise specified, the term "multiple" means two or more.

It should also be noted that, without conflict, the embodiments and features in the embodiments in the application can be combined with each other. The following will refer to the accompanying drawings and combine embodiments to illustrate the application in detail.

In related technologies, the development of interaction between users and virtual object mainly comprises the following:

- Facial synthesis and expression: Generating highly realistic virtual faces and various expressions to convey emotions. For example, by combining deep learning and computer graphics, precise facial expression synthesis can be achieved to make the virtual object more vivid and trustworthy.

- Voice synthesis and voice recognition: Generating a natural and uninterrupted voice, and achieving interaction between voice recognition and voice commands. For example, by combining deep learning and text to voice synthesis, high-quality voice synthesis can be achieved, enabling the virtual object to have natural voice interaction capabilities.

- Pose and action synthesis: Generating gestures and actions of the virtual object, enabling more realistic motion capabilities. For example, realistic pose and action

synthesis of virtual object can be achieved through physical simulation and motion capture, making virtual object more natural and flexible in the interaction process.

- Emotion and cognition: Simulating the emotional and cognitive abilities of the virtual object, and enabling the ability to express emotions and make considered decisions. For example, through emotional computing and cognitive modeling, the virtual object can become more intelligent and emotional, allowing for deeper interaction with users.

In related technologies, in order to ensure the fidelity of virtual object to a certain extent, moving images and audio are usually synchronized to match the lip movements of virtual object in the moving images with the voice content in the audio, so that the virtual object interacting with the user has realistic and natural lip movements. For example, the voice driven facial model in related technologies, as a deep learning based artificial intelligence model, is based on the core idea of learning the correspondence between audio and moving images through deep learning to achieve synchronization between audio and moving images. It can convert the linguistic information of a segment of audio into matching facial images, so as to match the lip movements of the virtual object with the voice content of the audio.

Although virtual objects in related technologies provide realistic and natural lip movements during interaction with users, these virtual objects are still relatively rigid and have a low degree of realism, which diminishes the user's interaction experience.

For this purpose, the application provides a user interaction method and device applied to a virtual object, applied to an interaction system with an image acquisition device and a screen. The screen is configured to display the virtual object in the interaction interface, and the image acquisition device is configured to collect real-time dynamic data within the first preset area in real time. The method comprises: for user object interacting with virtual object in a first preset area, obtaining real-time dynamic data comprising user object collected in real-time by the image acquisition device; based on the user position feature information and user size feature information of the user object in real-time dynamic data, the first relative position information of the user object relative to the image acquisition device can be determined; based on the first relative position information and the second relative position information of the image

acquisition device relative to the virtual object, the deviation position information of the user object relative to the virtual object can be determined; based on the deviation position information, the first posture rotation coefficient and the second posture rotation coefficient of the virtual object are determined, and the target pose of the virtual object is adjusted.

By using the above method, based on the user position feature information and user size feature information of the user object in real-time dynamic data, the deviation position information of the user object relative to the virtual object can be accurately determined, so that the target pose of the virtual object can be adjusted according to the deviation position information, so that the head of the virtual object can face the user object and the eyeballs of the virtual object can look directly at the user object, improving the realism of the virtual object and ensuring a desirable interaction experience between the user and the virtual object.

The user interaction method applied to virtual object provided in the embodiment of the application can be implemented through an interaction system, which can include a computer device, the computer device can be, for example, an interaction terminal or server. The server can be an independent physical server, a server cluster or a distributed system composed of multiple physical servers, or a cloud server providing cloud computing services. Interaction terminals include but are not limited to mobile phone, computer, intelligent voice interaction device, smart home appliance, vehicle terminal, aircraft, augmented reality device, etc. Interaction terminals and servers can be directly or indirectly connected through wired or wireless communication methods, and the application is not limited here.

It can be understood that in the specific implementation of the application, relevant data such as gender characteristic information and user basic information are involved. When the above embodiments are applied to specific products or technologies, user permission or consent is required, and the collection, use, and processing of relevant data must comply with relevant laws, regulations, and standards of relevant countries and regions.

FIG 1 is a schematic diagram of computing device displaying a virtual object based interaction scenario provided in an embodiment of the application. The computing device can be an interaction terminal 100, which comprises a screen 200 and an image acquisition device 300. The image acquisition device 300 can be detachable, and the screen 200 is configured to display the virtual object on the interaction interface. The image acquisition device 300 is configured to collect real-time dynamic data within the first preset area in real time, and the interaction terminal 100 is used to achieve virtual object based interaction.

Alternatively, the aforementioned computing device can also be servers and interaction terminals. In this case, the interactive system comprises interaction terminals and servers, which can include at least one screen and at least one image acquisition device. The screen is configured to display virtual objects on the interactive interface, and the image acquisition device is configured to collect real-time dynamic data in the first preset area in real time. The server is connected to the interaction terminal comprising the at least one screen and at least one image acquisition device, and the interaction terminal is configured send the real-time dynamic data collected by the image acquisition device to the server. The server is configured to determine the corresponding first posture rotation coefficient and second posture rotation coefficient based on the real-time dynamic data and send them to the interaction terminal, so that the interaction terminal can adjust the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

The following is an explanation of the user interaction method provided in the application through method embodiments, as shown in FIG 2. FIG 2 is a flowchart of a user interaction method applied to virtual object provided in embodiment of the application. The aforementioned computing device can be configured as an interaction terminal, and the method comprises at least steps 201-205.

At step 201, obtaining real-time dynamic data comprising user object collected by the image acquisition device.

A virtual object refers to a computer-generated object in virtual scenes, comprising digital humans and virtual humans. For different scenarios, virtual objects can meet

the information needs of users through real-time interaction. For example, in a shopping mall guide scene, virtual objects can be virtual service personnel, and users can obtain corresponding shopping mall discount information by interacting with virtual service personnel.

5

In related technologies, although realistic and natural lip movements of virtual objects can be achieved, the head orientation and eye direction of the virtual objects are not adjusted. However, in the actual interaction between users and virtual objects, users often do not specifically face the virtual object for interaction, which means the virtual
10 object may not be facing the users nor making eye contact with the users during the interaction. This can result in the virtual object appearing stiff and unnatural during the interaction.

In order to improve interaction with the virtual object and enhance the user's interactive
15 experience, in the embodiment, the interaction terminal may include at least a screen and at least an image acquisition device.

Based on FIG 1, the screen 200 is configured to display virtual objects in an interactive interface. In order to enable user to interact with virtual object, it is necessary to display
20 the virtual object on the interaction interface. In addition, in order to naturally depict the virtual object, the screen can also display a virtual background in the virtual space where the virtual object is located on the interaction interface. For example, when the virtual object is a virtual service personnel and the virtual background in the virtual space corresponding to the virtual object is a virtual service desk, the screen can not
25 only display the virtual service personnel, but also further display the virtual service desk.

Based on FIG 1, the at least one image acquisition device 300 is used to collect real-time dynamic data within a first preset area. The first preset area refers to the area
30 that can respond to the user's interaction behavior with virtual object, for example, the preset area in front of the screen 200 is the first preset area. In one example, the preset area is a circular area with the first preset distance as the radius, such as a circular area with a radius of 3 meters. In another example, the preset area is a square area with a second preset distance as the side length, such as a square area with a

side length of 3 meters. In another example, the preset area is a rectangular area with the third preset distance as the length and the fourth preset distance as the width, such as a rectangular area with a length of 4 meters and a width of 3 meters.

- 5 It should be noted that in the practical operation of the embodiment, the image acquisition device 300 can be arranged above the screen, allowing the image acquisition device to collect real-time dynamic data within the first preset area in real time.
- 10 In the embodiment, the user object refers to the user who interacts with the virtual object within the first preset area.

Multiple consecutive images can form non static real-time dynamic data.

- 15 Due to the fact that the user object does not always interact with the virtual object in the first preset area, when any images in the real-time dynamic data collected by the image acquisition device does not include the user object, it indicates that there is no user object interacting with the virtual object in the first preset area at the corresponding time of the video frame data. In such case, there is no need to adjust
- 20 the virtual object for the video frame data. In other words, processing the moving images that do not contain the user object in the real-time dynamic data collected by the image acquisition device is redundant.

- In order to avoid redundant operations, the interaction terminal can obtain real-time
- 25 dynamic data comprising user object collected by the image acquisition device in real-time. Real-time dynamic data refers to the images of the user object included in the data collected by the image acquisition device in real-time. In other words, the interaction terminal does not obtain all image data in the data collected by the image acquisition device in real-time, but selectively obtains real-time dynamic data
- 30 comprising user object, thereby avoiding redundant operations and reducing resource waste.

It should be noted that in order to ensure a desirable interaction experience between user object and virtual object, for a real-time dynamic data, a number of user objects

corresponding to the virtual objects is pre-defined as one. If there are multiple users interacting with the virtual object simultaneously in the first preset area at a corresponding time of real-time dynamic data, the interaction terminal is configured to select one user object from the multiple user objects interacting with the virtual object as the corresponding user object.

At step 202, determining the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in real-time dynamic data.

User position feature information is used to identify the position of user object in real-time dynamic data. As real-time dynamic data is a two-dimensional plane, user position feature information can be defined using two-dimensional coordinates.

User size feature information is used to identify the ratio between the size of user object in real-time dynamic data and the size of user object in the real world. As the size of user object in real-time dynamic data is usually smaller than the size of user object in the real world, user size feature information is usually less than 1.

The first relative position information is used to identify the position of the user object relative to the image acquisition device. Since the user object and the image acquisition device are both in three-dimensional space, the first relative position information can be defined using three-dimensional coordinates.

The position of the user object in real-time dynamic data and the ratio between the size of the user object in real-time dynamic data and size in the real world are closely related to the position of the user object relative to the image acquisition device. For example, when the user size feature information remains unchanged, the further away the user object is from the center in real-time dynamic data, the more offset the user object is relative to the image acquisition device. When the user object is at the center of the real-time dynamic data, it indicates that the user object is perpendicular to the image acquisition device. Alternatively, when the user position feature information remains unchanged, the smaller the proportion of the user object in real-time dynamic data, the further away the user object is from the image acquisition device.

After obtaining real-time dynamic data in step 201, the interaction terminal is configured to determine the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data. That is, the interaction terminal is configured to determine the position of the user object relative to the image acquisition device from the two dimensions of the position and proportion of the user object in the real-time dynamic data.

10 In one possible implementation, determining the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data, comprises:

15 determining the distance of the user object relative to the image acquisition device based on the user size feature information of the user object in the real-time dynamic data; and

determining the first relative position information of the user object relative to the image acquisition device based on the distance of the user object relative to the image acquisition device and the user position feature information of the user object in the real-time dynamic data.

Specifically, user size feature information is used to identify the ratio between the size of user object in real-time dynamic data and the size of user object in the real world, and the ratio between the size of user object in real-time dynamic data and the size of user object in the real world is closely related to the distance between user object and image acquisition devices. For example, the larger the ratio between the size of user object in real-time dynamic data and the size of user object in the real world, the closer the distance between user object and image acquisition devices; the smaller the ratio between the size of user object in real-time dynamic data and the size of user object in the real world, the farther the distance between user object and image acquisition devices. Therefore, the interaction terminal is configured to determine the distance

between the user object and the image acquisition device based on the user size feature information of the user object in real-time dynamic data.

As the user position feature information identifies the position of the user object in real-time dynamic data, the user position feature information is typically defined using two-dimensional coordinates. That is, the interaction terminal is configured to obtain the information of the user object relative to the image acquisition device in the x-axis direction and the information of the user object relative to the image acquisition device in the y-axis direction from the user position feature information. The first relative position information identifies the position of the user object relative to the image acquisition device, which is typically defined using three-dimensional coordinates. Therefore, the corresponding first relative position information cannot be obtained solely based on the user position feature information.

After determining the distance between the user object and the image acquisition device, the interaction terminal is configured to obtain the information of the target device in the z-axis direction relative to the image acquisition device from the distance between the user object and the image acquisition device. Therefore, the interaction terminal is configured to combine the distance between the user object and the image acquisition device and the user position feature information to obtain the information of the user object relative to the image acquisition device in the x-axis direction, the information of the user object relative to the image acquisition device in the y-axis direction, and the information of the user object relative to the image acquisition device in the z-axis direction, which can determine the first relative position information of the user object relative to the image acquisition device.

By using the user size feature information of the user object in real-time dynamic data, the distance between the user object and the image acquisition device can be determined. Based on the distance between the user object and the image acquisition device and the user position feature information, the first relative position information used to identify the position of the user object relative to the image acquisition device can be accurately determined.

In one possible implementation, user position feature information can be determined as follows:

- performing facial recognition detection on the real-time dynamic data, and
- 5 determining the object tracking box comprising the facial recognition area of the user object in the real-time dynamic data; and

determining the user position feature information based on the position of the object tracking box in the real-time dynamic data.

10

Specifically, in order to determine the position of the user object in the real-time dynamic data, the interaction terminal is configured to perform facial recognition detection on the real-time dynamic data, thereby determining the object tracking box that comprises the facial recognition area of the user object in the real-time dynamic

15 data. The object tracking box refers to the tracking box that comprises the facial recognition area of the user object, and the tracking box 300 refers to the box defined for facial recognition detection, as shown in FIG 3. The tracking box 300 is typically a rectangle.

- 20 After determining the object tracking box, the interaction terminal is configured to determine the user position feature information used to identify the position of the user object using the real-time dynamic data based on the position of the object tracking box in the real-time dynamic data.

- 25 Through the above method, it is possible to accurately determine the position of the user object using real-time dynamic data.

In one possible implementation, user size feature information can be determined as follows:

30

- determining the first facial size information of the user object in the real-time dynamic data; and

- determining the user size feature information based on the first facial size information and the actual facial size information of the user object.

Specifically, in the practical operation of the embodiment, real-time dynamic data may not include all of the user object, for example, as shown in FIG 3, the real-time dynamic data only comprises the portion 305 above the shoulder of the user object. In order to achieve interaction between the user object and the virtual object, the real-time dynamic data will include the facial recognition area of the user object. Therefore, the interaction terminal is configured to accurately determine the corresponding user size feature information based on the facial recognition area of the user object.

The interaction terminal is configured to determine the first facial size information of the user object in real-time dynamic data, which is used to identify the size of the facial recognition area of the user object in real-time dynamic data.

The real facial size information is used to identify the size of the facial recognition area of the user object in the real scene. The interaction terminal is configured to determine the user size feature information based on the first facial size information and the real facial size information. For example, the interaction terminal is configured to determine the ratio between the first facial size information and the real facial size information as the user size feature information.

The first facial size information corresponding to the facial recognition area of the user object and the real facial size information can accurately determine the position of the user object using real-time dynamic data.

At step 203, determining the deviation position information of the user object relative to the virtual object based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object.

The second relative position information is used to identify the position of the image acquisition device relative to the virtual object. Similar to the first relative position information, the second relative position information can also be defined using three-dimensional coordinates. It should be noted that in the actual operation of the

application, although the position of the image acquisition device is usually fixed, the second relative position information is usually not fixed since the display position of the virtual object on the screen may not be fixed.

- 5 In the practical operation of the embodiment, the interactive terminal is configured to first acquire the three-dimensional spatial positions of both the image acquisition device and the virtual object separately, and then determine the second relative position information of the image acquisition device relative to the virtual object based on these positions. For example, when the three-dimensional spatial position of the
- 10 image acquisition device is (x_1, y_1, z_1) and the three-dimensional spatial position of the virtual object is (x_2, y_2, z_2) , the second relative position information of the image acquisition device relative to the virtual object can be $(x_1-x_2, y_1-y_2, z_1-z_2)$.

- Deviation position information is used to identify the angle of the user object relative
- 15 to the virtual object. Deviation position information not only reflects the degree of deviation of the user object from the virtual object, but also reflects the direction of deviation of the user object from the virtual object. For example, the deviation position information of user object A from the virtual object can be 45 degrees to the right, or the deviation position information of user object B from the virtual object can be 45
- 20 degrees to the left.

- After the interaction terminal is configured to determine the first relative position information in step 202, as the first relative position information is used to identify the position of the user object relative to the image acquisition device, combined with the
- 25 second relative position information of the image acquisition device relative to the virtual object, the interaction terminal is configured to obtain the position relationship between the user object and the image acquisition device, as well as the position relationship between the image acquisition device and the virtual object, through the first relative position information and the second relative position information.
- 30 Therefore, the deviation position information between the user object and the virtual object can be determined. For example, when the first relative position information is the three-dimensional coordinates (x_1, y_1, z_1) of the user object relative to the image acquisition device, and the second relative position information is the three-dimensional coordinates (x_2, y_2, z_2) of the image acquisition device relative to the

virtual object, the three-dimensional coordinates of the user object relative to the virtual object can be determined as $(x_1+x_2, y_1+y_2, z_1+z_2)$. Based on the three-dimensional coordinates of the user object relative to the virtual object $(x_1+x_2, y_1+y_2, z_1+z_2)$, the deviation position information between the user object and the virtual object can be determined.

At step 204, determining the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation position information.

10 A head posture target parameter is used to identify the position and direction of the virtual object's head in the corresponding virtual scene. For example, the head posture target parameter can include a head position vector used to identify the head position and a head direction vector used to identify the head direction.

15 An eye posture target parameter is used to identify the position of the virtual object's eyeball in the corresponding virtual scene and the corresponding line of sight. For example, the eye posture target parameter can include an eye position vector used to identify the eye position, and a line of sight direction vector used to identify the line of sight of the eyeballs.

20 The first posture rotation coefficient is used to identify the change values for the head posture target parameter. For example, when the head posture target parameter include the head position vector and the head direction vector, the first posture rotation coefficient can include the head displacement and the head rotation coefficient.

25 The second posture rotation coefficient is used to identify the change values for the eye posture target parameter. For example, when the eye posture target parameter include the eye position vector and the line of sight direction vector, the second posture rotation coefficient can include the eye displacement and the eye rotation coefficient.

30 Due to the fact that the deviation position information is used to identify the angle of the user object relative to the virtual object, that is, the deviation position information can reflect the degree and direction of deviation of the user object relative to the virtual object, in order to make the head of the virtual object face the user object and the

eyeballs of the virtual object can look directly at the user object, the interaction terminal is configured to determine the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation position information.

- 5 It should be noted that since the first posture rotation coefficient and the second posture rotation coefficient determined by the interaction terminal both refer to the deviation position information, this will cause the determined first posture rotation coefficient to match the angle between the user object and the virtual object. At the same time, the determined second posture rotation coefficient will also match the
10 angle between the user object and the virtual object.

In one possible implementation, the deviation position information comprises the horizontal deviation relative angle and pitch deviation relative angle of the user object relative to the virtual object. In step 204, determining the first posture rotation
15 coefficient and the second posture rotation coefficient of the virtual object based on the deviation position information, comprises:

- determining the first posture rotation coefficient of the virtual object based on the horizontal deviation relative angle; and
20
- determining the second posture rotation coefficient of the virtual object based on the pitch deviation relative angle.

Specifically, the deviation position information can include the horizontal deviation
25 relative angle and pitch deviation relative angle of the user object relative to the virtual object. The horizontal deviation relative angle of the user object relative to the virtual object refers to the angle of the user object relative to the virtual object in the horizontal plane, and the pitch deviation relative angle of the user object relative to the virtual object refers to the angle of the user object relative to the virtual object in the vertical
30 plane.

In the embodiment, when the deviation position information comprises the horizontal deviation relative angle and pitch deviation relative angle of the user object relative to the virtual object, although the first posture rotation coefficient of the virtual object can

be accurately determined based on both the horizontal deviation relative angle and pitch deviation relative angle, the computational cost will be relatively large. In order to reduce the computational cost, the interaction terminal is configured to determine the first posture rotation coefficient of the virtual object based on the horizontal deviation relative angle, so that the first posture rotation coefficient matches the horizontal deviation relative angle. In such case, the first posture rotation coefficient determined by the interaction terminal (such as the head rotation coefficient) can be the same as the horizontal deviation relative angle. For example, when the horizontal deviation relative angle is 10 degrees to the right, the corresponding first posture rotation coefficient can be 10 degrees to the right.

Similarly, when the deviation position information comprises the horizontal deviation relative angle of the user object relative to the virtual object and the pitch deviation relative angle, although the second posture rotation coefficient of the virtual object can be accurately determined based on both the horizontal deviation relative angle and the pitch deviation relative angle, the computational cost will be large. To reduce the computational cost, the interaction terminal is configured to determine the second posture rotation coefficient of the virtual object based on the pitch deviation relative angle, so that the second posture rotation coefficient matches the pitch deviation relative angle. In such case, the second posture rotation coefficient determined by the interaction terminal (such as the eye rotation coefficient) can be the same as the pitch deviation relative angle. For example, when the pitch deviation relative angle is 10 degrees downward, the corresponding second posture rotation coefficient can be 10 degrees downward.

Through the above method, the first posture rotation coefficient matching the horizontal deviation relative angle and the second posture rotation coefficient matching the pitch deviation relative angle can be quickly obtained, which reduces the corresponding calculation cost.

At step 205, adjusting the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

After determining the first posture rotation coefficient and the second posture rotation coefficient of the virtual object in step 204, as the determined first posture rotation coefficient and second posture rotation coefficient match the angles between the user object and the virtual object respectively, the interaction terminal is configured to adjust the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient. Specifically, the interaction terminal is configured to adjust the head target pose of the virtual object based on the first posture rotation coefficient and adjust the eye target pose of the virtual object based on the second posture rotation coefficient. For example, the interaction terminal is configured to rotate the head of the virtual object based on the first posture rotation coefficient and the eyeball of the virtual object based on the second posture rotation coefficient, as shown in FIG 4, when the user object is located at the lower right corner of the virtual object, the head of the virtual object can be rotated to the right according to the corresponding first posture rotation coefficient, and the eyeball of the virtual object can be turned downwards according to the second posture rotation coefficient, as shown in FIG 5. When the user object is located at the lower left corner of the virtual object, the head of the virtual object can be turned to the left according to the corresponding first posture rotation coefficient, and the eyeball of the virtual object can be turned downwards according to the second posture rotation coefficient, so that the head of the virtual object can face the user object and the eyeballs of the virtual object can look directly at the user object, thereby improving the realism of the virtual object and enhancing the interaction experience between the user and the virtual object.

In one possible implementation, the interaction method may also include:

- determining the body posture rotation coefficient of the virtual object based on the first posture rotation coefficient; and
- adjusting the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient in step 205, comprising:
 - adjusting the target pose of the virtual object based on the body posture rotation coefficient, the first posture rotation coefficient, and the second posture rotation coefficient.

Specifically, the body posture target parameter is used to identify the position and direction of the virtual object's body in the corresponding virtual scene. For example, the body posture target parameter may include a body position vector for identifying
5 body position and a body direction vector for identifying body direction.

The body posture rotation coefficient is used to identify the change values for the body posture target parameter. For example, when the body posture target parameter include body position vector and body direction vector, the body posture rotation
10 coefficient can include body displacement and body rotation angle.

The interaction terminal is configured to determine the body posture rotation coefficient of the virtual object based on the first posture rotation coefficient, so that the body posture rotation coefficient can match the first posture rotation coefficient. That is, the
15 interaction terminal is configured to determine the same body posture rotation coefficient (such as body rotation angle) based on the first posture rotation coefficient (such as head rotation angle). For example, when the head rotation angle is offset to the left by 10 degrees, the corresponding body rotation angle can also be offset to the left by 10 degrees.

After determining the body posture rotation coefficient, the interaction terminal is configured to adjust the target pose of the virtual object based on the body posture rotation coefficient, first posture rotation coefficient, and second posture rotation coefficient. Specifically, the interaction terminal is configured to adjust the head target
25 pose of the virtual object based on the first posture rotation coefficient, adjust the eye target pose of the virtual object based on the second posture rotation coefficient, and adjust the body target pose of the virtual object based on the body posture rotation coefficient. In one example, the interaction terminal is configured to rotate the head of the virtual object based on the first posture rotation coefficient, rotate the eyeball of
30 the virtual object based on the second posture rotation coefficient, and rotate the body of the virtual object based on the body rotation angle. For example, when the user object is located in the lower right corner of the virtual object, the interaction terminal is configured to rotate the virtual object's head to the right based on the corresponding first posture rotation coefficient, make the virtual object's eyeball downward based on

the corresponding second posture rotation coefficient, and rotate the virtual object's body to the right based on the corresponding body posture rotation coefficient.

By adjusting the target pose of the virtual object through the body posture rotation coefficient, first posture rotation coefficient, and second posture rotation coefficient, the virtual object's body can follow the head to face the user object, and the virtual object's eyes can face the user object directly. This further improves engagement with the virtual object.

In summary, the application provides a user interaction method applied to a virtual object, which is applied to an interaction system with an image acquisition device and a screen. The screen is configured to display the virtual object in the interaction interface, and the image acquisition device is configured to collect real-time dynamic data within the first preset area in real time. By obtaining real-time dynamic data comprising user object collected in real-time by the image acquisition device; based on the user position feature information and user size feature information of the user object in real-time dynamic data, the first relative position information of the user object relative to the image acquisition device can be determined; based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object, the deviation position information of the user object relative to the virtual object can be determined; based on the deviation position information, the first posture rotation coefficient and the second posture rotation coefficient of the virtual object are determined, and the target pose of the virtual object is adjusted, so that the head of the virtual object can face the user object and the eyeballs of the virtual object can look directly at the user object, improving engagement with the virtual object and enhancing the interaction experience between the user and the virtual object.

In order to save resources, in the actual application of the application, the virtual object can be in a standby state. In one possible implementation, the interaction method may also include:

- performing facial recognition detection on the video frame data collected by the image acquisition device;

- if the video frame data comprises human face, the user corresponding to the face is determined as the user object, the video frame data comprising the user object is determined as the real-time dynamic data, and the virtual object is activated.

5

Specifically, in order to save resources, the virtual object of the application can be in standby mode in practical operation. In such case, the interaction terminal is configured to perform facial recognition detection on the video frame data collected by the image acquisition device.

10

If the video frame data comprises a face, it indicates that there may be interaction between user and virtual object in the first preset area. Therefore, the user corresponding to the face can be determined as the user object, and the video frame data comprising the user object can be determined as real-time dynamic data, and the virtual object can be activated to interact with the user object in a timely manner.

15

It should be noted that in order to avoid resource waste, if the video frame data does not include a face, it indicates that there is no interaction between the user and the virtual object in the first preset area, so the virtual object can be kept in standby mode.

20

By performing facial recognition detection on the video frame data obtained by the image acquisition device, the virtual object can be activated in a timely manner when there is interaction between the user and the virtual object in the first preset area.

25 In one possible implementation, when the video frame data comprises a human face, the virtual object is activated, comprising:

- determining the distance between the user object and the virtual object based on the user size feature information of the user object in the real-time dynamic data;

30

- if the distance between the user object and the virtual object is within a preset range, activating the virtual object.

Specifically, when the video frame data comprises a face, the interaction terminal can first determine the distance between the user object and the virtual object based on the user size feature information of the user object in the real-time dynamic data.

- 5 If the distance between the user object and the virtual object is within the preset range, the preset range refers to the preset interaction range located in the first preset area, indicating that the user object is close to the virtual object, and then the virtual object is activated.
- 10 If the distance between the user object and the virtual object is not within the preset range, it indicates that the user object is far from the virtual object. In such case, the interaction terminal is configured to withhold the virtual object.

15 In the practical operation of the embodiment, the preset range can be set by relevant personnel based on experience. The preset range is usually smaller than the first preset area. For example, when the first preset area is a square area with a side length of 3 meters in front of the screen, the preset range can be a range where the distance between the user object and the virtual object is no more than 2 meters. Alternatively, when the first preset area is a circular area with a radius of 3 meters in front of the screen, the preset range can also be a range where the distance between the user object and the virtual object is no more than 2 meters.

20

By measuring the distance between the user object and the virtual object, the virtual object can be restarted when it is closer to the user object, avoiding meaningless wake-up of the virtual object when it is further away from the user object.

25

In one possible implementation, the interaction system may also include an audio acquisition device and an audio playback device, and the interaction method may also comprise:

30

- collecting the first interactive voice of the user object towards the virtual object through the audio acquisition device;

- performing language recognition on the first interactive voice to determine the language of the first interactive voice;

5 - generating the second interactive voice corresponding to the first interactive voice based on the language of the first interactive voice and the semantics of the first interactive voice; the second interactive voice belongs to the same language as the first interactive voice;

- playing the second interactive voice through the audio playback device.

10

Specifically, the interaction terminal can also include an audio acquisition device and an audio playback device. The audio acquisition device is used to collect user voice targeting virtual object, and the audio playback device is used to play virtual object voice targeting user object.

15

The first interactive voice refers to the voice generated when a user object interacts with a virtual object. For example, the first interactive voice of user object A for the virtual object can be "Hello".

20 The interaction terminal is configured to collect the first interactive voice of the user object towards the virtual object through the audio acquisition device.

After obtaining the first interactive voice, the interaction terminal is configured to recognize the language of the first interactive voice and determine the language of the first interactive voice. For example, when the first interactive voice is "ni hao", the interaction terminal is configured to determine the corresponding language as Chinese, and when the first interactive voice is "Hello", the interaction terminal is configured to determine the corresponding language as English.

30 After determining the language of the first interactive voice at the interaction terminal, the second interactive voice corresponding to the first interactive voice can be generated based on the language of the first interactive voice and the semantics of the first interactive voice. In order to ensure the interaction experience of the user object, the second interactive voice and the first interactive voice belong to the same

language. For example, when the first interactive voice is "ni hao", the generated second interactive voice can be "ni hao", and when the first interactive voice is "Hello", the generated second interactive voice can be "Hello".

- 5 After the interaction terminal is configured to generate the second interactive voice, the second interactive voice can be played through an audio playback device.

After collecting the first interactive voice, the language of the first interactive voice can be determined first, so as to generate the second interactive voice that belongs to the same language as the first interactive voice, thereby helping the user object understand the semantics of the second interactive voice and further ensuring the user's interactive experience.

10

In one possible implementation, the user interaction method may also include:

15

- based on the real-time dynamic data, performing gender detection on the user object to determine the gender feature information of the user object;

- generating the second interactive voice corresponding to the first interactive voice based on the language of the first interactive voice and the semantics of the first interactive voice, comprising:

20

- generating corresponding second interactive voice based on the gender feature information of the user object, the semantics of the first interactive voice, and the language of the first interactive voice.

25

Specifically, the interaction terminal is configured to perform gender detection on user object based on real-time dynamic data, and configured to determine the gender feature information of user object. In some practical operations of the embodiment, gender detection can be performed on user object through classification models. For example, images comprising male and female users can be used as training samples first, and then the initial classification model can be trained through the training samples to obtain a classification model that can be used for gender detection.

30

After determining the gender feature information of the user object, in order to further improve the interaction experience of the user object, the interaction terminal can be configured to generate the corresponding second interaction voice based on the gender feature information of the user object, the semantics of the first interaction voice, and the language of the first interaction voice. For example, when the user object is male and the first interaction voice is "hello", the generated second interaction voice can be "hello, sir". When the user object is female and the first interaction voice is "hello", the generated second interaction voice can be "hello, Miss".

- By determining the gender feature information of the user object in advance, the second interactive voice targeting gender feature information can be generated to further ensure the user's interaction experience.

In one possible implementation, the interaction method also comprises:

15

- obtaining user information collection; the user information collection comprises basic information corresponding to multiple identified users, and the identified users are users registered in advance for the interaction system;

20

- based on the real-time dynamic data, identifying the user object through the user information collection, and determine the basic information of the user object;

- generating the second interactive voice corresponding to the first interactive voice based on the language of the first interactive voice and the semantics of the first interactive voice, comprising:

25

- generating the corresponding second interactive voice based on the basic information of the user object, the semantics of the first interactive voice, and the language of the first interactive voice.

30

Specifically, the interaction terminal is configured to obtain user information collection, which comprises basic information corresponding to multiple identified users. The identified users are users registered in advance for the interaction system, and the

basic information can include the name, age, and other information provided by the identified users during registration.

5 The interaction terminal is configured to identify user object based on real-time dynamic data and determine their basic information through user information collection. User's basic information refers to the basic information corresponding to the user object. It should be noted that since the user information collection only comprises basic information of identified users, the interaction terminal cannot obtain the basic information corresponding to the user object when the user object is not an identified
10 user.

After determining the basic information of the user object, in order to further improve the interaction experience of the user object, the interaction terminal is configured to generate the corresponding second interaction voice based on the user object's basic
15 information, the semantics of the first interaction voice, and the language of the first interaction voice. For example, when the name of user object A is "Mike" and the first interaction voice is "Hello", the generated second interaction voice can be "Hello, Mike".

By determining the basic information of the user object in advance, the second
20 interactive voice can be generated based on the user's basic information, thereby further enhancing the user's interaction experience.

In the description of the specification, the reference to terms such as "some possible embodiments", "some embodiments", "examples", "specific examples", or "some
25 examples" means that the specific features, structures, materials, or features described in combination with the embodiments or examples are included in at least one embodiment or example of the application, and the above terms may not necessarily represent the same embodiment or example. Moreover, the specific features, structures, materials, or features described can be combined in an
30 appropriate manner in any one or more embodiments or examples. In addition, those skilled in the art may combine and mix different embodiments or examples described in this specification, as well as features of different embodiments or examples, without contradicting each other.

- The method flowchart of the application describes certain operations as different steps executed in a certain order. The flowchart is explanatory rather than restrictive. Some steps described in the description can be grouped together and executed in a single operation, or some steps can be divided into multiple sub steps and executed in a different order from that shown in the description. The various steps shown in the flowchart can be implemented in any manner by any circuit structure and/or tangible mechanism, such as software running on a computer device, hardware (e.g., logic functions implemented by a processor or chip), etc., and/or any combination thereof.
- Those skilled in the art can understand that in the specific implementation methods described, the order of each step shown does not imply a strict execution order. The specific execution order of each step should be determined by its function and possible internal logic.
- Based on the aforementioned FIGs 1 to 5, the interaction system provided in the application will be explained through a system embodiment below. As shown in FIG 6, the interaction system 600 may include a processor 601, an image acquisition device 602 connected to the processor, and a screen 603.
- The screen 603 is configured to display virtual object in an interactive interface;
- the image acquisition device 602 is configured to collect real-time dynamic data within the first preset area in real time;
- the processor 601 is configured to obtain real-time dynamic data comprising user object collected in real-time by the image acquisition device; the user object is a user who interacts with the virtual object within the first preset area; determine the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data; determine the deviation position information of the user object relative to the virtual object based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object; determine the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation

position information; adjust the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

In one possible implementation, processor 601 is configured to:

5

- determine the distance of the user object relative to the image acquisition device based on the user size feature information of the user object in the real-time dynamic data;

10

- determine the first relative position information of the user object relative to the image acquisition device based on the distance of the user object relative to the image acquisition device and the user position feature information of the user object in the real-time dynamic data.

15

In one possible implementation, processor 601 is configured to:

- when the deviation position information comprises the horizontal deviation relative angle and pitch deviation relative angle of the user object relative to the virtual object, determine the first posture rotation coefficient of the virtual object based on the horizontal deviation relative angle; determine the second posture rotation coefficient of the virtual object based on the pitch deviation relative angle.

20

In one possible implementation, processor 601 is also configured to:

25

- perform facial recognition detection on the real-time dynamic data, and determine the object tracking box comprising the facial recognition area of the user object in the real-time dynamic data;

30

- determine the user position feature information based on the position of the object tracking box in the real-time dynamic data.

In one possible implementation, processor 601 is also configured to:

- determine the first facial size information of the user object in the real-time dynamic data;

5 - determine the user size feature information based on the first facial size information and the actual facial size information of the user object.

In one possible implementation, processor 601 is also configured to:

10 - determine the body posture rotation coefficient of the virtual object based on the first posture rotation coefficient;

15 - adjust the target pose of the virtual object based on the body posture rotation coefficient, the first posture rotation coefficient, and the second posture rotation coefficient.

In one possible implementation, processor 601 is also configured to:

20 - perform facial recognition detection on the video frame data collected by the image acquisition device;

if the video frame data comprises human face, the user corresponding to the face is determined as the user object, the video frame data comprising the user object is determined as the real-time dynamic data, and the virtual object is activated.

25

In one possible implementation, processor 601 is configured to:

30 - when the video frame data comprises human face, determine the distance between the user object and the virtual object based on the user size feature information of the user object in the real-time dynamic data;

- if the distance between the user object and the virtual object is within a preset range, activate the virtual object.

In one possible implementation, the interactive system further comprises an audio acquisition device and an audio playback device, and the processor 601 is further used for:

- 5 - collecting the first interactive voice of the user object towards the virtual object through the audio acquisition device;
- performing language recognition on the first interactive voice to determine the language of the first interactive voice;
- 10 - generating the second interactive voice corresponding to the first interactive voice based on the language of the first interactive voice and the semantics of the first interactive voice; the second interactive voice belongs to the same language as the first interactive voice;
- 15 - playing the second interactive voice through the audio playback device.

In one possible implementation, processor 601 is also configured to:

- 20 - based on the real-time dynamic data, perform gender detection on the user object to determine the gender feature information of the user object;
- generate corresponding second interactive voice based on the gender feature information of the user object, the semantics of the first interactive voice, and the language of the first interactive voice.
- 25

In one possible implementation, processor 601 is also configured to:

- 30 - obtain user information collection; the user information collection comprises basic information corresponding to multiple identified users, and the identified users are users registered in advance for the interaction system;

- based on the real-time dynamic data, identify the user object through the user information collection, and determine the basic information of the user object;

- 5 - generate the corresponding second interactive voice based on the basic information of the user object, the semantics of the first interactive voice, and the language of the first interactive voice.

10 It should be noted that the device in the implementation embodiment of the application can achieve the same effect and function by implementing various processes of the aforementioned method, and will not be repeated here.

15 In summary, the application provides a virtual object based interaction system, which comprises a processor, an image acquisition device connected to the processor, and a screen. The screen is configured to display the virtual object in the interaction interface, and the image acquisition device is configured to collect real-time dynamic data within the first preset area in real time. By using the interactive system, obtaining real-time dynamic data comprising user object collected in real-time by the image acquisition device; based on the user position feature information and user size feature information of the user object in real-time dynamic data, the first relative position information of the user object relative to the image acquisition device can be determined; based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object, the deviation position information of the user object relative to the virtual object can be determined; based on the deviation position information, the first posture rotation coefficient and the second posture rotation coefficient of the virtual object are determined, and the target pose of the virtual object is adjusted, so that the head of the virtual object can face the user object and the eyeballs of the virtual object can look directly at the user object, improving engagement with the virtual object and enhancing the interaction experience between the user and the virtual object.

20

25

30

The embodiment of the application also provides an electronic device comprising a processor, a memory, and a bus, wherein the memory stores machine-readable instructions executable by the processor. When the electronic device is operating, the

processor communicates with the memory via the bus, the machine-readable instructions are executed by the processor, the following processing is performed:

5 obtaining real-time dynamic data comprising user object collected by the image acquisition device, wherein the user object is a user who interacts with the virtual object within the first preset area;

10 determining the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data;

15 determining the deviation position information of the user object relative to the virtual object based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object;

determining the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation position information;

20 adjusting the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

25 The embodiment of the application also provides a computer-readable storage medium, the computer-readable storage medium stores a computer program, and when the computer program is run by a processor, it executes the user interaction method applied to virtual object described above. Wherein, the computer-readable storage medium can be a volatile or non-volatile computer-readable storage medium.

30 The embodiment of the application also provides a computer program product, comprising a computer program, which carries program code. The program code comprises instructions that can be used to execute the steps of the user interaction method applied to a virtual object described above. Details could refer to the above method embodiment, which will not be repeated here.

Wherein, the aforementioned computer program product can be specifically implemented through hardware, software, or a combination of them. In one optional embodiment, the computer program product is specifically embodied as a computer storage medium, while in another optional embodiment, the computer program product is specifically embodied as a software product, such as a Software Development Kit (SDK), etc.

The embodiments in the application are described in a progressive manner, and the same or similar parts between the embodiments can be referred to each other. Each embodiment focuses on the differences from other embodiment. In particular, for the embodiment of devices, equipment, and computer-readable storage medium, since they are basically similar to the method embodiment, their descriptions are simplified, and the relevant parts can be referred to the descriptions of the method embodiment.

The devices, equipment, and computer-readable storage medium provided in the embodiment of the application correspond one-to-one with the method. Therefore, the devices, equipment, and computer-readable storage medium also have beneficial technical effect similar to the corresponding method. Since the beneficial technical effect of the method have been described in detail above, the beneficial technical effect of the device, equipment, and computer-readable storage medium will not be repeated here.

Those skilled in the art will understand that the embodiment of the application can be implemented as method and device (equipment or system), or computer-readable storage medium. Therefore, the application can be implemented in a fully hardware implementation, a fully software implementation, or a combination of software and hardware. Furthermore, the application can be implemented in the form of a computer-readable storage medium containing one or more computer-usable program code (including but not limited to disk storage, CD-ROM, optical storage, etc.).

The application is described with reference to flowcharts and/or block diagrams of method, device (equipment or system), and computer-readable storage medium according to the embodiment of the application. It should be understood that each flow and/or block in the flowcharts and/or block diagrams, as well as combinations of flows

and/or blocks in the flowcharts and/or block diagrams, can be implemented by computer program instructions. These computer program instructions can be provided to a general-purpose computer, a special-purpose computer, an embedded processor, or other programmable data processing equipment to generate a machine, such that the instructions executed by the computer or other programmable data processing equipment generate devices that implement the functions specified in one or more flows in the flowchart and/or one or more blocks in the block diagram.

These computer program instructions can also be stored in a computer-readable memory that can direct a computer or other programmable data processing equipment to operate in a specific manner, such that the instructions stored in the computer-readable memory generate a product including an instruction device, where the instruction device implements the functions specified in one or more flows in the flowchart and/or one or more blocks in the block diagram.

These computer program instructions can also be loaded onto a computer or other programmable data processing equipment, such that a series of operational steps are performed on the computer or other programmable equipment to generate computer-implemented processing, thereby providing steps for implementing the functions specified in one or more flows in the flowchart and/or one or more blocks in the block diagram.

In a typical configuration, a computing device includes one or more processors (CPU), input/output interfaces, network interfaces, and memory.

Memory may include non-permanent memory, random access memory (RAM), and/or non-volatile memory in the form of Computer-readable medium, such as read-only memory (ROM) or Flash RAM. Memory is an example of a computer-readable medium.

Computer-readable medium include permanent and non-permanent, removable and non-removable media, which can be implemented by any method or technology for storing information. The information can be computer-readable instructions, data structures, program modules, or other data. Examples of computer-readable storage medium include, but are not limited to, phase change memory (PRAM), static random

access memory (SRAM), dynamic random access memory (DRAM), other types of random access memory, read-only memory, electrically erasable programmable read-only memory (EEPROM), flash memory, or other memory technologies, CD-ROM, digital versatile disc (DVD), or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage, or other magnetic storage devices, or any other non-transmission medium that can be used to store information accessible by a computing device. Additionally, although the operations of the method of the application are described in a specific order in the accompanying drawings, this does not require or imply that these operations must be performed in that specific order, or that all illustrated operations must be performed to achieve the desired result. Furthermore, certain steps may be omitted, multiple steps may be combined into one step for execution, and/or one step may be decomposed into multiple sub-steps for execution.

Although the spirit and principles of the application have been described above with reference to several specific embodiments, it should be understood that the application is not limited to the disclosed specific embodiments, and the division of various aspects does not imply that the features in these aspects cannot be combined. the application is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

Through the description of the above implementation modes, technical personnel in the field can clearly understand that the application can be implemented by software plus necessary general hardware, and of course, it can also be implemented through special hardware, comprising application-specific integrated circuits, special CPUs, special memories, special components, etc. In general, any function completed by a computer program can be easily implemented by corresponding hardware, and the specific hardware structure used to implement the same function can also be diverse, such as analog circuits, digital circuits, or special circuits. However, for the application, software program implementation is a better implementation mode in more cases. Based on this understanding, the technical solution of the application, or the part that contributes to the prior art, can be embodied in the form of a software product, which is stored in a readable storage medium, such as a computer's floppy disk, U disk, mobile hard disk, ROM, RAM, magnetic disk, or optical disk, etc., and comprises

several instructions for enabling a computer device to execute the method described in the embodiment of the application.

In the above embodiment, all or part of the implementation can be achieved through software, hardware, firmware, or any combination of them. When implemented using software, it can be fully or partially implemented in the form of computer program products.

The computer program product comprises one or more computer instructions. When the computer program instructions are loaded and executed on a computer, all or part of the processes or functions described in the embodiment of the application are generated. The computer can be a general-purpose computer, a special-purpose computer, a computer network, or other programmable devices. The computer instructions can be stored in a computer-readable storage medium or transmitted from one computer-readable storage medium to another. The available medium can be a magnetic medium (e.g., a floppy disk, a hard disk, or a magnetic tape), an optical medium (e.g., a DVD), or a semiconductor medium (e.g., a Solid State Disk (SSD)), etc.

It should be understood that the "an embodiment" or "an embodiment" mentioned throughout the specification means that specific features, structures, or characteristics related to the embodiment are included in at least one embodiment of the application. Therefore, the appearance of "in an embodiment" or "in an embodiment" in various places throughout the specification does not necessarily refer to the same embodiment. Furthermore, these specific features, structures, or characteristics can be combined in one or more embodiment in any suitable manner. It should be understood that in the various embodiment of the application, the size of the sequence numbers of the above processes does not necessarily imply the order of execution, and the execution order of the processes should be determined by their functions and internal logic rather than limiting the implementation process of the embodiment of the application.

In addition, the terms "system" and "network" are often used interchangeably in this article. The term "and/or" in this article is merely used to describe the association

relationship between associated objects and indicates that there can be three relationships, for example, A and/or B can indicate: A exists alone, A and B exist at the same time, and B exists alone. In addition, the character "/" in this article generally indicates an "or" relationship between the associated objects before and after it.

5

Ordinary technical personnel in this field can realize that the modules and algorithm steps of the examples described in conjunction with the embodiment disclosed in this article can be implemented through electronic hardware, computer software, or a combination of both. To clearly illustrate the interchangeability of hardware and software, the composition and steps of each example have been generally described in terms of their functions in the above description. Whether these functions are performed through hardware or software depends on specific applications and design constraints of the technical solutions. Professional technicians can use different methods for each specific application to implement the described functions, but such implementation should not be considered as the scope of the application.

In summary, what has been described above is only a preferred embodiment of the technical solution of the application, and is not used to limit the protection scope of the application. Any modifications, equivalent substitutions, improvements, etc., made within the spirit and principles of the application should be included within the protection scope of the application.

Throughout this specification and claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers or steps but not the exclusion of any other integer or group of integers.

Persons skilled in the art will appreciate that numerous variations and modifications will become apparent. All such variations and modifications which become apparent to persons skilled in the art, should be considered to fall within the spirit and scope that the invention broadly appearing before described.

CLAIMS

1. A user interaction method applied to virtual object, characterized in that the method
5 is applied to an interaction system comprising an image acquisition device and a
screen, wherein the screen is configured to display virtual object in the interaction
interface, the image acquisition device is configured to collect real-time dynamic data
within the first preset area in real time, and the method comprises:
obtaining real-time dynamic data comprising user object collected by the image
10 acquisition device, wherein the user object is a user who interacts with the virtual
object within the first preset area;
determining the first relative position information of the user object relative to the image
acquisition device based on the user position feature information and user size feature
information of the user object in the real-time dynamic data;
15 determining the deviation position information of the user object relative to the virtual
object based on the first relative position information and the second relative position
information of the image acquisition device relative to the virtual object;
determining the first posture rotation coefficient and the second posture rotation
coefficient of the virtual object based on the deviation position information; and
20 adjusting the target pose of the virtual object based on the first posture rotation
coefficient and the second posture rotation coefficient.
2. The method according to claim 1, characterized in that determine the first relative
position information of the user object relative to the image acquisition device based
25 on the user position feature information and user size feature information of the user
object in the real-time dynamic data, comprising:
determining the distance of the user object relative to the image acquisition device
based on the user size feature information of the user object in the real-time dynamic
data;
30 determining the first relative position information of the user object relative to the image
acquisition device based on the distance of the user object relative to the image
acquisition device and the user position feature information of the user object in the
real-time dynamic data.

3. The method according to claim 1, characterized in that the deviation position information comprises the horizontal deviation relative angle and pitch deviation relative angle of the user object relative to the virtual object; determining the first posture rotation coefficient and the second posture rotation coefficient of the virtual

5 object based on the deviation position information, comprising:

determining the first posture rotation coefficient of the virtual object based on the horizontal deviation relative angle;

determining the second posture rotation coefficient of the virtual object based on the pitch deviation relative angle.

10

4. The method according to claim 1, characterized in that the user position feature information is determined as follows:

performing facial recognition detection on the real-time dynamic data, and determine the object tracking box comprising the facial recognition area of the user object in the

15 real-time dynamic data;

determining the user position feature information based on the position of the object tracking box in the real-time dynamic data.

5. The method according to claim 1, characterized in that the user size feature information is determined as follows:

20 determining the first facial size information of the user object in the real-time dynamic data;

determining the user size feature information based on the first facial size information and the actual facial size information of the user object.

25

6. The method according to claim 1, characterized in that the method further comprises: determining the body posture rotation coefficient of the virtual object based on the first posture rotation coefficient;

adjusting the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient, comprising:

30 adjusting the target pose of the virtual object based on the body posture rotation coefficient, the first posture rotation coefficient, and the second posture rotation coefficient.

7. The method according to claim 1, characterized in that the method further comprises:
perform facial recognition detection on the video frame data collected by the image
acquisition device;

5 if the video frame data comprises human face, the user corresponding to the face is
determined as the user object, the video frame data comprising the user object is
determined as the real-time dynamic data, and the virtual object is activated.

8. The method according to claim 7, characterized in that when the video frame data
comprises human face, the virtual object is activated comprises:

10 determining the distance between the user object and the virtual object based on the
user size feature information of the user object in the real-time dynamic data;
if the distance between the user object and the virtual object is within a preset range,
activating the virtual object.

15 9. The method according to claim 1, characterized in that the interaction system further
comprises an audio acquisition device and an audio playback device, and the method
further comprises:

collecting the first interactive voice of the user object towards the virtual object through
the audio acquisition device;

20 performing language recognition on the first interactive voice to determine the
language of the first interactive voice;

generating the second interactive voice corresponding to the first interactive voice
based on the language of the first interactive voice and the semantics of the first
interactive voice; the second interactive voice belongs to the same language as the

25 first interactive voice;

playing the second interactive voice through the audio playback device.

10. The method according to claim 9, characterized in that the method further
comprises:

30 based on the real-time dynamic data, performing gender detection on the user object
to determine the gender feature information of the user object;

generating the second interactive voice corresponding to the first interactive voice
based on the language of the first interactive voice and the semantics of the first
interactive voice, comprising:

generating corresponding second interactive voice based on the gender feature information of the user object, the semantics of the first interactive voice, and the language of the first interactive voice.

- 5 11. The method according to claim 9, characterized in that the method further comprises:
- obtaining user information collection; the user information collection comprises basic information corresponding to multiple identified users, and the identified users are users registered in advance for the interaction system;
- 10 based on the real-time dynamic data, identifying the user object through the user information collection, and determine the basic information of the user object;
- generating the second interactive voice corresponding to the first interactive voice based on the language of the first interactive voice and the semantics of the first interactive voice, comprising:
- 15 generating the corresponding second interactive voice based on the basic information of the user object, the semantics of the first interactive voice, and the language of the first interactive voice.

12. An interactive system characterized in that comprises a processor, an image
- 20 acquisition device connected to the processor, and a screen:
- the screen being configured to display virtual object in an interactive interface;
- the image acquisition device being configured to collect real-time dynamic data within the first preset area in real time;
- the processor being configured to obtain real-time dynamic data comprising user
- 25 object collected in real-time by the image acquisition device; the user object is a user who interacts with the virtual object within the first preset area; determine the first relative position information of the user object relative to the image acquisition device based on the user position feature information and user size feature information of the user object in the real-time dynamic data; determine the deviation position information
- 30 of the user object relative to the virtual object based on the first relative position information and the second relative position information of the image acquisition device relative to the virtual object; determine the first posture rotation coefficient and the second posture rotation coefficient of the virtual object based on the deviation

position information; adjust the target pose of the virtual object based on the first posture rotation coefficient and the second posture rotation coefficient.

13. An electronic device, characterized in that comprising a processor, a memory, and a bus; wherein the memory is configured to store machine-readable instructions executable by the processor; when the electronic device is operating, the processor communicates with the memory via the bus; when the machine-readable instructions are executed by the processor, the user interaction method applied to virtual object which described in any one of claims 1 to 11 is performed.

14. A computer-readable storage medium, characterized in that the computer-readable storage medium stores a computer program, and when the computer program is run by a processor, it executes the user interaction method applied to virtual object described in any one of claims 1 to 11.

15. A computer program product, characterized in that comprises a computer program, and when the computer program is run by a processor, it executes the user interaction method applied to virtual object described in any one of claims 1 to 11.