BIOMETRIC RECOGNITION METHOD, SERVER AND CLIENT

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to Chinese Patent Application CN202210108922.9, titled "BIOMETRIC RECOGNITION METHOD, SERVER AND CLIENT", filed on January 28, 2022, and International Application No. PCT/CN2022/113675, titled "BIOMETRIC RECOGNITION METHOD, SERVER AND CLIENT", filed on August 19, 2022, the disclosure of which is incorporated herein by reference in its entirety for all purposes.

TECHNICAL FIELD

[0002] The present application pertains to an identity recognition solution, more specifically, a biometric recognition method, a biometric recognition server, a client, a computer storage medium and a computer program product.

BACKGROUND

[0003] Over recent years, biometric features (e.g., face, fingerprint, etc.), owing to their close correlation with individual identities, have increasingly emerged as a mainstream method of identity verification and, thereby, have seen a considerable number of applications for important financial scenarios such as withdrawal, payment and mobile banking login.

[0004] Recently, the Technology Department of the People's Bank of China issued a "Financial Technology Applications Risk Warning" referring to media reports about breached facial recognition systems of multiple Android phones, which warns of the potential risks associated with the application of facial technology to the financial payment sector. As noted in the warning, the method employed to breach the facial recognition systems is based on the adversarial sample attack technique of Generative Adversarial Networks.

[0005] Accordingly, anticipation builds up for the development of an identity verification solution based on biometric features that can effectively prevent the adversarial sample attack technique.

SUMMARY

[0006] According to one aspect of the disclosure, a biometric recognition method is provided, the method comprising: Step A: receiving biometric data; Step B: performing

a random transformation of low-bit data in received biometric data; Step C: randomly selecting one or more filtering algorithms to pre-process the randomly transformed biometric data; and Step D: performing identity verification based on the pre-processed biometric data.

[0007] As a supplement or alternative to the solution, the biometric data of the method includes facial image data.

[0008] As a supplement or alternative to the solution, "performing a random transformation of low-bit data in the received biometric data" of the method includes: performing a random transformation of the lowest bits in the biometric data.

[0009] As a supplement or alternative to the solution, "performing a random transformation of the lowest bits in the biometric data" of the method includes any one of the following: setting the lowest bits of the biometric data all to 0; setting the lowest bits of the biometric data all to 1; randomly re-arranging the lowest bits of the biometric data; and randomly generating the lowest bits of the biometric data.

[0010] As a supplement or alternative to the solution, the identity verification based on the pre-processed biometric data performed in the Step D to obtain a first comparison result further includes: repeating execution of Steps B to D to obtain a second comparison result; and directly performing identity verification based on the received biometric data to obtain a third comparison result.

[0011] As a supplement or alternative to the solution, the method further includes: comparing the first comparison result, the second comparison result and the third comparison result for consistency; and feeding back the first comparison result when the first comparison result, the second comparison result, and the third comparison result are consistent.

[0012] As a supplement or alternative to the solution, the method further includes: feeding back a recognition process anomaly when the first comparison result, the second comparison result, and the third comparison result are inconsistent; and adding the corresponding data collection terminal to a blacklist.

[0013] As a supplement or alternative to the solution, according to the method, the biometric data further includes a timestamp, and the method further comprises between the Steps A and B: decrypting the biometric data to obtain the facial image data and the timestamp; and feeding back an anomaly when a difference between system time and the timestamp does not meet requirements.

[0014] As a supplement or alternative to the solution, between Steps A and B, the method further comprises: detecting a facial position in the facial image data when the difference between the system time and the timestamp meets the requirements.

[0015] As a supplement or alternative to the solution, between Steps A and B, the method further includes: adding an adversarial sample to the facial position, the adversarial sample devised to counteract action of a negative adversarial sample added at the data collection terminal, wherein the negative adversarial sample is a negative value of the adversarial sample.

[0016] As a supplement or alternative to the solution, according to the method, the adversarial sample is pre-generated by performing a back-propagation training on a generator based on facial recognition results.

[0017] According to another aspect of the disclosure, a biometric recognition server is provided, the server comprising: a receiving device to receive biometric data; a low-bit random filter to randomly transform low-bit data in received biometric data; a preprocessor to randomly select one or more filtering algorithms to pre-process randomly transformed biometric data; and an identity verification device to perform identity verification based on pre-processed biometric data.

[0018] As a supplement or alternative to the solution, in the server, the biometric data comprises facial image data.

[0019] As a supplement or alternative to the solution, the low-bit random filter of the server is configured to randomly transform lowest bits in the biometric data.

[0020] As a supplement or alternative to the solution, the low-bit random filter of the server is configured to perform any one of the following: set the lowest bits of the biometric data all to 0; set the lowest bits of the biometric data all to 1; randomly re-arrange the lowest bits of the biometric data; and randomly generate the lowest bits of the biometric data.

[0021] As a supplement or alternative to the solution, the server further comprises: a recognition decision device to compare multiple recognition results for consistency and to feed back the results when all the recognition results are consistent.

[0022] As a supplement or alternative to the solution, the recognition decision device of the server is further configured to: feed back a recognition process anomaly when the recognition results are inconsistent; and add the corresponding data collection terminal to a blacklist.

[0023] As a supplement or alternative to the solution, the server further comprises the biometric data further comprises a timestamp.

[0024] As a supplement or alternative to the solution, the server further comprises: a decryption device to decrypt the biometric data to obtain the facial image data and the timestamp; and a verification device to feed back an anomaly when a difference between system time and the timestamp does not meet requirements.

[0025] As a supplement or alternative to the solution, the verification device of the server is further configured to: detect a facial position in the facial image data when the difference between the system time and the timestamp meets the requirements.

[0026] As a supplement or alternative to the solution, the verification device of the server is further configured to: the verification device is further configured to: add in the facial position an adversarial sample for counteracting action of a negative adversarial sample added at the data collection terminal, wherein the negative adversarial sample is a negative value of the adversarial sample.

[0027] As a supplement or alternative to the solution, in the server, the adversarial sample is pre-generated by performing a backpropagation training on a generator based on facial recognition results.

[0028] According to another aspect of the disclosure, a client for use in conjunction with the server is provided, the client comprising: a facial data collection device to collect facial image data; and an embedding device to embed a negative adversarial sample in a facial position, wherein the negative adversarial sample is the additive inverse of the adversarial sample, and the adversarial sample is additive noise superimposed on real facial image data, causing a facial recognition model to produce an erroneous recognition result.

[0029] As a supplement or alternative to the solution, the embedding device of the client is configured to: detect the facial position based on the facial image data; embed the negative adversarial sample in the facial position; and send a processed facial image and a timestamp in an encrypted manner.

[0030] According to another aspect of the disclosure, a method for generating an adversarial sample using a biometric recognition server as discussed previously is provided, the method comprising: Step A: constructing an adversarial generative network comprising a generator and a discriminator; Step B: generating noise information with the generator, and adding the noise information to real facial image data to obtain generated facial data; Step C: recognizing the generated facial data with a facial recognition model; Step D: performing a backpropagation training on the generator based on recognition results; Step E: repeating the Steps B to D until the facial recognition model produces an erroneous recognition result; and Step F: determining the noise information corresponding to the erroneous recognition result as an adversarial sample.

[0031] According to another aspect of the disclosure, a computer storage medium is provided, the medium comprising instructions, wherein the instructions, when running, implement the method as discussed previously.

[0032] According to another aspect of the disclosure, a computer program product is provided, the product comprising a computer program, wherein the computer program, when executed by a processor, implements the method as discussed previously.

[0033] Compared to existing biometric recognition solutions, the biometric recognition solution, according to one or more embodiments of the disclosure, performs a random transformation on low-bit data in received biometric data and carries out identity verification based on pre-processed biometric data. The solution significantly disrupts adversarial noise in images, neutralizing adversarial attacks. Consequently, even if attackers were to obtain the parameters of a biometric recognition model, the training process for an adversarial attack model would fail to achieve stable convergence, thereby failing to develop an effective attack model.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] The aforementioned and other objectives and advantages of the disclosure will become more comprehensive and clear through the detailed description below, in conjunction with the appended drawings, in which identical or similar elements are denoted by the same reference signs.

[0035] Figure 1 presents a flowchart of a biometric recognition method according to an embodiment of the disclosure;

[0036] Figure 2 illustrates the principle of an adversarial sample attack;

[0037] Figure 3 shows a schematic structure of a biometric recognition server according to an embodiment of the disclosure;

[0038] Figure 4 depicts a flowchart of facial recognition according to an embodiment of the disclosure;

[0039] Figure 5 displays a schematic diagram of a facial recognition system including a data collection terminal and a facial authentication service platform according to an embodiment of the disclosure; and

[0040] Figure 6 demonstrates a flowchart of facial recognition according to another embodiment of the disclosure.

DETAILED DESCRIPTION

[0041] The implementation mode of the disclosure is further described in details below, with reference to the drawings and embodiments. The following embodiments are intended to illustrate the disclosure and should not be seen as limitation on the scope of the disclosure.

[0042] In the description of the Specification, the terms, such as "an embodiment," "some embodiments", "example", "specific examples" or "some examples", indicate that a particular feature, structure, material, or characteristic described in connection with the embodiment or example is included in at least one embodiment or example of the present disclosure. According to the Specification, the symbolic representation of the aforementioned terms does not necessarily refer to the same embodiment or example. Moreover, the specific features, structures, materials or characteristics described may be combined by suitable means in one or more embodiments or examples. Furthermore, the terms "first", "second" and "third" are devised for descriptive purposes only, which should not be construed as indicating or implying relative importance. Besides, where not mutually exclusive, those skilled in the art may integrate and combine different embodiments or examples described in the Specification, as well as the features of different embodiments or examples.

[0043] In the context of the disclosure, the term "client," also referred to as a user end, denotes a device or apparatus that provides local services to customers in correspondence with servers. In one or more embodiments, the client possesses basic functionalities, including the capability to collect the biometric information of users. Such a client can be a smart device belonging to a user, including, but not limited to, a user's mobile phone, laptop and head-mounted devices.

[0044] The term "server," also referred to as a remote server or cloud server, denotes a device or apparatus that provides remote services to the client, corresponding to the "client." In one or more embodiments, the server can offer biometric recognition functionalities resistant to adversarial sample attacks (e.g., facial recognition), meaning that the server performs biometric recognition on the biometric data received from the client after undergoing certain preprocessing steps (such as denoising).

[0045] Figure 1 presents a flowchart of a biometric recognition method 1000 according to an embodiment of the disclosure. As illustrated in Figure 1, the method 1000 comprises the following steps:

[0046] At step S110, receiving biometric data;

[0047] At step S120, performing a random transformation of low-bit data in received biometric data;

[0048] At step S130, randomly selecting one or more filtering algorithms to pre-process randomly transformed biometric data; and

[0049] At step S140, performing identity verification based on pre-processed biometric data.

[0050] In the context of the disclosure, the term "biometric data" refers to any inherent data of a human body that can be used for personal identification, as well as auxiliary information related to the noted data. For example, biometric data includes, but is not limited to, physiological characteristics (such as fingerprints, irises, facial features and DNA) and behavioral characteristics (such as gait and typing habits).

[0051] In one embodiment, the biometric data includes facial image data. In another embodiment, in addition to facial image data, the biometric data also includes a timestamp associated with the facial image data. The timestamp could be either the time when the facial image data was recorded or the time when the facial image data was uploaded.

[0052] Referencing Figure 2, the figure illustrates the principle of adversarial sample attacks. The "adversarial sample" refers to a sample that can mislead classifiers into making incorrect predictions without affecting human visual recognition, e.g., adversarial samples could be anormal samples created by attackers through slight perturbations of normal samples. In one embodiment, an adversarial sample can cause a CNN model to yield a highly confident incorrect output. Figure 2 shows an adversarial sample in a facial comparison algorithm. After superimposing adversarial noise, the facial comparison score dramatically drops from 100 to below 60. That is, the facial comparison algorithm can no longer successfully recognize these two individuals in this case.

[0053] Adversarial sample attacks can be categorized into white-box attacks and black-box attacks. A white-box attack refers to the scenario where an attacker knows all information about the recognition model, including training data, model architecture, hyper-parameters, the number of layers, activation functions and model weights, and generates adversarial samples for attack through calculation of model gradients, while a black-box attack refers to attacks conducted without knowledge of the aforementioned information.

[0054] From the principle of adversarial sample attacks, it is understood that the fundamental cause of misrecognition in a biometric recognition system (e.g., a facial recognition system) is the addition of additive adversarial noise to the biometric data (e.g., facial image data) by attackers.

[0055] The inventor of the present application discovers that most of the data values of adversarial noise or adversarial samples are extremely low, for example, only

the low bits of 8-bit image data are changed. Thus, performing random transformations on the low-bit data of the biometric data can disrupt most data of adversarial noise. Here, "low-bit data" refers to the lowest bits, the last two bits, or the last three bits of data inherent to the human body that can be used for personal identification.

[0056] In one embodiment, step S120 includes performing a random transformation on the lowest bits of the biometric data. Certainly, those skilled in the art can understand that, in one embodiment, a random transformation can also be applied to the last two bits or the last three bits of the biometric data.

[0057] In one embodiment, performing a random transformation on the lowest bits of the biometric data includes any one of the following: setting the lowest bits of the biometric data all to 0; setting the lowest bits of the biometric data all to 1; randomly re-arranging the lowest bits of the biometric data; and randomly generating the lowest bits of the biometric data. For example, for 8-bit image data, the random transformation of the lowest bits can include the following four algorithms: (1) setting the 8-bit lowest bits all to 0; (2) setting the 8-bit lowest bits all to 1; (3) randomly re-arranging the 8-bit lowest bits; and (4) randomly generating the 8-bit lowest bits. In one or more embodiments, the algorithm for the random transformation of the lowest bits is not fixed but is randomly selected from multiple algorithms (e.g., the said four algorithms) to pre-process an input image.

[0058] After performing a random transformation on the lowest bits of the biometric data, it is advantageous to filter (pre-process) the randomly transformed biometric data again, thereby smoothing the biometric data (e.g., smoothing facial image data). For example, by randomly selecting one or more filtering algorithms, the randomly transformed biometric data is preprocessed.

[0059] Taking image data (e.g., facial image data) as an example, there are various image denoising/filtering algorithms, including, including but not limited to, BM3D denoising; DCT denoising; PCA denoising; K-SVD denoising; non-local means denoising; WNNM denoising; image denoising algorithms based on principal component analysis and bilateral filtering; wavelet transformation; wavelet threshold denoising; Contourlet transform; SAR image denoising based on shift-invariant Contourlet transform, etc. In one or more embodiments, one or several of these 11 filtering algorithms are randomly selected to pre-process an input image. Normally, removing noise does not affect the recognition results of a facial recognition system, which instead significantly disrupts the adversarial noise in the image, rendering the adversarial attack ineffective.

[0060] In one embodiment, in step S140, the identity verification is performed based on pre-processed biometric data to obtain a first comparison result. Although not shown in Figure 1, the aforementioned method 1000 may also comprise: repeating the aforementioned steps S120 to S140 to obtain a second comparison result (wherein,

the filtering algorithm employed in re-executing steps S120 and S130 is different or partially different from that used in the first execution); and directly performing identity verification based on received biometric data to obtain a third comparison result.

[0061] In the aforementioned embodiment, the method 1000 also comprises: comparing the first, second, and third comparison results for consistency; and feeding back the first comparison result when the first comparison result, the second comparison result and the third comparison result are consistent. However, on condition that there is inconsistency among the first comparison result, the second comparison result and the third comparison result, a recognition process anomaly is fed back; and a corresponding data collection terminal is added to a blacklist.

[0062] Evidently, the third recognition is an identity verification carried out directed at the original biometric data (i.e., without undergoing the random transformation of lower bits), while the first and the second recognitions are both of identity verifications performed on biometric data undergone random transformation of lower bits (where different filtering algorithms may be employed). On condition that the results of these three identity verifications are consistent, the identity verification result can be directly fed back. However, when the results of multiple identity verifications (e.g., three times) differ, a recognition process anomaly (likely due to an adversarial sample attack) may be fed back, with the terminal where the anomalous data are collected being added to a blacklist. Subsequently, identity verification services would be denied to the terminal.

[0063] Thus, through the addition of a multi-image verification strategy and a blacklist rejection strategy, it is capable of effectively preventing attackers from conducting black-box mode attack training.

[0064] At present, the most significant issue affecting facial recognition is camera hijacking attacks, which severely damage the security of facial recognition systems. The primary principle of such attacks is as follows: firstly, an attacker "cracks" a payment app to hijack camera data; next, the attacker completes the live detection process required for facial recognition; finally, the attacker replaces video data uploaded by a camera with pre-fabricated victim video data and uploads falsified data to a backend facial recognition system, thereby completing the entire attack process. Currently, the vast majority of facial recognition systems lack the capability to defend against such camera hijacking attacks.

[0065] To address the issue, in one embodiment, between Steps S110 and S120, the method 1000 may further comprise: decrypting the biometric data to obtain the facial image data and the timestamp; feeding back an anomaly when the difference between the system time and the timestamp does not meet the requirements; when the difference

between the system time and the timestamp meets the requirements, detecting a facial position in the facial image data and adding an adversarial sample to the facial position for facial recognition.

[0066] Here, adding an adversarial sample at the facial position is intended to counteract the impact of embedding a negative adversarial sample during a facial image collection process. That is, during the facial image collection process, the facial collection terminal (e.g., client) needs to detect the facial position, embed a negative adversarial sample in the facial detection area, and transmit the processed facial image, facial position, and timestamp in an encrypted form to a facial recognition authentication server. Thus, even in the event of camera hijacking, images not embedded with a negative adversarial sample at the time of collection will also be embedded with an adversarial sample before an authentication service performs facial recognition, thereby failing to yield correct results in the process of facial recognition.

[0067] In one embodiment, adversarial samples are pre-generated through back-propagation training of a generator based on the results of facial recognition. For instance, adversarial samples can be generated through the following steps: (1) Constructing an adversarial generative network, where a generative model is responsible for creating a fake facial image while a discriminator serves as a facial recognition model, with the parameters of the facial recognition model fixed and only the parameters of a generator optimized; (2) generating noise information with the generator, and adding the noise information to real facial image data to form fake data and fake labels; (3) recognizing generated facial data with a facial recognition model; (4) performing a back-propagation training on the generator based on recognition results; (5) repeating Steps (2) to (4) until the facial recognition model produces an erroneous recognition result. At this point, the image produced by the generator is adversarial sample A, and its additive inverse, -A, is the negative adversarial sample. It is understood that an adversarial sample belongs to additive noise while a negative adversarial sample is the negative value of the adversarial sample, intended to counteract the effect of the adversarial sample.

[0068] In one embodiment, a negative adversarial sample needs to be generated only once, with subsequent embedding and removal processes in images being straightforward and efficient. Moreover, facial data embedded with a negative adversarial sample do not affect human vision, meaning that even if the data are cracked during transmission, the data cannot be removed.

[0069] Turning to Figure 3, it illustrates a schematic structure of the biometric recognition server 3000 according to an embodiment of the disclosure. As shown in Figure 3, the biometric recognition server 3000 comprises: a receiving device 310, a low-bit random filter 320, a preprocessor 330, and an identity verification device 340,wherein, specifically, the receiving device 310 is devised to receive biometric data, the low-bit random filter 320 is used to randomly transform low-bit data in received

biometric data, the preprocessor 330 is configured to randomly select one or more filtering algorithms to pre-process randomly transformed biometric data, and the identity verification device 340 is intended to perform identity verification based on pre-processed biometric data.

[0070] In the context of the disclosure, the term "biometric data" refers to any inherent data of a human body that can be used for personal identification, as well as auxiliary information related to the noted data. For example, biometric data includes, but is not limited to, physiological characteristics (such as fingerprints, irises, facial features and DNA) and behavioral characteristics (such as gait and typing habits).

[0071] In one embodiment, the biometric data includes facial image data. In another embodiment, in addition to facial image data, the biometric data also includes a timestamp associated with the facial image data. The timestamp could be either the time when the facial image data was recorded or the time when the facial image data was uploaded.

[0072] An "adversarial sample" refers to a sample that can mislead classifiers into making incorrect predictions without affecting human visual recognition, e.g., adversarial samples could be anormal samples created by attackers through slight perturbations of normal samples. In one embodiment, the adversarial sample can cause a CNN model to yield a highly confident incorrect output. Figure 2 shows an adversarial sample in a facial comparison algorithm. After superimposing adversarial noise, the facial comparison score dramatically drops from 100 to below 60. That is, the facial comparison algorithm can no longer successfully recognize these two individuals in this case. Adversarial sample attacks are categorized into white-box attacks and black-box attacks. A white-box attack refers to the scenario where an attacker knows all information about the recognition model, including training data, model architecture, hyper-parameters, the number of layers, activation functions and model weights, and generates adversarial samples for attack through calculation of model gradients, while a black-box attack refers to attacks conducted without knowledge of the aforementioned information.

[0073] From the principle of adversarial sample attacks, it is understood that the fundamental cause of misrecognition in a biometric recognition system (e.g., a facial recognition system) is the addition of additive adversarial noise to the biometric data (e.g., facial image data) by attackers.

[0074] The inventor of the present application discovers that most of the data values of adversarial noise or adversarial samples are extremely low, for example, only the low bits of 8-bit image data are changed. Thus, performing random transformations on the low-bit data of the biometric data can disrupt most data of adversarial noise. Here, "low-bit data" refers to the lowest bits, the last two bits, or the last three bits of data inherent to the human body that can be used for personal identification.

[0075] In one embodiment, the low-bit random filter 320 is configured to randomly transform lowest bits in the biometric data. Certainly, those skilled in the art can understand that, in one embodiment, the random filter device 320 can perform a random transformation on the last two bits or the last three bits of the biometric data, not limited to the lowest bits.

[0076] In one embodiment, the low-bit random filter 320 is configured to perform any one of the following: setting the lowest bits of the biometric data all to 0; setting the lowest bits of the biometric data all to 1; randomly re-arranging the lowest bits of the biometric data; and randomly generating the lowest bits of the biometric data. For example, for 8-bit image data, the random transformation of the lowest bits can include the following four algorithms: (1) setting the 8-bit lowest bits all to 0; (2) setting the 8-bit lowest bits all to 1; (3) randomly re-arranging the 8-bit lowest bits; and (4) randomly generating the 8-bit lowest bits. In one or more embodiments, the algorithm employed by the low-bit random filter 320 is not fixed but is randomly selected from multiple algorithms (such as the said four algorithms) to pre-process an input image.

[0077] After the low-bit random filter 320 performs a random transformation on the lowest bits of the biometric data, it is advantageous that the preprocessor 330 filters (pre-process) randomly transformed biometric data again, thereby smoothing the biometric data (e.g., smoothing facial image data). For instance, the preprocessor 330 is configured to pre-process the randomly transformed biometric data by randomly selecting one or more filtering algorithms.

[0078] Taking image data (e.g., facial image data) as an example, there are various image denoising/filtering algorithms, including, including but not limited to, BM3D denoising; DCT denoising; PCA denoising; K-SVD denoising; non-local means denoising; WNNM denoising; image denoising algorithms based on principal component analysis and bilateral filtering; wavelet transformation; wavelet threshold denoising; Contourlet transform; SAR image denoising based on shift-invariant Contourlet transform, etc. In one or more embodiments, the preprocessor is configured to randomly select one or several of these 11 filtering algorithms to pre-process the input image. Normally, noise removal does not affect the recognition results of a facial recognition system but significantly disrupts the adversarial noise in the image, rendering the adversarial attack ineffective. In one or more embodiments, one or several of these 11 filtering algorithms are randomly selected to pre-process an input image. Normally, removing noise does not affect the recognition results of a facial recognition system, which instead significantly disrupts the adversarial noise in the image, rendering the adversarial attack ineffective.

[0079] In one embodiment, the server 3000 may also comprise: a recognition decision device 350 to compare multiple recognition results for consistency and to feed back the

results when all the recognition results are consistent. For example, the first recognition is an identity verification carried out directed at the original biometric data (i.e., without undergoing the random transformation of lower bits), while the first and the second recognitions are both of identity verifications performed on biometric data undergone random transformation of lower bits (where different filtering algorithms may be employed). On condition that the results of these three identity verifications are consistent, the identity verification result can be directly fed back.

[0080] In one embodiment, the recognition decision device 350 is further configured to: feed back a recognition process anomaly when the recognition results are inconsistent; add the corresponding data collection terminal to a blacklist. This means, when the results of multiple identity verifications (e.g., three times) differ, a recognition process anomaly (likely due to an adversarial sample attack) may be fed back, with the terminal where the anomalous data are collected being added to a blacklist. Subsequently, identity verification services would be denied to the terminal. Thus, through the addition of a multi-image verification strategy and a blacklist rejection strategy, it is capable of effectively preventing attackers from conducting black-box mode attack training.

[0081] At present, the most significant issue affecting facial recognition is camera hijacking attacks, which severely damage the security of facial recognition systems. The primary principle of such attacks is as follows: firstly, an attacker "cracks" a payment app to hijack camera data; next, the attacker completes the live detection process required for facial recognition; finally, the attacker replaces video data uploaded by a camera with pre-fabricated victim video data and uploads falsified data to a backend facial recognition system, thereby completing the entire attack process. Currently, the vast majority of facial recognition systems lack the capability to defend against such camera hijacking attacks.

[0082] To address this issue, in one embodiment, the server 3000 may also comprise: a decryption device 360to decrypt the biometric data to obtain the facial image data and the timestamp; and a verification device 370 to feed back an anomaly when a difference between system time and the timestamp does not meet requirements. In one embodiment, the verification device 370 is further configured, when the difference between the system time and the timestamp meets the requirements, to detect a facial position in the facial image data and to add an adversarial sample at the facial position for facial recognition.

[0083] The verification device 370 adds an adversarial sample at the facial position to counteract the effect of a negative adversarial sample embedded during the facial image collection process. That is, during the facial image collection process, the facial collection terminal (e.g., client) needs to detect the facial position, embed a negative adversarial sample in the facial detection area, and transmit the processed facial image,

facial position, and timestamp in an encrypted form to a facial recognition authentication server. Thus, even in the event of camera hijacking, images not embedded with a negative adversarial sample at the time of collection will also be embedded with an adversarial sample before an authentication service performs facial recognition, thereby failing to yield correct results in the process of facial recognition.

[0084] In one embodiment, adversarial samples are pre-generated by performing back-propagation training on a generator based on the results of facial recognition. For instance, adversarial samples can be generated by a biometric recognition server, with the specific generation method comprising: Step A: constructing an adversarial generative network comprising a generator and a discriminator; Step B: generating noise information with the generator, and adding the noise information to real facial image data to obtain generated facial data; Step C: recognizing the generated facial data with a facial recognition model; Step D: performing a back-propagation training on the generator based on recognition results; Step E: repeating the Steps B to D until the facial recognition model produces an erroneous recognition result; and Step F: determining the noise information corresponding to the erroneous recognition result as an adversarial sample A. The additive inverse, -A, is the negative adversarial sample. It is understood that an adversarial sample belongs to additive noise while a negative adversarial sample is the negative value of the adversarial sample, intended to counteract the effect of the adversarial sample.

[0085] In one embodiment, a negative adversarial sample needs to be generated only once, with subsequent embedding and removal processes in images being straightforward and efficient. Moreover, facial data embedded with a negative adversarial sample do not affect human vision, meaning that even if the data are cracked during transmission, the data cannot be removed.

[0086] Figure 4 depicts a flowchart of facial recognition according to an embodiment of the disclosure. As shown in Figure 4, the facial recognition system may include multiple devices or modules, such as the facial data collection module, the pixel low-bit random filter, the random noise filter, the facial recognition function module, and recognition decision module, as depicted in Figure 4. In one embodiment, the facial data collection module may be located on a client side, for example, included in a user's smartphone. The pixel low-bit random filter, the random noise filter, the facial recognition function module and the recognition decision module may be located on a server side, for example, within a facial authentication service platform.

[0087] The facial data collection module is responsible for collecting facial data. Given that the majority of adversarial noise data values are extremely low, affecting only the lowest bits of 8-bit image data, the pixel low-bit random filter transforms the lowest bits of the image data randomly, thus disrupting the majority of the adversarial noise data. The filtering algorithms contained in the pixel low-bit random filter primarily consist of: setting the 8-bit lowest bits all to 0; setting the 8-bit lowest bits all

to 1; randomly re-arranging the 8-bit lowest bits; and randomly generating the 8-bit lowest bits. During the filtering process, the pixel low-bit random filter can randomly select from the above 4 algorithms to pre-process the input image.

[0088] The random noise filter's denoising algorithms include, but are not limited to, the following methods: BM3D denoising; DCT denoising; PCA denoising; K-SVD denoising; NL-Means non-local denoising; WNNM denoising; image denoising algorithms based on principal component analysis and bilateral filtering; wavelet transformation; wavelet threshold denoising; Contourlet transform; and SAR image denoising based on shift-invariant Contourlet transform. The random noise filter can randomly select one or several from the above 11 filtering algorithms to preprocess the input image.

[0089] The facial recognition function module is responsible for completing the process of facial recognition in images. The recognition decision module compares the results of multiple facial comparisons for consistency, wherein, on condition that the results are consistent, the results are fed back; on condition that the results are inconsistent, a recognition process anomaly is fed back and a corresponding data collection terminal is added to a blacklist.

[0090] In one embodiment, after the facial data collection module completes data collection, the facial recognition function module performs a first facial recognition on the original facial data, feeding back the recognition results to the recognition decision module. next, the original facial data undergoes first pixel low-bit random filtering and first random noise filtering, followed by a second facial comparison on the filtered data. Subsequently, the original facial data undergoes second pixel low-bit random filtering and second random noise filtering (ensuring that the method of the second filtering is different from that of the first filtering), followed by a third facial comparison on the filtered data. Finally, the recognition decision module compares the results of the three filtrations: on condition that the three recognition results are consistent, the recognition decision module feeds back a normal value for facial recognition; on condition that the three recognition results are inconsistent, the recognition decision module feeds back that there is a recognition anomaly (the anomaly value is S) and adds the terminal that collects the anomalous data to a blacklist. Subsequently, identity verification services would be denied to the terminal.

[0091] Figure 5 displays a schematic diagram of a facial recognition system including a data collection terminal and a facial authentication service platform according to an embodiment of the disclosure. The data collection terminal can be located on the user side, while the facial authentication service platform be on the remote server side.

[0092] In the embodiment shown in Figure 5, the data collection terminal provides operational resources for a camera and a Trusted Execution Environment (TEE), where the camera is responsible for collecting facial data and the TEE for providing a secure

operation and storage environment for the applications and drivers of the data collection terminal. As shown in Figure 5, the data collection terminal also includes a negative adversarial sample embedding module, the embedding module devised to detect a facial position, embed a negative adversarial sample in a facial detection area and to transmit a processed facial image, facial position and timestamp in an encrypted form to the facial authentication service platform located remotely.

[0093] In an embodiment, the facial authentication service platform is devised to fulfill facial recognition authentication services, a negative adversarial sample verification module devised to decrypt the facial image, facial position and timestamp, while a facial recognition model is intended to recognize a verified facial recognition image, feeding back the identity information of an identified individual.

[0094] Figure 6 demonstrates a flowchart of facial recognition using the facial recognition system shown in Figure 5. As illustrated in Figure 6, before facial recognition is conducted, the data collection terminal carries out application initialization, storing a facial detection model and a negative adversarial sample -A in a Trusted Execution Environment. Next, the camera captures a facial image; the negative adversarial sample embedding module detects a facial position P in the facial image, and after embedding the negative adversarial sample -A in P area, encrypts and uploads the facial image along with an embedded timestamp t to the facial authentication service platform. Following this, the server-side negative adversarial sample verification module decrypts the facial data and the embedded timestamp t, verifies the embedded timestamp t. On condition that the difference between the system time and the t does not meet the requirements, an anomaly is fed back to the platform; otherwise, the facial position P is detected with the adversarial sample A embedded in the position P. Finally, the facial recognition model executes the facial recognition process and feeds back the recognition result.

[0095] In one embodiment, the process of embedding a negative adversarial sample may include the following steps: detecting a facial position P (x1, y1, x2, y2) in an input image F, wherein x and y represent the horizontal and vertical coordinates of a facial frame, respectively, and pixel block content determined by the position P in the image is denoted as S; setting S1 = (S - A) mod 256, and replacing the value S with S1, at which point the image transforms into F1; recording negative sample embedding time t; and encrypting (F1, t).

[0096] In one embodiment, the process of verifying a negative adversarial sample may include the following steps: decrypting (F1, t); verifying that the difference between system time t1 and a timestamp t is within a range (-a, a); detecting a facial position P (x1, y1, x2, y2) in an input image F, wherein the pixel block content determined by the position P in the image denoted as S1; setting S = (S1 + A) mod 256; replacing the value S1 with S, at which point the image reverts to F.

[0097] It should be noted that although, in Figure 6, the negative adversarial sample embedding module and the negative adversarial sample verification module are shown as single modules, those skilled in the art can understand that these modules can be further subdivided according to function or need, for example, the negative adversarial sample verification module can be further divided into a decryption module and a verification module.

[0098] Furthermore, the device embodiments described above are merely illustrative, where the modules described as separate components may not be physically separate, that is, they may be located in one place or be distributed across multiple network modules. Part or all of the modules can be selected according to practical requirements to achieve the purpose of this embodiment, which those skilled in the art can understand and implement without investment of creative work.

[0099] Through the description of various embodiments referenced above, those skilled in the art can clearly learn that all the embodiments can be implemented either through software in conjunction with the necessary general hardware platform, or, of course, by hardware. Based on such understanding, the essential or contributory part of the above technical solutions to the prior art can be embodied in the form of a software product, which can be stored on a computer-readable storage medium. The computer-readable storage medium comprises any mechanism for storing or transmitting information in a form readable by a computer (e.g., a computer), such as read-only memory (ROM), random access memory (RAM), disk storage media, optical storage media, flash storage media, electrical, optical, acoustic, or other forms of propagation signals (e.g., carrier waves, infrared signals, digital signals, etc.). The computer software product comprises several instructions to enable a computing device (which can be a personal computer, server, or network device, etc.) to execute the method described in various embodiments or some parts thereof.

[0100] The present application is described with reference to the flowcharts and/or block diagrams of the methods, apparatuses (devices), and computer program products according to the embodiments of the present application. It should be understood that each process and/or block within the flowcharts and/or block diagrams, as well as combinations of the processes and/or blocks in the flowcharts and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general-purpose computer, a special purpose computer, an embedded processor, or other programmable data processing apparatuses to create a machine, such that the instructions executed via the processor of the computer or other programmable data processing apparatuses establish means for implementing the functions specified in one or more processes within the flowcharts and/or one or more blocks within the block diagrams.

[0101] Moreover, these computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data

processing apparatuses to function in a particular manner, such that the instructions stored in the computer-readable memory create a manufactured product including an instruction means, the instruction means implementing the functions specified in one or more processes within the flowcharts and/or one or more blocks within the block diagrams.

[0102] These computer program instructions can also be loaded onto a computer or other programmable data processing apparatuses, enabling a series of operational steps to be executed on the computer or other programmable device to produce computer-implemented processing. As such, the instructions executed on the computer or other programmable apparatuses provide steps for implementing the functions specified in one or more processes of the flowchart and/or one or more blocks of the block diagram.

[0103] In summary, according to one or more embodiments of the disclosure, the biometric recognition solution is intended to perform identity verification through random transformation of low-bit data in received biometric data and based on pre-processed biometric data. The solution significantly disrupts the adversarial noise in images, rendering the adversarial attack ineffective. As such, even if attackers were to obtain the parameters of a biometric recognition model, the training process of an adversarial attack model would fail to achieve stable convergence, thereby failing to develop an effective attack model.

[0104] Additionally, in one or more embodiments, by incorporating a multiple image verification strategy and blacklist denial strategy, it is capable of effectively preventing attackers from conducting black-box mode attack training. Moreover, the process of embedding a negative adversarial sample is added during a facial image collection process and the process of verifying (removing) a negative adversarial sample is added at a facial recognition system end. Due to such design, images not embedded with a negative adversarial sample at the time of collection will also be embedded with an adversarial sample before an authentication service performs facial recognition, thereby failing to yield correct results in the process of facial recognition. A negative adversarial sample needs to be generated only once, with subsequent embedding and removal processes in an image being computationally simple and efficient. Moreover, facial data embedded with a negative adversarial sample do not affect human vision, meaning that even if the data are cracked during transmission, the data cannot be removed.

[0105] Although preferred embodiments are already described, those skilled in the art, upon acquiring the primary creative concepts, could make additional changes and modifications to such embodiments. For this reason, the appended claims are intended to be explained as those encompassing the preferred embodiments and all the changes and modifications falling within the scope of the present application. Evidently, those skilled in the art are allowed to make various changes and modifications to the present application without departing from its spirit and scope of the application. Consequently,

when these modifications and variations of the present application fall within the scopes of the claims of the present application and their technological equivalents, the present application also intends to include such modifications and variations.

What is claimed is:

1. A biometric recognition method, comprising:
Step A: receiving biometric data;
Step B: performing a random transformation of low-bit data in received biometric data;
Step C: randomly selecting one or more filtering algorithms to pre-process randomly transformed biometric data; and
Step D: performing identity verification based on the pre-processed biometric data.

2. The method according to claim 1, wherein the biometric data includes facial image data.

3. The method according to claim 1, wherein the step of " performing a random transformation of low-bit data in received biometric data" comprises: performing a random transformation of lowest bits in the biometric data.

4. The method according to claim 3, wherein the step of "performing a random transformation of lowest bits in the biometric data" comprises any one of the following:
setting the lowest bits of the biometric data all to 0;
setting the lowest bits of the biometric data all to 1;
randomly re-arranging the lowest bits of the biometric data; and
randomly generating the lowest bits of the biometric data.

5. The method according to claim 1, wherein the identity verification based on the pre-processed biometric data is performed in the Step D to obtain a first comparison result, and the method further comprises:
repeating execution of the Steps B to D to obtain a second comparison result; and
directly performing identity verification based on received biometric data to obtain a third comparison result.

6. The method according to claim 5, further comprising:
comparing the first comparison result, the second comparison result and the third comparison result for consistency; and
feeding back the first comparison result when the first comparison result, the second comparison result and the third comparison result are consistent.

7. The method according to claim 6, further comprising:
feeding back a recognition process anomaly when the first comparison result, the second comparison result and the third comparison result are inconsistent; and
adding a corresponding data collection terminal to a blacklist.

8. The method according to claim 2, wherein the biometric data further includes a timestamp, and the method further comprises between the Steps A and B:

decrypting the biometric data to obtain the facial image data and the timestamp; and
feeding back an anomaly when a difference between system time and the timestamp does not meet requirements.

9. The method according to claim 8, further comprising between the Steps A and B:
detecting a facial position in the facial image data when the difference between the system time and the timestamp meets the requirements.

10. The method according to claim 9, further comprising between the Steps A and B:
adding an adversarial sample to the facial position, the adversarial sample devised to counteract action of a negative adversarial sample added at the data collection terminal, wherein the negative adversarial sample is a negative value of the adversarial sample.

11. The method according to claim 10, wherein the adversarial sample is pre-generated by performing a backpropagation training on a generator based on facial recognition results.

12. A biometric recognition server, comprising:
a receiving device to receive biometric data;
a low-bit random filter to randomly transform low-bit data in received biometric data;
a preprocessor to randomly select one or more filtering algorithms to pre-process randomly transformed biometric data; and
an identity verification device to perform identity verification based on pre-processed biometric data.

13. The server according to claim 12, wherein the biometric data comprises facial image data.

14. The server according to claim 12, wherein the low-bit random filter is configured to randomly transform lowest bits in the biometric data.

15. The server according to claim 14, wherein the low-bit random filter is configured to perform any one of the following:
set the lowest bits of the biometric data all to 0;
set the lowest bits of the biometric data all to 1;
randomly re-arrange the lowest bits of the biometric data; and
randomly generate the lowest bits of the biometric data.

16. The server according to claim 12, further comprising:
a recognition decision device to compare multiple recognition results for consistency and to feed back the results when all the recognition results are consistent.

17. The server according to claim 16, wherein the recognition decision device is further configured to:

feed back a recognition process anomaly when the recognition results are inconsistent; and
add the corresponding data collection terminal to a blacklist.

18. The server of claim 13, wherein the biometric data further comprises a timestamp.

19. The server of claim 18, further comprising:
a decryption device to decrypt the biometric data to obtain the facial image data and the timestamp; and
a verification device to feed back an anomaly when a difference between system time and the timestamp does not meet requirements.

20. The server of claim 19, wherein the verification device is further configured to:
detect a facial position in the facial image data when the difference between the system time and the timestamp meets the requirements.

21. The server of claim 20, wherein the verification device is further configured to:
add in the facial position an adversarial sample for counteracting action of a negative adversarial sample added at the data collection terminal, wherein the negative adversarial sample is a negative value of the adversarial sample.

22. The server of claim 21, wherein the adversarial sample is pre-generated by performing a backpropagation training on a generator based on facial recognition results.

23. A client for use in conjunction with the server according to claim 21, comprising:
a facial data collection device to collect facial image data; and
an embedding device to embed a negative adversarial sample in a facial position.

24. The client of claim 23, wherein the embedding device is configured to:
detect the facial position based on the facial image data;
embed the negative adversarial sample in the facial position; and
send a processed facial image and a timestamp in an encrypted manner.

25. A method for generating an adversarial sample using a biometric recognition server as claimed in any one of claims 12 to 22, comprising:
Step A: constructing an adversarial generative network comprising a generator and a discriminator;
Step B: generating noise information with the generator, and adding the noise information to real facial image data to obtain generated facial data;
Step C: recognizing the generated facial data with a facial recognition model;
Step D: performing a backpropagation training on the generator based on recognition results;

Step E: repeating the Steps B to D until the facial recognition model produces an erroneous recognition result; and
Step F: determining the noise information corresponding to the erroneous recognition result as an adversarial sample.

26. A computer storage medium comprising instructions, wherein the instructions, when running, implement the method as claimed in any one of claims 1 to 11.

27. A computer program product comprising a computer program, wherein the computer program, when executed by a processor, implements the method as claimed in any one of claims 1 to 11.