wherein z_i^l denotes features obtained by performing principal component analysis on the left-eye iris feature samples and the right-eye iris feature samples.

The projection matrix calculation module is configured to optimize by the following formula, to obtain the projection matrices P^o and P^1:

$$(P^o, P^1) = \underset{P^o, P^1}{\operatorname{argmax}} \frac{(P^o)^T S_{Z^0 Z^1} P^1}{\sqrt{((P^o)^T S_{Z^0 Z^0} P^0)((P^1)^T S_{Z^1 Z^1} P^1)}}$$

wherein $Z^0 = [z_0^0, z_1^0, \ldots, z_n^0,]$, $Z^1 = [z_0^1, z_1^1, \ldots, z_n^1,]$, $S_{Z^0 Z^0}$and $S_{Z^1 Z^1}$ are within-class variance matrices of Z^0 and Z^1 respectively, and $S_{Z^0 Z^1}$ is a between-class variance matrix of Z^0 and Z^1.

According to a third aspect, the present disclosure provide a non-transitory computer-readable storage medium for multimodal identity authentication, including a memory for storing instructions executable by a processor, which when executed by the processor, implement the steps of the multimodal identity authentication method in the first aspect.

According to a fourth aspect, the present disclosure provide a device for multimodal identity authentication, including at least one processor and a memory storing computer-executable instructions, which when executed by the processor, implement the steps of the multimodal identity authentication method in the first aspect.

A multimodal identity authentication method and apparatus, a storage medium and a device for multimodal identity authentication in the present disclosure have the following beneficial effects:

In the present disclosure, multimodal comparison results are fused by set weights, and identity authentication is performed according to a decision fusion result obtained by fusion. The weight of each modality is determined according to the distribution of comparison scores of a training sample set of the modality and the false rejection rate of the training sample set of the modality. The use of the distribution of comparison scores can increase the distribution difference between comparison scores of same person and comparison scores of different persons, thereby improving the authentication classification effect. The false rejection rate can be used to guide the generation of weights, such that the weights are optimized towards a target direction beneficial to decreasing the false rejection rate. The fused decision fusion result is used for identity authentication, thereby improving the accuracy.

In order to solve the defects in the related art, the present disclosure further provide a multimodal liveness detection method and apparatus, a storage medium and a device for multimodal liveness detection, which improve the effect of face liveness detection.

the accompanying drawings and specific embodiments. Obviously, the embodiments as described are only some rather than all the embodiments of the present disclosure. Generally, assemblies of embodiments of the present disclosure as described and illustrated in the accompanying drawings herein may be arranged and designed in various different configurations. Therefore, the following detailed illustration of embodiments of the present disclosure provided in the accompanying drawings is not intended to limit the claimed scope of the present disclosure, but merely represent selected embodiments of the present disclosure. On the basis of the embodiments in the present disclosure, all other embodiments obtained by a person skilled in the art without any inventive effort shall all fall within the scope of protection of the present disclosure.

In some embodiments of the present disclosure, a multimodal identity authentication method is provided; as shown in Fig. 1, the method includes steps S100 to S400.

S100: two groups of multi-modal biometric data are acquired.

This step is used to acquire two groups of multi-modal biometric data, for identity authentication. For example, two groups, i.e. groups A and B of multi-modal biometric data are acquired, each group biometric data includes multiple modalities, and the multiple modalities may include a face feature and an iris feature, etc. The purpose of authentication is to judge whether the two groups of multi-modal biometric data belong to same person.

In present disclosure, the manner of acquiring the biometric data is not limited, for example, features of a biometric image may be extracted through methods such as image processing, machine learning, or deep learning, etc., so as to obtain eigenvectors, i.e. the biometric data.

S200: same modality biometric data among the two groups of multi-modal biometric data is compared, to obtain inter-group comparison scores of the multiple modalities.

The inter-group comparison scores reflect the similarity between two pieces of same modality biometric data among the two groups of multi-modal biometric data, and the higher the similarity, the greater the possibility that the two pieces of biometric data belong to same person.

In the present disclosure, the manner of calculating the inter-group comparison scores is not limited, for example, Euclidean distance between two pieces of biometric data can be calculated, to obtain the inter-group comparison score.

Inter-group comparison scores of the multiple modalities may be denoted as $s(k)$, where $s(k)$ is an inter-group comparison score of a *kth* modality, $k = 1, 2, ..., m$, and m is the total number of modalities.

excellent the performance. Therefore, a multimodal fusion method needs to be used to decrease the score overlapping portion, and to increase the difference between the two.

S20: the comparison score set of each modality is normalized according to a non-overlapping region of the distribution of the comparison scores of same person of each modality and the distribution of the comparison scores of different persons of each modality.

As described above, the larger the overlapping region of distributions between the comparison scores of same person and the comparison scores of different persons (i.e. the smaller a non-overlapping region), the less obvious the difference between the comparison scores of same person and the comparison scores of different persons; and on the contrary, the higher the distinction degree between the comparison scores of same person and the comparison scores of different persons. Therefore, in the present disclosure, the comparison score set of each modality is normalized by using a non-overlapping region, which can unify the dimensions and distribution of the comparison scores, and increase the distribution difference between comparison scores of same person and comparison scores of different persons, thereby obtaining a better identity authentication effect.

Specific implementations of step S20 may include S21 and S22:

S21: the non-overlapping region of the distribution of the comparison scores of same person of each modality and the distribution of the comparison scores of different persons of each modality is calculated by the following formula, as a normalized dimension of each modality.

$$q(k) = \mu_a(k) - min_a(k) + max_r(k) - \mu_r(k)$$

where $q(k)$ is a normalized dimension of the *kth* modality, $k = 1, 2, ..., m$, m is a total number of modalities, $max_r(k)$ is a maximum value of comparison scores of different persons in a comparison score set of the *kth* modality, $\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the *kth* modality, $min_a(k)$ is a minimum value of comparison scores of same person in the comparison score set of the *kth* modality, and $\mu_a(k)$ is an average value of the comparison scores of same person in the comparison score set of the *kth* modality.

This step is used for performing a normalization operation on comparison scores judged under different modalities, wherein the normalization operation needs to use a parameter of the non-overlapping region in the distribution of comparison scores of a training sample set as a normalized dimension, for performing an operation on the comparison scores. The calculation of the normalized dimension is as shown in the formula above.

The larger $q(k)$, the larger the distribution difference between the comparison scores of same person and the comparison scores of different persons in the *kth* modality, the higher the distinction degree, and the better the effect, such that the normalization on the comparison scores by the normalized dimension can reduce the overlapping region of the distributions, thereby obtaining a better identity authentication effect.

S22: the comparison score set of each modality is normalized according to a normalized dimension of each modality:

$$\hat{s}_i^j(k) = \frac{\left|s_i^j(k) - q(k)\right|}{q(k)}$$

where $s_i^j(k)$ and $\hat{s}_i^j(k)$ are respectively *ith* comparison scores in the comparison score set of the *kth* modality before normalization and after normalization, *i = 1, 2, 3, ..., N_k*, *N_k* is a total number of comparison scores in the comparison score set of the *kth* modality, *j = 0* denotes that the comparison score $s_i^j(k)$ is a comparison score of a different person, and *j = 1* denotes that the comparison score $s_i^j(k)$ is a comparison score of same person.

S30: the weight of each modality is calculated according to a normalized comparison score set of each modality and the false rejection rate of the training sample set of each modality.

In one example, a specific calculation formula of the weight of each modality is as follows:

$$w(k) = \frac{e_k}{\sum_{k=1}^{m} e_k}$$

where *w(k)* is the weight of the *kth* modality;

$$e_k = \frac{\sum_{i=1,2,..,N_k, j=0,1} \hat{s}_i^j(k)}{N_k} + log\frac{1 - FRR_k}{FRR_k}$$

FRR_k is the false rejection rate of the training sample set of the *kth* modality.

Two indexes for evaluating an identity recognition authentication method are respectively a false rejection rate (FRR) and a false accept rate (FAR), in which FRR denotes the rate of biometric data of same person being judged to be biometric data of different persons, and the FAR denotes the rate of biometric data of different persons being judged to be biometric data of same person. In the evaluation of the performance of recognition or authentication, it is expected that regarding the two indexes, the smaller, the better, and if the two indexes are excessively high, a certain service risk will be caused. In a multimodal fusion method, not only the effect of score dimensions needs to be considered, but also the difference between the performance of different modality classifications may reduce the performance of a fused algorithm.

Therefore, in the present disclosure, not only the normalized dimension denoting the distribution of comparison scores is introduced into the calculation of the weight, but also FRR is introduced into the calculation of the weight, and the weight *w(k)* of each modality is calculated by the formula above.

From the content above, it can be determined that in the present disclosure, a method for determining a weight by using the distribution of comparison scores of a training sample set is used, and the use of the distribution of comparison scores can increase the distribution difference between comparison scores of same person and those of different persons, thereby improving the classification effect. In addition, a false rejection rate is introduced when determining the weight, and can be used to guide the generation of weights, such that optimization is performed towards a target

the overlapping portion, the less obvious the distinction between the liveness samples and the spoof samples, and the poorer the performance; on the contrary, the smaller the overlapping portion, the higher the distinction degree between the liveness scores and the spoof scores, and the more excellent the performance. Therefore, a multimodal fusion method needs to be used to decrease the score overlapping portion, and to increase the difference between the two.

S220: the detection score set of each modality is normalized on the basis of a non-overlapping region of a distribution of the liveness scores of each modality and a distribution of the spoof scores of each modality.

From the content above, the larger the overlapping region of the distribution of the liveness scores and the distribution of the spoof scores (i.e. the smaller the non-overlapping region), the less obvious the distinction between the liveness samples and the spoof samples; and on the contrary, the higher the distinction degree between the liveness scores and the spoof scores. Therefore, in the present disclosure, the detection score set of each modality is normalized by using a non-overlapping region, which can unify the dimensions and distribution of the detection scores, and increase the distribution difference between the liveness scores and the spoof scores, thereby obtaining a better liveness detection effect.

This step S220 specifically includes steps S221 and S222.

S221: The non-overlapping region of the distribution of the liveness scores of each modality and the distribution of the spoof scores of each modality is calculated by the following formula, as a normalized dimension of each modality.

$$q_s(k) = \mu_{as}(k) - min_{as}(k) + max_{rs}(k) - \mu_{rs}(k)$$

where $q_s(k)$ is a normalized dimension of the *kth* modality, $k = 1, 2, ..., m$, m is a total number of modalities, $max_{rs}(k)$ is a maximum value of comparison scores of different persons in a comparison score set of the *kth* modality, $\mu_{rs}(k)$ is an average value of the comparison scores of different persons in the comparison score set of the *kth* modality, $min_{as}(k)$ is a minimum value of comparison scores of same person in the comparison score set of the *kth* modality, and $\mu_{as}(k)$ is an average value of the comparison scores of same person in the comparison score set of the *kth* modality.

For fusion of the detection scores, a normalization operation needs to be performed on detection scores judged under different modalities, wherein the normalization operation needs to use a parameter of the non-overlapping region in the distribution of detection scores of a liveness detection training sample set as a normalized dimension, for performing an operation on the detection scores. The calculation of the normalized dimension is as shown in the formula above.

The larger $q_s(k)$, the larger the distribution difference between the detection scores of the liveness scores and the spoof scores in the *kth* modality, the higher the distinction degree, and the better the effect, such that the normalization on the detection scores by the normalized dimension can reduce the overlapping region of the distributions, thereby obtaining a better

liveness detection effect.

S222: the detection score set of each modality is normalized according to the normalized dimension of each modality:

$$\hat{s}_{c_i}^{\ j}(k) = \frac{\left| s_{c_i}^{\ j}(k) - q_s(k) \right|}{q_s(k)}$$

where $s_i^j(k)$ and $\hat{s}_{c_i}^{\ j}(k)$ are respectively *ith* comparison scores in the comparison score set of the *kth* modality before normalization and after normalization, *i = 1, 2, 3, ..., N_k*, *N_k* is a total number of comparison scores in the comparison score set of the *kth* modality, *j = 0* denotes that the comparison score $s_i^j(k)$ is a comparison score of a different person, and *j = 1* denotes that the comparison score $s_i^j(k)$ is a comparison score of same person.

S230: the weight of each modality is calculated according to a normalized detection score set of each modality and the liveness rejection rate of the liveness detection training sample set of each modality.

In one example, a specific calculation formula of the weight of each modality is as follows:

$$w(k) = \frac{E_k}{\sum_{k=1}^{m} E_k}$$

where *w(k)* is the weight of the *kth* modality;

$$E_k = \frac{\sum_{i=1,2,..,N_k, j=0,1} \hat{s}_{c_i}^{\ j}(k)}{M_k} + log\frac{1-LRR_k}{LRR_k},$$

LRR_k is a liveness rejection rate of a liveness detection training sample set of the *kth* modality.

Two indexes for evaluating the liveness detection performance are a liveness rejection rate (LRR) and a prosthesis pass rate (PPR), respectively. LRR denotes the rate of a liveness sample being judged as a spoof sample, and is an evaluation for misjudging the liveness sample; and PPR denotes the rate of a spoof sample being judged as a liveness sample, and is the proportion of misjudging a spoof. In the performance evaluation of liveness detection, it is expected that regarding the two indexes, the smaller, the better. In the multimodal liveness detection method of the present disclosure, not only the effect of the detection score dimensions needs to be considered, but also the differences in the liveness detection performance of different modalities need to be considered, which may also reduce the performance of the algorithm after fusion.

Therefore, in the present disclosure, not only the normalized dimension denoting the distribution of detection scores is introduced into the calculation of the weight, but also LRR is introduced into the calculation of the weight, and the weight *w(k)* of each modality is calculated by the formula above.

In the present disclosure, there may be multiple situations of the multi-modal facial data, which are illustrated by two specific examples below.

a magnetic tape, a magnetic core memory, a magnetic bubble memory and a USB flash drive; and an apparatus that stores information in an optically manner, such as a CD or DVD. Of course, there are other forms of readable storage media, such as quantum memory, graphene memory, etc.

According to the illustration of the embodiments of the multimodal liveness detection method, the storage medium can further include other embodiments. The implementation principle and technical effects of this embodiment are the same as those of the embodiments of the multimodal liveness detection method, and for details, reference may be made to the illustration of the embodiments of the multimodal liveness detection method, and details will not be repeated herein again.

In the present disclosure, further provided is a device for multimodal liveness detection. The device may be a separate computer, and may also include an actual operation apparatus or the like using one or more of the method of the present description or the apparatus of one or more of embodiments. The device for multimodal liveness detection may include at least one processor and a memory storing computer-executable instructions, which when executed by the processor, implement the steps of any one or more embodiments of the multimodal liveness detection method.

According to the illustration of the embodiments of the multimodal liveness detection method, the device for multimodal liveness detection can further include other embodiments. The implementation principle and technical effects of this embodiment are the same as those of the embodiments of the multimodal liveness detection method, and for details, reference may be made to the illustration of the embodiments of the multimodal liveness detection method, and details will not be repeated herein again.

In some embodiments of the present disclosure, a computer device may be a terminal, and the internal structure thereof may be as shown in Fig. 7. The computer device includes a processor, a memory, a communication interface, a display screen, and an input apparatus which are connected by a system bus. The processor of the computer device is used to provide computing and control capabilities. The memory of the computer device includes a non-transitory storage medium and an internal memory. The non-transitory storage medium stores an operating system and a computer program. The internal memory provides an environment for operation of the operating system and the computer program in the non-transitory storage medium. The communication interface of the computer device is used for communicating with an external terminal in a wired or wireless manner, and the wireless manner can be realized by Wi-Fi, a mobile cellular network, NFC (Near Field Communication) or other techniques. The computer program implements a multimodal liveness detection method when executed by the processor. The display screen of the computer device may

Claims

1. A multimodal identity authentication method, the method comprising:

acquiring two groups of multi-modal biometric data;

comparing same modality biometric data among the two groups of multi-modal biometric data, to obtain inter-group comparison scores of the multiple modalities;

determining a comparison result of each modality according to the inter-group comparison score of each modality and a comparison threshold set for each modality, and performing a weighted voting on the comparison result of each modality according to a weight set for each modality to obtain a decision fusion result,

wherein the weight of each modality is determined according to a distribution of comparison scores of a training sample set of the modality and a false rejection rate of the training sample set of the modality; and

judging whether the two groups of multi-modal biometric data belong to same person according to the decision fusion result.

2. The multimodal identity authentication method according to claim 1, wherein a training process of determining the weight of each modality comprises:

performing pairwise comparison on biometric samples of the training sample set of each modality to obtain a comparison score set of each modality,

wherein the comparison score set comprises comparison scores of same person and comparison scores of different persons;

normalizing the comparison score set of each modality according to a non-overlapping region of a distribution of the comparison scores of same person of each modality and a distribution of the comparison scores of different persons of each modality; and

calculating the weight of each modality according to a normalized comparison score set of each modality and the false rejection rate of the training sample set of each modality.

3. The multimodal identity authentication method according to claim 2, wherein the normalizing the comparison score set of each modality according to a non-overlapping region of a distribution of the comparison scores of same person of each modality and a distribution of the comparison scores of different persons of each modality, comprises:

calculating the non-overlapping region of the distribution of the comparison scores of same

person of each modality and the distribution of the comparison scores of different persons of each modality as a normalized dimension of each modality:

$$q(k) = \mu_a(k) - min_a(k) + max_r(k) - \mu_r(k)$$

wherein *q(k)* is a normalized dimension of a *kth* modality, *k = 1, 2, ..., m, m* is a total number of modalities, $max_r(k)$ is a maximum value of comparison scores of different persons in a comparison score set of the *kth* modality, $\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the *kth* modality, $min_a(k)$ is a minimum value of comparison scores of same person in the comparison score set of the *kth* modality, and $\mu_a(k)$ is an average value of the comparison scores of same person in the comparison score set of the *kth* modality;

normalizing the comparison score set of each modality according to the normalized dimension of each modality:

$$\hat{s}_i^j(k) = \frac{\left| s_i^j(k) - q(k) \right|}{q(k)}$$

wherein $s_i^j(k)$ and $\hat{s}_i^j(k)$ are respectively *ith* comparison scores in the comparison score set of the *kth* modality before normalization and after normalization, *i = 1, 2, 3, ..., N_k*, N_k is a total number of comparison scores in the comparison score set of the *kth* modality, *j = 0* denotes that the comparison score $s_i^j(k)$ is a comparison score of a different person, and *j = 1* denotes that the comparison score $s_i^j(k)$ is a comparison score of same person; and

the calculating the weight of each modality according to a normalized comparison score set of each modality and the false rejection rate of the training sample set of each modality, comprises:

calculating the weight of each modality as:

$$w(k) = \frac{e_k}{\sum_{k=1}^{m} e_k}$$

wherein *w(k)* is the weight of the *kth* modality;

$$e_k = \frac{\sum_{i=1,2,..,N_k, j=0,1} \hat{s}_i^j(k)}{N_k} + log\frac{1 - FRR_k}{FRR_k}$$

FRR_k is the false rejection rate of the training sample set of the *kth* modality.

4. The multimodal identity authentication method according to claim 3, wherein the decision fusion result is calculated as:

$$\delta = \sum_{i=1}^{m} w(k) \cdot \delta(k)$$

wherein δ is the decision fusion result, and $\delta(k)$ is a comparison result of k modalities:

$$\delta(k) = \begin{cases} 0 & s(k) > T(k) \\ 1 & s(k) \leq T(k) \end{cases}$$

$s(k)$ is an inter-group comparison score of the *kth* modality, and $T(k)$ is a comparison threshold of the *kth* modality.

5. The multimodal identity authentication method according to any one of claims 1-4, wherein the multi-modal biometric data comprises a face feature and an iris feature, wherein the iris feature is obtained by fusing a left-eye iris feature and a right-eye iris feature.

6. The multimodal identity authentication method according to claim 5, wherein fusing the left-eye iris feature and the right-eye iris feature to obtain the iris feature, comprises:

performing principal component analysis on the left-eye iris feature and the right-eye iris feature, to obtain:

$$z^l = {W^l}^T \cdot (x^l - \mu^l), l = 0,1$$

wherein l = *0,1* denotes a left eye and a right eye, respectively, x^l denotes the left-eye iris features or the right-eye iris features, z^l denotes features obtained by performing principal component analysis on the left-eye iris features or the right-eye iris features, μ^l denotes an average value of the left-eye iris features and an average value of the right-eye iris features, W^l denotes a principal component analysis matrix of the left-eye iris features and a principal component analysis matrix of the right-eye iris features, ${W^l}^T$ denotes a transposed matrix of W^l; and μ^l is obtained by training a left-eye iris features training set or a right-eye iris features training set, and W^l is obtained by training a left-eye iris features training set or a right-eye iris features training set; and

fusing features obtained by performing principal component analysis on the left-eye iris feature and the right-eye iris feature to obtain the iris feature as:

$$Z = \begin{pmatrix} P^o \\ P^1 \end{pmatrix}^T \begin{pmatrix} z^0 \\ z^1 \end{pmatrix}$$

wherein Z is the iris feature, P^o and P^1 are projection matrices, and P^o and P^1 are obtained by training the left-eye iris features training set and the right-eye iris features training set.

7. The multimodal identity authentication method according to claim 6, wherein obtaining the μ^l, W^l, P^o and P^1 by training, comprises:

calculating the average value of the left-eye iris features and the average value of the right-eye iris features as:

$$\mu^l = \frac{1}{n}\sum_{i=1}^{n} x_i^l, \qquad l = 0,1$$

wherein x_i^l is an *ith* left-eye iris feature sample in the left-eye iris features training set or an *ith* right-eye iris feature sample in the right-eye iris features training set, *i = 1, 2, 3, ..., n*, *n* is a total number of left-eye iris feature samples or right-eye iris feature samples;

calculating a covariance matrix of the left-eye iris features samples and the right-eye iris feature samples as:

$$S^l = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^l - \mu^l\right)(x_i^l - \mu^l)^T$$

wherein S^l is the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples, and

decomposing eigenvalues of the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples as:

$$S^l \cdot p_t^l = \lambda_t^l \cdot p_t^l$$

wherein λ_t^l and p_t^l are respectively S^l eigenvalues and eigenvectors, *t = 0, 1, 2, …, d*, *d* is a number of rows and a number of columns of S^l;

selecting the first *r* eigenvectors accounting for 95% according to a percentage φ^l of principal components to constitute the principal component analysis matrix W^l,

wherein$\varphi^l = \frac{\sum_{i=1}^{r}\lambda_i^l}{\sum_{i=1}^{d}\lambda_i^l}, W^l = [p_0^l,\ p_1^l, \dots, p_r^l]$;

performing principal component analysis on the left-eye iris feature samples and the right-eye iris feature samples, to obtain:

$$z_i^l = {W^l}^T \cdot x_i^l, l = 0,1$$

wherein z_i^l denotes features obtained by performing principal component analysis on the left-eye iris feature samples and the right-eye iris feature samples; and

optimizing to obtain the projection matrices P^o and P^1 as:

$$(P^o, P^1) = \underset{P^o,P^1}{\operatorname{argmax}} \frac{(P^o)^T S_{Z^0Z^1} P^1}{\sqrt{((P^o)^T S_{Z^0Z^0} P^0)((P^1)^T S_{Z^1Z^1} P^1)}}$$

wherein $Z^0 = [z_0^0, z_1^0, \dots, z_n^0,]$, $Z^1 = [z_0^1, z_1^1, \dots, z_n^1,]$, $S_{Z^0Z^0}$ and $S_{Z^1Z^1}$ are within-class variance matrices of Z^0 and Z^1 respectively, and $S_{Z^0Z^1}$ is a between-class variance matrix of Z^0

and Z^1.

8. A multimodal identity authentication apparatus, the apparatus comprising:

a data acquisition module, configured to acquire two groups of multi-modal biometric data;

a comparison score calculation module, configured to compare same modality biometric data among the two groups of multi-modal biometric data, to obtain inter-group comparison scores of the multiple modalities;

a decision fusion module, configured to determine a comparison result of each modality according to the inter-group comparison score of each modality and a comparison threshold set for each modality, and perform weighted voting on the comparison result of each modality according to a weight set for each modality to obtain a decision fusion result,

wherein the weight of each modality is determined according to a distribution of comparison scores of a training sample set of the modality and a false rejection rate of the training sample set of the modality; and

a judgment module, configured to judge whether the two groups of multi-modal biometric data belong to same person according to the decision fusion result.

9. A non-transitory computer-readable storage medium for multimodal identity authentication, comprising a memory for storing instructions executable by a processor, which when executed by the processor, implement the steps of the multimodal identity authentication method according to any one of claims 1-7.

10. A device for multimodal identity authentication, a device for multimodal identity authentication, comprising at least one processor and a memory storing computer-executable instructions, which when executed by the processor, implement the steps of the multimodal identity authentication method according to any one of claims 1-7.

11. A multimodal liveness detection method, the method comprising:

acquiring multi-modal facial data of a same user.

performing liveness detection on the multi-modal facial data of the same user, so as to obtain detection scores of the multiple modalities.

performing weighted summation on the detection scores of the multiple modalities according to a weight set for each modality, to obtain a fusion score;

wherein the weight of each modality is determined according to a distribution of detection scores

of a liveness detection training sample set of the modality and a liveness rejection rate of the liveness detection training sample set of the modality;

determining a liveness detection result of each modality according to the detection scores of each modality and a liveness detection threshold set for each modality, and performing weighted voting on the liveness detection result of each modality according to the weight set for each modality, so as to obtain a liveness detection decision fusion result.

judging whether the facial data belongs to a live person or a spoof according to at least one of the fusion score and the liveness detection decision fusion result.

12. The multimodal liveness detection method according to claim 11, wherein a training process of determining the weight of each modality comprises:

performing liveness detection on the liveness detection training sample set of each modality, to obtain a detection score set of each modality;

wherein the liveness detection training sample set comprises liveness samples and spoof samples, and the detection score set comprises liveness scores and spoof scores;

normalizing the detection score set of each modality on the basis of a non-overlapping region of a distribution of the liveness scores of each modality and a distribution of the spoof scores of each modality; and

calculating the weight of each modality according to a normalized detection score set of each modality and the liveness rejection rate of the liveness detection training sample set of each modality.

13. The multimodal liveness detection method according to claim 12, wherein the normalizing the detection score set of each modality on the basis of a non-overlapping region of a distribution of the liveness scores of each modality and a distribution of the spoof scores of each modality, comprises:

calculating the non-overlapping region of the distribution of the liveness scores of each modality and the distribution of the spoof scores of each modality as a normalized dimension of each modality:

$q_s(k) = \mu_{as}(k) - min_{as}(k) + max_{rs}(k) - \mu_{rs}(k)$, wherein $q_s(k)$ is a normalized dimension of a *kth* modality, $k = 1, 2, ..., m$, m is a total number of modalities, $max_{rs}(k)$ is a maximum value of liveness scores in a detection score set of the *kth* modality, $\mu_{rs}(k)$ is an average value of the liveness scores in the detection score set of the *kth* modality, $min_{as}(k)$ is a minimum value of spoof scores in the detection score set of the *kth* modality, and $\mu_{as}(k)$ is an average value of the

spoof scores in the detection score set of the *kth* modality;

and normalizing the detection score set of each modality according to the normalized dimension of each modality:

$\hat{s}_{c_i}^{\ j}(k) = \frac{\left|s_{c_i}^{\ j}(k) - q_s(k)\right|}{q_s(k)}$, wherein $s_{c_i}^{\ j}(k)$ and $\hat{s}_{c_i}^{\ j}(k)$ are respectively *ith* detection scores in the detection score set of the *kth* modality before normalization and after normalization, *i = 1, 2, 3, ..., M_k*, *M_k* is a total number of detection scores in the detection score set of the *kth* modality, *j = 0* denotes that the detection score $s_{c_i}^{\ j}(k)$ is a spoof score, and *j = 1* denotes that the detection score $s_{c_i}^{\ j}(k)$ is a liveness score; and

calculating the weight of each modality according to a normalized detection score set of each modality and the liveness rejection rate of the liveness detection training sample set of each modality, comprises:

calculating the weight of each modality as:

$$w(k) = \frac{E_k}{\sum_{k=1}^{m} E_k}$$

wherein *w(k)* is the weight of the *kth* modality;

$E_k = \frac{\sum_{i=1,2,..,N_k, j=0,1} \hat{s}_{c_i}^{\ j}(k)}{M_k} + log\frac{1-LRR_k}{LRR_k}$, *$LRR_k$* is a liveness rejection rate of a liveness detection training sample set of the *kth* modality.

14. The multimodal liveness detection method according to claim 13, wherein the fusion score is calculated as:

$s_c = \sum_{i=1}^{m} w(k) \cdot s_c(k)$, wherein s_c is the fusion score, and $s_c(k)$ is a detection score of the *kth* modality;

and the decision fusion result is calculated as:

$\delta_f = \sum_{i=1}^{m} w(k) \cdot \delta_d(k)$, where δ_f is the liveness detection decision fusion result, and $\delta_d(k)$ is a liveness detection result of *k* modalities;

$\delta_d(k) = \begin{cases} 0 & s_c(k) > T_d(k) \\ 1 & s_c(k) \leq T_d(k) \end{cases}$, $T_d(k)$ is a liveness detection threshold of the *kth* modality.

15. The multimodal liveness detection method according to any one of claims 11-14, wherein judging whether the facial data belongs to a live person or a spoof according to at least one of the fusion score and the liveness detection decision fusion result, comprises:

judging that the facial data belongs to a live person if the fusion score is greater than a set first

threshold, and if the fusion score is less than or equal to the set first threshold, judging that the facial data belongs to a spoof; or

judging that the facial data belongs to a live person if the liveness detection decision fusion result is greater than a set second threshold, and if the liveness detection decision fusion result is less than or equal to the set second threshold, judging that the facial data belongs to a spoof; or

judging that the facial data belongs to a live person if the fusion score is greater than the set first threshold and the liveness detection decision fusion result is greater than the set second threshold, if the fusion score is less than or equal to the set first threshold, judging that the facial data belongs to a spoof, and if the liveness detection decision fusion result is less than or equal to the set second threshold, judging that the facial data belongs to a spoof.

16. The multimodal liveness detection method according to any one of claims 11-14, wherein the multi-modal facial data comprises face images under different types of light sources; or the multi-modal facial data comprises multiple different regions of a face image.

17. The multimodal liveness detection method according to claim 16, wherein the multiple different regions of a face image comprise a whole region and local regions of the face image, and the face image is a visible-light face image or a near-infrared face image.

18. A multimodal liveness detection apparatus, the apparatus comprising:

a data acquisition module, configured to acquire multi-modal facial data of a same user;

a detection score calculation module, configured to perform liveness detection on the multi-modal facial data of the same user, so as to obtain detection scores of the multiple modalities;

a fusion score calculation module, configured to perform weighted summation on the detection scores of the multiple modalities according to a weight set for each modality, to obtain a fusion score;

wherein the weight of each modality is determined according to a distribution of detection scores of a liveness detection training sample set of the modality and a liveness rejection rate of the liveness detection training sample set of the modality;

a decision fusion result calculation module, configured to determine a liveness detection result of each modality according to the detection scores of each modality and a liveness detection threshold set for each modality, and perform weighted voting on the liveness detection result of each modality according to the weight set for each modality, so as to obtain a liveness detection decision fusion result; and

a liveness judgment module, configured to judge whether the facial data belongs to a live person or a spoof according to at least one of the fusion score and the liveness detection decision fusion result.

19. A non-transitory computer-readable storage medium for multimodal liveness detection, comprising a memory for storing instructions executable by a processor, which when executed by the processor, implement the steps of the multimodal liveness detection method according to any one of claims 11-17.

20. A device for multimodal liveness detection, comprising at least one processor and a memory storing computer-executable instructions, which when executed by the processor, implement the steps of the multimodal liveness detection method according to any one of claims 11-17.