

**THREE-DIMENSIONAL LIVING-BODY FACE DETECTION METHOD, FACE
AUTHENTICATION RECOGNITION METHOD, AND APPARATUSES**

Cross-Reference to Related Applications

- 5 **[0001]** This international application is based upon and claims priority to Chinese Patent Application No. 201810777429.X, filed on July 16, 2018, the entire content of all of which is incorporated herein by reference.

Technical Field

- 10 **[0002]** Embodiments of this specification relate to the field of computer technologies, and in particular, to a three-dimensional living-body face detection method, a face authentication recognition method, and apparatuses.

Technical Background

- 15 **[0003]** Currently popular face recognition and detection technologies have been used to improve the security of authentication.

- [0004]** In face recognition systems, the most common cheating manner is counterfeiting attacks, in which an imposter intrudes a face recognition system with a counterfeit feature of the same representation form. At present, common counterfeiting attacks mainly include photos,
20 videos, three-dimensional models, and so on.

- [0005]** Currently, living-body detection technologies are mainly used to defend against similar attacks, in which instructions are delivered to instruct completion of specific living-body actions such as blinking, turning the head, opening the mouth, or other physiological behaviors, thereby determining whether these living-body actions are completed by a living body. However,
25 these living-body detection methods cannot achieve desirable detection performance, which affects

the living-body detection results, thereby affecting the accuracy of authentication recognition.

Summary of the Invention

[0006] Embodiments of this specification provide a three-dimensional living-body face
5 detection method, a face authentication recognition method, and apparatuses, for solving the
problem of poor living-body detection performance in the prior art.

[0007] The embodiments of this specification adopt the following technical solutions to
solve the above technical problem.

[0008] In a first aspect, a three-dimensional living-body face detection method is provided,
10 comprising:

[0009] acquiring multiple frames of depth images for a target detection object;

[0010] pre-aligning the multiple frames of depth images to obtain pre-processed point
cloud data;

[0011] normalizing the point cloud data to obtain a grayscale depth image; and

15 [0012] performing living-body detection based on the grayscale depth image and a living-
body detection model.

[0013] In a second aspect, a face authentication recognition method is provided,
comprising:

[0014] acquiring multiple frames of depth images for a target detection object;

20 [0015] pre-aligning the multiple frames of depth images to obtain pre-processed point
cloud data;

[0016] normalizing the point cloud data to obtain a grayscale depth image;

[0017] performing living-body detection based on the grayscale depth image and a living-
body detection model; and

25 [0018] determining whether the authentication recognition succeeds according to the

living-body detection result.

[0019] In a third aspect, a three-dimensional face detection apparatus is provided, comprising:

[0020] an acquisition module configured to acquire multiple frames of depth images for a
5 target detection object;

[0021] a first pre-processing module configured to pre-align the multiple frames of depth images to obtain pre-processed point cloud data;

[0022] a normalization module configured to normalize the point cloud data to obtain a grayscale depth image; and

10 [0023] a detection module configured to perform living-body detection based on the grayscale depth image and a living-body detection model.

[0024] In a fourth aspect, a face authentication recognition apparatus is provided, comprising:

[0025] an acquisition module configured to acquire multiple frames of depth images for a
15 target detection object;

[0026] a first pre-processing module configured to pre-align the multiple frames of depth images to obtain pre-processed point cloud data;

[0027] a normalization module configured to normalize the point cloud data to obtain a grayscale depth image;

20 [0028] a detection module configured to perform living-body detection based on the grayscale depth image and a living-body detection model; and

[0029] a recognition module configured to determine whether the authentication recognition succeeds according to the living-body detection result.

[0030] In a fifth aspect, an electronic device is provided, the electronic device comprising
25 a memory, a processor, and a computer program stored in the memory and executable on the

processor, wherein the computer program is executed by the processor for:

- [0031] acquiring multiple frames of depth images for a target detection object;
- [0032] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;
- 5 [0033] normalizing the point cloud data to obtain a grayscale depth image; and
- [0034] performing living-body detection based on the grayscale depth image and a living-body detection model.

[0035] In a sixth aspect, an electronic device is provided, the electronic device comprising a memory, a processor, and a computer program stored in the memory and executable on the processor, wherein the computer program is executed by the processor for:

- [0036] acquiring multiple frames of depth images for a target detection object;
- [0037] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;
- [0038] normalizing the point cloud data to obtain a grayscale depth image;
- 15 [0039] performing living-body detection based on the grayscale depth image and a living-body detection model; and
- [0040] determining whether the authentication recognition succeeds according to the living-body detection result.

[0041] In a seventh aspect, a computer-readable storage medium storing one or more programs is provided, wherein when executed by an electronic device comprising multiple applications, the one or more programs enable the electronic device to perform the following operations:

- [0042] acquiring multiple frames of depth images for a target detection object;
- [0043] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;
- 25

[0044] normalizing the point cloud data to obtain a grayscale depth image; and

[0045] performing living-body detection based on the grayscale depth image and a living-body detection model.

[0046] In an eighth aspect, a computer-readable storage medium storing one or more
5 programs is provided, wherein when executed by a server comprising multiple applications, the one or more programs enable the server to perform the following operations:

[0047] acquiring multiple frames of depth images for a target detection object;

[0048] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;

10 [0049] normalizing the point cloud data to obtain a grayscale depth image;

[0050] performing living-body detection based on the grayscale depth image and a living-body detection model; and

[0051] determining whether the authentication recognition succeeds according to the living-body detection result.

15 [0052] At least one of the above technical solutions adopted in the embodiments of this specification can achieve the following beneficial effects.

[0053] With the above technical solution, multiple frames of depth images for a target detection object are acquired to ensure the overall performance of an image input as detection data; the multiple frames of depth images are pre-aligned and the point cloud data is normalized to obtain
20 a grayscale depth image, which can ensure the integrity and accuracy of the grayscale depth image and compensate for the image quality problem; and finally, the living-body detection is performed based on the grayscale depth image and a living-body detection model, thereby improving the accuracy of the living-body detection. Then, more effective security verification or attack defense can be implemented based on the detection results.

Brief Description of the Drawings

[0054] In order to illustrate the technical solutions in the embodiments of this specification or in the prior art more clearly, the accompanying drawings used in descriptions of the embodiments or the prior art will be briefly described below. It is apparent that the accompanying drawings in the following description are only some embodiments recorded in the embodiments of this specification, and other accompanying drawings can be obtained by those of ordinary skill in the art according to these accompanying drawings without creative efforts.

[0055] FIG. 1a is a first schematic diagram of steps of a three-dimensional living-body face detection method according to an embodiment of this specification;

10 [0056] FIG. 1b is a second schematic diagram of steps of a three-dimensional living-body face detection method according to an embodiment of this specification;

[0057] FIG. 2a is a first schematic diagram of steps of a living-body detection model generation method according to an embodiment of this specification;

[0058] FIG. 2b is a second schematic diagram of steps of a living-body detection model generation method according to an embodiment of this specification;

[0059] FIG. 3 is a schematic diagram of a human living-body face detection method according to an embodiment of this specification;

[0060] FIG. 4 is a schematic diagram of steps of a face authentication recognition method according to an embodiment of this specification;

20 [0061] FIG. 5 is a schematic diagram of a hardware structure of an electronic device according to an embodiment of this specification;

[0062] FIG. 6a is a first schematic structural diagram of a three-dimensional living-body face detection apparatus according to an embodiment of this specification;

[0063] FIG. 6b is a second schematic structural diagram of a three-dimensional living-body face detection apparatus according to an embodiment of this specification;

[0064] FIG. 6c is a third schematic structural diagram of a three-dimensional living-body face detection apparatus according to an embodiment of this specification;

[0065] FIG. 6d is a fourth schematic structural diagram of a three-dimensional living-body face detection apparatus according to an embodiment of this specification; and

5 [0066] FIG. 7 is a schematic structural diagram of a face authentication recognition apparatus according to an embodiment of this specification.

Detailed Description

[0067] In order to make objectives, technical solutions and advantages of the embodiments
10 of this specification clearer, the technical solutions of the embodiments of this specification will be described clearly and completely below with reference to specific embodiments of this specification and the corresponding accompanying drawings. It is apparent that the described embodiments are only a part of rather than all the embodiments of this specification. All other embodiments obtained by those of ordinary skill in the art based on the embodiments of this
15 specification without creative efforts fall within the protection scope of the embodiments of this specification.

[0068] The technical solutions provided by the embodiments of this specification are described in detail below with reference to the accompanying drawings.

[0069] Embodiment 1

20 [0070] Referring to FIG. 1a, a schematic diagram of steps of a three-dimensional living-body face detection method according to an embodiment of this specification is shown. The method may be executed by a three-dimensional living-body face detection apparatus or a mobile terminal installed with the three-dimensional living-body face detection apparatus.

[0071] The three-dimensional living-body face detection method may include the
25 following steps.

[0072] In step 102, multiple frames of depth images for a target detection object are acquired.

[0073] It should be understood that the three-dimensional living-body face detection involved in the embodiment of this specification is mainly three-dimensional living-body face
5 detection for a human. It is determined according to analysis on a three-dimensional human face image whether a target detection object is a living body, i.e., whether it is the person corresponding to the target detection object in the image. In fact, the target detection object of the three-dimensional living-body face detection is not limited to a human, but can be an animal having a recognizable face, which is not limited in the embodiment of this specification.

10 [0074] The living-body detection can determine whether a current operator is a living human or a non-human such as a picture, a video, a mask, or the like. The living-body detection can be applied to scenarios using face swiping verification such as clock in and out and face swiping payment.

[0075] The multiple frames of depth images described in the embodiment of this
15 specification refer to images acquired for a face region of the target detection object by means of photographing, infrared, or the like, and specifically depth images that can be acquired by a depth camera that measures a distance between an object (the target detection object) and the camera. The depth camera involved in the embodiment of this specification may include: a depth camera based on a structured light imaging technology, or a depth camera based on a light time-of-flight
20 imaging technology. Further, while the depth image is acquired, a color image for the target detection object, that is, an RGB image is also acquired. Since color images are generally acquired during image acquisition, it is set by default in this specification that the color image is also acquired while the depth image is acquired.

[0076] Considering that the depth camera based on the structured light imaging technology
25 is sensitive to illumination and cannot be used in an outdoor scene with strong light, an active

binocular depth camera is preferably used in the embodiment of this specification to acquire a depth image of the target detection object.

[0077] It should be understood that in the embodiment of this specification, the multiple frames of depth images may be acquired from a depth camera device (such as various types of depth cameras mentioned above) externally mounted on the three-dimensional living-body face detection apparatus, that is, these depth images are acquired by the depth camera and transmitted to the three-dimensional living-body face detection apparatus; or acquired from a depth camera device built in the three-dimensional living-body face detection apparatus, that is, the depth images are acquired by the three-dimensional living-body face detection apparatus through a built-in depth camera. This is not limited in this specification.

[0078] In step 104, the multiple frames of depth images are pre-aligned to obtain pre-processed point cloud data.

[0079] It should be understood that the depth images acquired in step 102 are mostly acquired based on depth cameras, and are generally incomplete, limited in accuracy, etc. Therefore, the depth images may be pre-processed before use.

[0080] In the embodiment of this specification, the multiple frames of depth images may be pre-aligned, thereby effectively compensating for the acquisition quality problem of the depth camera, having better robustness to subsequent three-dimensional living-body face detection, and improving the overall detection accuracy.

[0081] In step 106, the point cloud data is normalized to obtain a grayscale depth image.

[0082] In the embodiment of this specification, the pre-alignment of the depth images can be regarded as a feature extraction process. After the feature extraction and the pre-alignment, the point cloud data needs to be normalized to a grayscale depth image that can be used by the subsequent algorithm. Thus, the integrity and accuracy of the image are further improved.

[0083] In step 108, living-body detection is performed based on the grayscale depth image

and a living-body detection model.

[0084] It should be understood that in the embodiment of this specification, when living-body detection is performed on a target, depth images may vary for a living target detection object and a non-living target detection object. Taking the human living-body face detection as an example, if the target detection object is a face photo, a video, a three-dimensional model, or the like, instead of a living human face, a distinction is made at the time of detection. Based on this idea, it is determined whether the target detection object is a living body or a non-living body by detecting the acquired depth image of the target detection object in this specification.

[0085] With the above technical solution, multiple frames of depth images for a target detection object are acquired to ensure the overall performance of an image input as detection data; the multiple frames of depth images are pre-aligned and the point cloud data is normalized to obtain a grayscale depth image, which can ensure the integrity and accuracy of the grayscale depth image and compensate for the image quality problem; and finally, the living-body detection is performed based on the grayscale depth image and a living-body detection model, thereby improving the accuracy of the living-body detection. Then, more effective security verification or attack defense can be implemented based on the detection results.

[0086] The living-body detection model in the embodiment of this specification may be a preset normal living-body detection model. Referring to FIG. 2a, it may preferably be obtained based on the following methods.

[0087] In step 202, multiple frames of depth images for a target training object are acquired.

[0088] It should be understood that the multiple frames of depth images for the target training object in this step may be a historical depth image extracted from an existing depth image database or other storage spaces. Unlike the depth image in step 102, the type of the target training object (living body or non-living body) is known.

[0089] In step 204, the multiple frames of depth images are pre-aligned to obtain pre-

processed point cloud data.

[0090] The specific implementation of the step 204 may be obtained with reference to step 104.

[0091] In step 206, the point cloud data is normalized to obtain a grayscale depth image sample.

[0092] The point cloud data obtained after the pre-alignment based on the above step 204 is normalized to obtain a gray-scale depth image sample. As the sample, the depth image subjected to the pre-alignment and the normalization is mainly used as data of a known type that is input to a training model subsequently. The normalization here is the same as the implementation of step 106.

[0093] In step 208, training is performed based on the grayscale depth image sample and label data of the grayscale depth image sample to obtain the living-body detection model.

[0094] Label data of the grayscale depth image sample may be a type label of the target training object. In the embodiment of this specification, the type label may be simply set to be: living body or non-living body.

[0095] It should be understood that in the solution involved in the embodiment of this specification, a convolutional neural network (CNN) structure may be selected as a training model, and the CNN structure mainly includes a convolution layer and a pooling layer. A construction process thereof may include: convolution, activation, pooling, full connection, and the like. The CNN structure can perform binary training on the input image data and the label of the training object, thereby obtaining a classifier. For example, the grayscale depth image samples A1 (label data: living body), B1 (label data: living body), A2 (label data: non-living body), B2 (label data: living body), A3 (label data: living body), B3 (label data: non-living body), etc. after normalization are used as data input to the training model, i.e., the CNN structure. After that, the CNN structure performs model training according to the input data, and finally obtains a classifier, which can

accurately identify whether the target detection object corresponding to the input data is a living body and output the detection result.

[0096] It should be noted that in the actual model training process, the quantity of data (grayscale depth image samples) input to the training model can be enough to support the training model for effective training. Some of them are listed in the embodiment of this specification only for illustration.

[0097] In fact, the classifier mentioned above can be understood as a living-body detection model obtained by training. As there are only two types (living or non-living) of the labels (i.e., the label data) input during training, the classifier can be a binary classifier.

10 [0098] According to the living-body detection model obtained in the above FIG. 2a, the CNN model is trained based on the grayscale depth image sample after the pre-processing and the normalization used as the input data, and therefore, a more accurate living-body detection model can be obtained and, further, the living-body detection based on the living-body detection model is more accurate.

15 [0099] Optionally, in the embodiment of this specification, step 104 may specifically include:

[0100] roughly aligning the multiple frames of depth images based on three-dimensional key facial points; and

[0101] finely aligning the roughly aligned depth images based on an iterative closest point (ICP) algorithm to obtain the point cloud data.

[0102] It can be seen that step 104 mainly includes rough alignment and fine alignment, and the pre-alignment is briefly introduced below.

[0103] The multiple frames of depth images are roughly aligned based on three-dimensional key facial points. In specific implementation, an RGB image detection mode may be used to determine the face key points in the depth image, and then the determined face key points

are subjected to point cloud rough-alignment. The face key points can be five key points in the human face including the two corners of eyes, the tip of the nose, and the two corners of the mouth. With the point cloud rough-alignment, the multiple frames of depth images are only roughly registered to ensure that the depth image is substantially aligned.

5 **[0104]** The point cloud data is obtained by finely aligning the depth images after the rough alignment based on the ICP algorithm. In specific implementation, the depth images processed by the rough alignment may be used as the initialization of the ICP algorithm, and then the iterative process of the ICP algorithm is used to perform fine alignment. In the embodiment of this specification, in the process of the ICP algorithm selecting key points, random sample consensus
10 (RANSAC) point selection is performed with reference to position information of five key points of the human face including the two corners of eyes, the tip of the nose, and the two corners of the mouth. At the same time, the number of iterations is limited so that the iterations are not excessive, thereby ensuring the processing speed of the system.

[0105] Optionally, in the embodiment of this specification, as shown in FIG. 1b, before
15 performing step 104, the method further includes:

[0106] step 110: bilaterally filtering each frame of depth image in the multiple frames of depth images.

[0107] It should be understood that in the embodiment of this specification, the multiple frames of depth images are acquired, and each frame of depth image may have an image quality
20 problem. Therefore, each frame of depth image in the multiple frames of depth images may be bilaterally filtered, thereby improving the integrity of each frame of depth image.

[0108] Specifically, each frame of depth image can be bilaterally filtered with reference to the following formula:

$$g(i, j) = \frac{\sum_{k,l} f(k, l) \omega(i, j, k, l)}{\sum_{k,l} \omega(i, j, k, l)} \quad (1)$$

[0109] wherein $g(i, j)$ represents a depth value of a pixel (i, j) in the depth image after the bilateral filtering, $f(k, l)$ is a depth value of a pixel (k, l) in the depth image before the bilateral filtering, and $\omega(i, j, k, l)$ is a weight value of the bilateral filtering.

[0110] Further, the weight value $\omega(i, j, k, l)$ of the bilateral filtering can be calculated by
 5 the following formula:

$$\omega(i, j, k, l) = \exp \left(-\frac{(i-k)^2 + (j-l)^2}{2\sigma_d^2} - \frac{\|f_c(i, j) - f_c(k, l)\|^2}{2\sigma_r^2} \right) \quad (2)$$

[0111] wherein $f_c(i, j)$ represents a color value of a pixel (i, j) in the color image, $f_c(k, l)$ represents a color value of a pixel (k, l) in the color image, σ_d^2 is a filtering parameter corresponding to the depth image, and σ_r^2 is a filtering parameter corresponding to the color image.

10 [0112] Optionally, in step 106, when the point cloud data is normalized to obtain a grayscale depth image, the method may be specifically implemented as follows.

[0113] In step 1, an average depth of the face region is determined according to three-dimensional key facial points in the point cloud data.

[0114] Taking the three-dimensional face being a human face as an example, the average
 15 depth of the human face region is calculated by average weighting or the like according to the five key points of the human face.

[0115] In step 2, the face region is segmented, and a foreground and a background in the point cloud data are deleted.

[0116] Image segmentation is performed on the face region, for example, key points such
 20 as nose, mouth, and eyes are obtained by segmentation, and then the point cloud data corresponding to a foreground image and the point cloud data corresponding to a background image other than the human face in the point cloud data are deleted, thereby eliminating the interference of the

foreground image and the background image with the point cloud data.

[0117] In step 3, the point cloud data from which the foreground and background have been deleted is normalized to preset value ranges before and after the average depth that take the average depth as the reference to obtain a grayscale depth image.

5 [0118] The depth values of the face region having the interference from the foreground and the background excluded are normalized to preset value ranges before and after the average depth determined in step 1 that take the average depth as the reference, wherein the preset value ranges before and after the average depth that take the average depth as the reference refer to a depth range between the average depth and a front preset value and a depth range between the average depth
10 and a rear preset value. The front refers to the side of a human face that faces the depth camera, and the rear refers to the side of a human face that opposes the depth camera.

[0119] For example, if the average depth of the face region previously determined is D1 and the preset value is D2, the depth value range of the face region normalized is $[D1-D2, D1+D2]$. It should be understood that considering that the thickness of the contour of the human face is
15 limited and is substantially within a certain range, the preset value may be set to any value between 30 mm and 50 mm, preferably 40 mm.

[0120] It should be understood that in the embodiment of this specification, the normalization involved in the above step 106 can be applied to the normalization of the model training shown in FIG. 2a.

20 [0121] Optionally, referring to FIG. 2b, before step 208 is performed, the method further includes:

[0122] step 210: performing data augmentation on the grayscale depth image sample, wherein the data augmentation includes at least one of the following: a rotation operation, a shift operation, and a zoom operation.

25 [0123] It should be understood that by the above data augmentation, the quantity of the

grayscale depth image samples (living body, non-living body) can be increased, the robustness of model training can be improved, and the accuracy of living-body detection can be further improved.

[0124] Preferably, during the augmentation, the rotation, shift, and zoom operations may be respectively performed according to three-dimensional data information of the grayscale depth
5 image sample.

[0125] Optionally, in order to improve the robustness of model training and subsequent living-body detection, the living-body detection model is a model obtained by training based on a convolutional neural network structure.

[0126] The three-dimensional living-body face detection scheme involved in this
10 specification will be described in detail below through a specific example.

[0127] It should be noted that, in the three-dimensional living-body face detection scheme, the three-dimensional face is, for example, a human face, and the training model is, for example, a CNN model.

[0128] Referring to FIG. 3, a schematic diagram of training of a living-body detection
15 model and living-body face detection according to an embodiment of this specification is shown. Here,

[0129] a training phase may include historical depth image acquisition, historical depth image pre-processing, point cloud data normalization, data augmentation, and binary model training. A detection phase may include online depth image acquisition, online depth image pre-
20 processing, point cloud data normalization, detection of whether it is a living body based on a binary model, or the like. In fact, the specific training phase and the detection phase may include other processes, which are not completely shown in the embodiment of this specification.

[0130] It should be understood that the binary model in the embodiment of this specification is the living-body detection model shown in FIG. 1a. In fact, the operations of the
25 training phase and the detection phase may be performed by a mobile terminal having a depth

image acquisition function or another terminal device. In the following, for example, the operations are performed by a mobile terminal. Specifically, the process shown in FIG. 3 mainly includes the following.

[0131] (1) Historical depth image acquisition

5 **[0132]** The mobile terminal acquires historical depth images. Some of these historical depth images are acquired by a depth camera for a living human face, and some are acquired by the depth camera for a non-living (such as a picture and a video) human face image. The historical depth images may be acquired based on an active binocular depth camera and stored as historical depth images in a historical database. The mobile terminal triggers the acquisition of historical
10 depth images from the historical database when model training and/or living-body detection are/is required.

[0133] It should be understood that the historical depth images involved in the embodiment of this specification are the multiple frames of depth images for the target training object described in FIG. 2a. When a historical depth image is acquired, a label corresponding to the historical depth
15 image (i.e., the label data) is also acquired, and the label is used to indicate that a target training object corresponding to the historical depth image is a living body or a non-living body.

[0134] (2) Historical depth image pre-processing

[0135] After the completion of the historical depth image acquisition, a single-frame depth image in the historical depth images can be bilaterally filtered, then the multiple frames of depth
20 images after bilateral filtering are roughly aligned according to the human face key points, and finally the ICP algorithm is used to finely align the results after the rough alignment, thus implementing accurate registration of the point cloud data. Therefore, more complete and accurate training data can be obtained. The specific implementation of the operations such as bilateral filtering, rough alignment of the human face key points, and fine alignment by the ICP algorithm
25 can be obtained with reference to the related description of the foregoing embodiments, and details

are not described here.

[0136] (3) Point cloud data normalization

[0137] In order to obtain more accurate training data, the registered point cloud data can also be normalized into a grayscale depth image for subsequent use. Firstly, the human face key points and the depth image D are detected according to the human face RGB image, and the average depth df of the face region is calculated. The df can be a numerical value in mm. Secondly, image segmentation is performed on the face region to exclude the interference from the foreground and the background. For example, only all point clouds with depth values in the range of $df-40\text{mm}$ to $df+40\text{mm}$ are reserved as the point cloud $P\{(x,y,z) | df+40 > z > df-40\}$ of the human face. Finally, the depth values of the face region having the interference from the foreground and the background excluded are normalized to a range of 40 mm before and after the average depth (this can be a value range at this time).

[0138] (4) Data augmentation

[0139] Considering that the quantity of acquired historical depth images may be limited, the normalized grayscale depth image may be augmented to increase the quantity of input data required for model training. The augmentation may be specifically implemented as at least one of a rotation operation, a shift operation, and a zoom operation.

[0140] For example, assuming that the normalized grayscale depth images are $M1$, $M2$, and $M3$, the grayscale depth images after the rotation operation are $M1(x)$, $M2(x)$, and $M3(x)$, the grayscale depth images after the shift operation are $M1(p)$, $M2(p)$, and $M3(p)$, and the grayscale depth images after the zoom operation are $M1(s)$, $M2(s)$, and $M3(s)$. As such, the original three grayscale depth images are augmented into twelve grayscale depth images, thereby increasing the input data of living body and non-living body and improving the robustness of model training. At the same time, the detection performance of subsequent living-body detection can further be improved.

[0141] It should be understood that the number of the normalized grayscale depth images described above is only an example, and is not limited to three. The specific acquisition quantity may be set as required.

[0142] (5) Binary model training

5 [0143] In the model training, the depth images obtained in step (1) may be used as training data, or the depth images obtained by the pre-processing in step (2) may be used as training data, or the grayscale depth images obtained by the normalization in step (3) may be used as training data, or the grayscale depth images obtained by the augmentation in step (4) may be used as the training data.

10 [0144] Obviously, the living-body detection model trained by inputting the grayscale depth images obtained by the augmentation in step (4) as the training data to the CNN model is more accurate.

[0145] After the normalized grayscale depth images are processed by data augmentation, the CNN structure can be used to extract image features from the augmented grayscale depth
15 images, and then model training is performed based on the extracted image features and the CNN model.

[0146] In fact, during training, the training data also includes a label of the grayscale depth image, which may be labeled as "living body" or "non-living body" in the embodiment of this specification. As such, after the training is completed, a binary model that can output "living body"
20 or "non-living body" according to the input data can be obtained.

[0147] (6) Online depth image acquisition

[0148] Specific implementation of step (6) can be obtained with reference to the acquisition process in step (1).

[0149] (7) Online depth image pre-processing

25 [0150] Specific implementation of step (7) can be obtained with reference to the pre-

processing process of step (2).

[0151] (8) Point cloud data normalization

[0152] Specific implementation of step (8) can be obtained with reference to the normalization process of step (3).

5 [0153] (9) Detection of whether it is a living body based on the binary model

[0154] In the embodiment of this specification, the online depth images acquired in step (6) may be used as an input of the binary model, or the online depth images pre-processed in step (7) may be used as an input of the binary model, or the online grayscale depth images normalized in step (8) may be used as an input of the binary model to detect whether the target detection target is
10 a living body.

[0155] It should be understood that in the embodiment of this specification, the processing manner of inputting the data of the detection model in the detection phase may be the same as the processing manner of inputting the data of the training model in the training phase. For example, if the binary model is obtained by training based on the acquired historical depth images, the online
15 depth images acquired in step (6) are used as an input of the binary model for detection.

[0156] In the embodiment of this specification, in order to ensure the accuracy of the living-body detection, a binary model obtained by training based on the augmented grayscale depth images is preferably selected, the online grayscale depth image normalized in step (8) is selected as an input, and the binary model can output a detection result of "living body" or "non-living
20 body" based on the input data.

[0157] (10) Output the detection result to a living-body detection apparatus

[0158] The test result can be obtained based on the binary model.

[0159] At this time, the detection result can be fed back to a living-body detection system so that the living-body detection system performs a corresponding operation. For example, in a
25 payment scenario, if the detection result is "living body," the detection result is fed back to a

payment system, so that the payment system performs payment; if the detection result is "non-living body," the detection result is fed back to the payment system, so that the payment system refuses to perform the payment. Thus, the authentication security can be improved by a more accurate living-body detection method.

5 **[0160]** The specific embodiments of this specification have been described above. In some cases, the actions or steps recited in this specification can be performed in an order different from that in the embodiments and the desired results can still be achieved. In addition, the processes depicted in the accompanying drawings are not necessarily required to be in the shown particular order or successive order to achieve the expected results. In some implementation manners, 10 multitasking and parallel processing are also possible or may be advantageous.

[0161] Embodiment 2

[0162] Referring to FIG. 4, a schematic diagram of steps of a face authentication recognition method according to an embodiment of this specification is shown. The method may be performed by a face authentication recognition apparatus or a mobile terminal provided with a 15 face authentication recognition apparatus.

[0163] The face authentication recognition method may include the following steps.

[0164] In step 402, multiple frames of depth images for a target detection object are acquired.

[0165] Specific implementation of step 402 may be obtained with reference to step 102.

20 **[0166]** In step 404, the multiple frames of depth images are pre-aligned to obtain pre-processed point cloud data.

[0167] Specific implementation of step 404 may be obtained with reference to step 104.

[0168] In step 406, the point cloud data is normalized to obtain a grayscale depth image.

[0169] Specific implementation of step 406 may be obtained with reference to step 106.

25 **[0170]** In step 408, living-body detection is performed based on the grayscale depth image

and a living-body detection model.

[0171] Specific implementation of step 408 may be obtained with reference to step 108.

[0172] In step 410, it is determined whether the authentication recognition succeeds according to the living-body detection result.

5 [0173] In the embodiment of this specification, the detection result of step 408: living body or non-living body, may be transmitted to an authentication recognition system, so that the authentication recognition system determines whether the authentication succeeds. For example, if the detection result is a living body, the authentication succeeds; and if the detection result is a non-living body, the authentication fails.

10 [0174] With the above technical solution, multiple frames of depth images for a target detection object are acquired to ensure the overall performance of an image input as detection data; the multiple frames of depth images are pre-aligned and the point cloud data is normalized to obtain a grayscale depth image, which can ensure the integrity and accuracy of the grayscale depth image and compensate for the image quality problem; and finally, the living-body detection is performed
15 based on the grayscale depth image and a living-body detection model, thereby improving the accuracy of the living-body detection. Then, more effective security verification or attack defense can be implemented based on the detection results.

[0175] Embodiment 3

[0176] An electronic device according to an embodiment of this specification is introduced
20 in detail below with reference to FIG. 5. Referring to FIG. 5, at the hardware level, the electronic device includes a processor and optionally further includes an internal bus, a network interface, and a memory. The memory may include a memory such as a high-speed Random-Access Memory (RAM), or may further include a non-volatile memory such as at least one magnetic disk memory. Definitely, the electronic device may further include hardware required by other services.

25 [0177] The processor, the network interface, and the memory may be interconnected

through the internal bus, and the internal bus may be an Industry Standard Architecture (ISA) bus, a Peripheral Component Interconnect (PCI) bus, an Extended Industry Standard Architecture (EISA) bus, or the like. The bus may be classified into an address bus, a data bus, a control bus, and the like. For ease of representation, only one double-sided arrow is shown in FIG. 5, but it does
5 not mean that there is only one bus or one type of bus.

[0178] The memory is configured to store a program. Specifically, the program may include program codes including a computer operation instruction. The memory may include memory and a non-volatile memory and provides an instruction and data to the processor.

[0179] The processor reads, from the non-volatile memory, the corresponding computer
10 program into the memory and runs the computer program, thus forming a three-dimensional face detection apparatus at the logic level. The processor executes the program stored in the memory, and is specifically configured to perform the following operations:

[0180] acquiring multiple frames of depth images for a target detection object;

[0181] pre-aligning the multiple frames of depth images to obtain pre-processed point
15 cloud data;

[0182] normalizing the point cloud data to obtain a grayscale depth image; and

[0183] performing living-body detection based on the grayscale depth image and a living-body detection model.

[0184] Alternatively, the processor performs the following operations:

20 **[0185]** acquiring multiple frames of depth images for a target detection object;

[0186] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;

[0187] normalizing the point cloud data to obtain a grayscale depth image;

[0188] performing living-body detection based on the grayscale depth image and a living-body detection model; and
25

[0189] determining whether the authentication recognition succeeds according to the living-body detection result.

[0190] The three-dimensional living-body face detection method disclosed in the embodiments shown in FIG. 1a to FIG. 3 according to the embodiments of this specification or the face authentication recognition method disclosed in FIG. 4 can be applied to the processor or implemented by the processor. The processor may be an integrated circuit chip having a signal processing capability. In the process of implementation, various steps of the above methods may be completed by an integrated logic circuit of hardware in the processor or an instruction in the form of software. The processor may be a general-purpose processor, including a Central Processing Unit (CPU), a Network Processor (NP), etc.; or may be a Digital Signal Processor (DSP), an Application Specific Integrated Circuit (ASIC), a Field-Programmable Gate Array (FPGA) or another programmable logic device, discrete gate or transistor logic device, or discrete hardware component. The methods, steps, and logical block diagrams disclosed in the embodiments of this specification can be implemented or performed. The general-purpose processor may be a microprocessor, or the processor may be any conventional processor or the like. The steps of the method disclosed in the embodiments of this specification may be directly performed by a hardware decoding processor, or may be performed by a combination of hardware and software modules in the decoding processor. The software module can be located in a storage medium mature in the field, such as a random-access memory, a flash memory, a read-only memory, a programmable read-only memory or electrically erasable programmable memory, a register, and the like. The storage medium is located in the memory, and the processor reads the information in the memory and implements the steps of the above method in combination with its hardware.

[0191] The electronic device can also perform the methods of FIG. 1a to FIG. 3, implement the functions of the three-dimensional living-body face detection apparatus in the embodiments shown in FIG. 1a to FIG. 3, perform the method in FIG. 4, and implement the functions of the face

authentication recognition apparatus in the embodiment shown in FIG. 4, which will not be elaborated here in the embodiments of this specification.

[0192] Definitely, in addition to the software implementation, the electronic device in the embodiment of this specification does not exclude other implementation manners, such as a logic device or a combination of software and hardware, etc. In other words, the following processing flow is not limited to being executed by various logic units and can also be executed by hardware or logic devices.

[0193] Embodiment 4

[0194] A computer-readable storage medium storing one or more programs is further provided in an embodiment of this specification, wherein when executed by a server including multiple applications, the one or more programs enable the server to perform the following operations:

[0195] acquiring multiple frames of depth images for a target detection object;

[0196] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;

[0197] normalizing the point cloud data to obtain a grayscale depth image; and

[0198] performing living-body detection based on the grayscale depth image and a living-body detection model.

[0199] A computer-readable storage medium storing one or more programs is further provided in an embodiment of this specification, wherein when executed by a server including multiple applications, the one or more programs enable the server to perform the following operations:

[0200] acquiring multiple frames of depth images for a target detection object;

[0201] pre-aligning the multiple frames of depth images to obtain pre-processed point cloud data;

[0202] normalizing the point cloud data to obtain a grayscale depth image;

[0203] performing living-body detection based on the grayscale depth image and a living-body detection model; and

[0204] determining whether the authentication recognition succeeds according to the
5 living-body detection result.

[0205] The computer-readable storage medium is, for example, a Read-Only Memory (ROM), a Random Access Memory (RAM), a magnetic disk, an optical disc, or the like.

[0206] Embodiment 5

[0207] Referring to FIG. 6a, a schematic structural diagram of a three-dimensional living-
10 body face detection apparatus according to an embodiment of this specification is shown. The apparatus mainly includes:

[0208] an acquisition module 602 configured to acquire multiple frames of depth images for a target detection object;

[0209] a first pre-processing module 604 configured to pre-align the multiple frames of
15 depth images to obtain pre-processed point cloud data;

[0210] a normalization module 606 configured to normalize the point cloud data to obtain a grayscale depth image; and

[0211] a detection module 608 configured to perform living-body detection based on the grayscale depth image and a living-body detection model.

20 [0212] With the above technical solution, multiple frames of depth images for a target detection object are acquired to ensure the overall performance of an image input as detection data; the multiple frames of depth images are pre-aligned and the point cloud data is normalized to obtain a grayscale depth image, which can ensure the integrity and accuracy of the grayscale depth image and compensate for the image quality problem; and finally, the living-body detection is performed
25 based on the grayscale depth image and a living-body detection model, thereby improving the

accuracy of the living-body detection. Then, more effective security verification or attack defense can be implemented based on the detection results.

[0213] Optionally, as an embodiment, when the living-body detection model is obtained,

[0214] the acquisition module 602 is configured to acquire multiple frames of depth images
5 for a target detection object;

[0215] the first pre-processing module 604 is configured to pre-align the multiple frames of depth images to obtain pre-processed point cloud data; and

[0216] the normalization module 606 is configured to normalize the point cloud data to obtain a grayscale depth image sample.

10 [0217] Moreover, referring to FIG. 6b, the apparatus further includes:

[0218] a training module 610 configured to train based on the grayscale depth image sample and label data of the grayscale depth image sample to obtain the living-body detection model.

[0219] Optionally, the first pre-processing module 604 is specifically configured to:

15 [0220] roughly align the multiple frames of depth images based on three-dimensional key facial points; and

[0221] finely align the roughly aligned depth images based on an ICP algorithm to obtain the point cloud data.

[0222] Optionally, referring to FIG. 6c, the three-dimensional living-body face detection
20 apparatus further includes:

[0223] a second pre-processing module 612 configured to bilaterally filter each frame of depth image in the multiple frames of depth images.

[0224] Optionally, the normalization module 604 is specifically configured to:

[0225] determine an average depth of the face region according to three-dimensional key
25 facial points in the point cloud data;

[0226] segment the face region, and delete a foreground and a background in the point cloud data; and

[0227] normalize the point cloud data from which the foreground and background have been deleted to preset value ranges before and after the average depth that take the average depth
5 as the reference to obtain the grayscale depth image.

[0228] Optionally, the preset value ranges from 30 mm to 50 mm.

[0229] Optionally, referring to FIG. 6d, the three-dimensional living-body face detection apparatus further includes:

[0230] an augmentation module 614 configured to perform data augmentation on the
10 grayscale depth image sample, wherein the data augmentation comprises at least one of the following: a rotation operation, a shift operation, and a zoom operation.

[0231] Optionally, the living-body detection model is a model obtained by training based on a convolutional neural network structure.

[0232] Optionally, the multiple frames of depth images are acquired based on an active
15 binocular depth camera.

[0233] Referring to FIG. 7, a schematic structural diagram of a face authentication recognition apparatus according to an embodiment of this specification is shown. The apparatus mainly includes:

[0234] an acquisition module 702 configured to acquire multiple frames of depth images
20 for a target detection object;

[0235] a first pre-processing module 704 configured to pre-align the multiple frames of depth images to obtain pre-processed point cloud data;

[0236] a normalization module 706 configured to normalize the point cloud data to obtain a grayscale depth image;

25 [0237] a detection module 708 configured to perform living-body detection based on the

grayscale depth image and a living-body detection model; and

[0238] a recognition module 710 configured to determine whether the authentication recognition succeeds according to the living-body detection result.

[0239] With the above technical solution, multiple frames of depth images for a target
5 detection object are acquired to ensure the overall performance of an image input as detection data; the multiple frames of depth images are pre-aligned and the point cloud data is normalized to obtain a grayscale depth image, which can ensure the integrity and accuracy of the grayscale depth image and compensate for the image quality problem; and finally, the living-body detection is performed based on the grayscale depth image and a living-body detection model, thereby improving the
10 accuracy of the living-body detection. Then, more effective security verification or attack defense can be implemented based on the detection results.

[0240] In brief, the above description is merely preferred embodiments of the embodiments of this specification and is not intended to limit the protection scope of the embodiments of this specification. Any modification, equivalent replacement, improvement and the like made without
15 departing from the spirit and principle of the embodiments of this specification should be included in the protection scope of the embodiments of this specification.

[0241] The system, apparatus, module or unit illustrated in the above embodiments may be implemented by a computer chip or an entity, or by a product having a certain function. A typical implementation device is a computer. For example, the computer may be a personal computer, a
20 laptop computer, a cellular phone, a camera phone, a smart phone, a personal digital assistant, a media player, a navigation device, an email device, a game console, a tablet computer, a wearable device, or a combination of any of these devices.

[0242] The computer-readable medium includes non-volatile and volatile media as well as movable and non-movable media and may implement information storage by means of any method
25 or technology. The information may be a computer-readable instruction, a data structure, a module

of a program or other data. An example of the storage medium of a computer includes, but is not limited to, a phase change memory (PRAM), a static random access memory (SRAM), a dynamic random access memory (DRAM), other types of RAMs, a ROM, an electrically erasable programmable read-only memory (EEPROM), a flash memory or other memory technologies, a compact disk read-only memory (CD-ROM), a digital versatile disc (DVD) or other optical storages, a cassette tape, a magnetic tape/magnetic disk storage or other magnetic storage devices, or any other non-transmission medium, and can be used to store information accessible to the computing device. According to the definition in this text, the computer-readable medium does not include transitory media, such as a modulated data signal and a carrier.

10 **[0243]** It should be further noted that the terms "include," "comprise" or any other variations thereof are intended to cover non-exclusive inclusion, so that a process, method, article or device including a series of elements not only includes the elements, but also includes other elements not expressly listed, or further includes elements inherent to the process, method, article or device. In the absence of more limitations, an element defined by "including a/an..." does not
15 exclude that the process, method, article or device including the element further has other identical elements.

[0244] Various embodiments in the embodiments of this specification are described in a progressive manner. The same or similar parts between the embodiments may be referenced to one another. In each embodiment, the part that is different from other embodiments is mainly described.
20 Particularly, the system embodiment is described in a relatively simple manner because it is similar to the method embodiment, and for related parts, reference can be made to the parts described in the method embodiment.