

Multimodal Identity Authentication Method and Apparatus, Storage Medium, and Device for Multimodal Identity Authentication

Cross-Reference to Related Application

The present disclosure claims priority to Chinese Patent Application No. 202210100246.0, filed to the China National Intellectual Property Administration on January 27, 2022 and entitled "Multimodal Identity Authentication Method and Apparatus, Storage Medium, and Device", and Chinese Patent Application No. 202210100939.X, filed to the China National Intellectual Property Administration on January 27, 2022 and entitled "Multimodal Liveness Detection Method and Apparatus, Storage Medium, and Device", which are incorporated herein by reference in their entireties.

Technical Field

The present disclosure relates to the field of biometric recognition, and in particular to a multimodal identity authentication method and apparatus, a storage medium, and a device for multimodal identity authentication, and a multimodal liveness detection method and apparatus, a storage medium, and a device for multimodal liveness detection.

Background

Currently, biometric features commonly used for identity authentication include face, fingerprint, iris, finger vein, palm print, voiceprint, gait, and the like. Due to the continuous development of technology, single biometric feature recognition has reached a high precision, for example, face recognition has reached a precision of 99.8% on a common LFW (Labeled Faces in the Wild) test library. However, in practical applications, the precision of the single biometric feature recognition will be decreased due to some inherent limitations, for example, face recognition will be affected by illumination, posture, makeup, etc., fingerprint recognition will be affected by wearing, desquamation, etc., and voiceprint recognition will be affected by having a cold and hoarse voice, etc. Existing research has shown that, by using complementary information between different biometric features to perform multimodal biometric recognition, the disadvantage of a single kind of biometric feature can be eliminated, thereby helping to improve the recognition accuracy and security.

Data information fusion needs to be performed in the multimodal biometric recognition, and decision fusion is a common fusion manner in the relevant art. The decision fusion includes: biometric feature comparison is performed by multiple modalities to obtain comparison results of the

multiple modalities; and then the comparison results of the multiple modalities are fused to obtain a final comparison result.

Due to the difference between authentication methods of different modalities, different modalities have different contributions to a final authentication result, and therefore when the comparison results of different modalities are subjected to decision fusion, the weights of the comparison results of different modalities should be different. In the related art, when weights of different modalities are determined, they are generally set according to experience, causing the settings of the weights to be inaccurate, thereby reducing the accuracy of multimodal identity authentication.

In recent years, face recognition is highly favored by people due to advantages such as convenient acquisition and non-contact, and has been widely used in fields such as finance and security. However, just because of the easy acquisition thereof, it is easy to be used by other people, a spoof face is produced by manners such as printing a picture or recording a video, and a face recognition system is attacked, such that a person passes the face recognition system as an impostor. Thus, face image-based liveness detection is extremely important and is the prerequisite and guarantee of face recognition.

Currently, there are two types of common face liveness detection technologies: user-cooperative liveness detection and silent liveness detection. The user-cooperative liveness detection means that a liveness detection system randomly sends a plurality of facial action instructions, and requires a user to make sequential response actions within a specified period of time, and the system provides a judgment as to whether the face is a liveness face according to a judgment as to whether the user's response actions are accurate. The silent liveness detection means that a user completes a liveness detection task without perception. The silent face liveness detection method mainly includes several methods such as a method based on traditional statistical learning, a 3D-based method, and a near-infrared-based method, etc.

Summary

To solve the defects in the related art, the present disclosure provide a multimodal identity authentication method and apparatus, a storage medium, and a device for multimodal identity authentication, which improve the performance of identity authentication.

The present disclosure provide the following technical solutions:

according to a first aspect, the present disclosure provide a multimodal identity authentication method, the method includes:

two groups of multi-modal biometric data are acquired;

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

a comparison result of each modality is determined according to the inter-group comparison score of each modality and a comparison threshold set for each modality, and a weighted voting is performed 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

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

In some embodiments, a training process of determining the weight of each modality includes:

pairwise comparison is performed 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 includes comparison scores of same person and comparison scores of different persons;

the comparison score set of each modality is normalized 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

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 some embodiments, the comparison score set of each modality is normalized 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, includes:

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 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 k th modality, $k = 1, 2, \dots, m$, m is the total number

of modalities, $\max_r(k)$ is a maximum value of comparison scores of different persons in a comparison score set of the k th modality, $\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the k th modality, $\min_a(k)$ is a minimum value of comparison scores of same person in the comparison score set of the k th modality, and $\mu_a(k)$ is an average value of the comparison scores of same person in the comparison score set of the k th modality;

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

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

wherein $s_i^j(k)$ and $\hat{s}_i^j(k)$ are respectively i th comparison scores in the comparison score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, N_k$, N_k is a total number of comparison scores in the comparison score set of the k th 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 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, includes: the weight of each modality is calculated as:

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

wherein $w(k)$ is the weight of the k th modality;

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

FRR_k is the false rejection rate of the training sample set of the k th modality.

In some embodiments, 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 k th modality, and $T(k)$ is a comparison threshold of the k th modality.

In some embodiments, whether the two groups of multi-modal biometric data belong to same person is judged according to the decision fusion result.

if the decision fusion result is greater than a set decision threshold, it is judged that the two groups of multi-modal biometric data belong to same person; otherwise, it is judged that the two groups of multi-modal biometric data do not belong to same person.

In some embodiments, the multi-modal biometric data includes 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.

In some embodiments, the left-eye iris feature and the right-eye iris feature are fused to obtain the iris feature, includes:

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

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

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^{lT} 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;

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

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

wherein Z is the iris feature, P^0 and P^1 are projection matrices, and P^0 and P^1 are obtained by training the left-eye iris features training set and the right-eye iris features training set.

In some embodiments, μ^l , W^l , P^0 and P^1 are obtained by training, includes:

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

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

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

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

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

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

eigenvalues of the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples are decomposed 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, $t = 0, 1, 2, \dots, d$, d is a number of rows and a number of columns of S^l ;

the first r eigenvectors accounting for 95% are selected according to a percentage φ^l of principal components to constitute the principal component analysis matrix W^l ;

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

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

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

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;

by optimization, the projection matrices P^0 and P^1 are obtained as:

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

wherein $Z^0 = [z_0^0, z_1^0, \dots, z_n^0]$, $Z^1 = [z_0^1, z_1^1, \dots, z_n^1]$, $S_{Z^0 Z^0}$ 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 second aspect, the present disclosure provide a multimodal identity authentication apparatus, the apparatus includes a data acquisition module, a comparison score

calculation module, a decision fusion module and a judgment module.

The data acquisition module is configured to acquire two groups of multi-modal biometric data.

The comparison score calculation module is 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.

The decision fusion module is 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.

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

In some embodiments, the weight of each modality is determined by a training data preparation module, a normalization module and a weight calculation module.

The training data preparation module is configured to perform 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 includes comparison scores of same person and comparison scores of different persons.

The normalization module is configured to normalize 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.

The weight calculation module is configured to calculate 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.

In some embodiments, the normalization module includes a normalized-dimension calculation unit and a normalization unit.

The normalized-dimension calculation unit is configured to calculate, by the following formula, the 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 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 k th modality, $k = 1, 2, \dots, m$, 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 k th modality, $\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the k th modality, $\min_a(k)$ is a minimum value of comparison scores of same person in the comparison score set of the k th modality, and $\mu_a(k)$ is an average value of the comparison scores of same person in the comparison score set of the k th modality.

The normalization unit is configured to normalize the comparison score set of each modality according to the normalized dimension of each modality:

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

wherein $s_i^j(k)$ and $\hat{s}_i^j(k)$ are respectively i th comparison scores in the comparison score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, N_k$, N_k is a total number of comparison scores in the comparison score set of the k th 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.

The weight calculation module is configured to calculate the weight of each modality by the following formula:

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

wherein $w(k)$ is the weight of the k th modality:

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

FRR_k is the false rejection rate of the training sample set of the k th modality.

In some embodiments, the decision fusion result is calculated by the following formula:

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

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 k th modality, and $T(k)$ is a comparison threshold of the k th modality.

In some embodiments, the judgment module is further configured to judge that the two groups of multi-modal biometric data belong to same person if the decision fusion result is greater than a set decision threshold; otherwise, judge that the two groups of multi-modal biometric data do not belong to same person.

In some embodiments, the multi-modal biometric data includes 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.

In some embodiments, the left-eye iris feature and the right-eye iris feature are fused to obtain the iris feature by a first principal component analysis module and a feature fusion module.

The first principal component analysis module is configured to perform principal component analysis on the left-eye iris feature and the right-eye iris feature, to obtain:

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

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^{lT} 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.

The feature fusion module is configured to fuse the features obtained by performing principal component analysis on the left-eye iris feature and the right-eye iris feature by the following formula, to obtain the iris feature:

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

wherein Z is the iris feature, P^0 and P^1 are projection matrices, and P^0 and P^1 are obtained by training the left-eye iris features training set and the right-eye iris features training set.

In some embodiments, the μ^l , W^l , P^0 and P^1 are obtained by training by an average value

calculation module, a covariance matrix calculation module, an eigenvalue decomposition module, a principal component analysis matrix acquisition module, a second principal component analysis module and a projection matrix calculation module.

The average value calculation module is configured to calculate the average value of the left-eye iris features and the average value of the right-eye iris features by the following formula:

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

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

the covariance matrix calculation module is configured to calculate a covariance matrix of the left-eye iris features samples and the right-eye iris feature samples by the following formula:

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

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

the eigenvalue decomposition module is configured to decompose eigenvalues of the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples by the following formula:

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

wherein λ_t^l and p_t^l are respectively eigenvalues and eigenvectors of a covariance matrix S^l , $t = 0, 1, 2, \dots, d$, d is a number of rows and a number of columns of the covariance matrix S^l .

The principal component analysis matrix acquisition module is configured to select 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 ;

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

The second principal component analysis module is configured to perform principal component analysis on the left-eye iris feature samples and the right-eye iris feature samples by the following formula, to obtain:

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

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.

The projection matrix calculation module is configured to optimize by the following formula, to obtain the projection matrices P^0 and P^1 :

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

wherein $Z^0 = [z_0^0, z_1^0, \dots, z_n^0]$, $Z^1 = [z_0^1, z_1^1, \dots, z_n^1]$, $S_{Z^0 Z^0}$ 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 liveness detection 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.

According to a fifth aspect, some embodiments of the present disclosure provide a multimodal liveness detection method, the method includes:

multi-modal facial data of a same user is acquired;

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

weighted summation is performed 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 liveness detection result of each modality is determined according to the detection scores of each modality and a liveness detection threshold set for each modality, and weighted voting is performed 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

it is judged 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.

In some embodiments, a training process of determining the weight of each modality includes:

liveness detection is performed 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 includes liveness samples and spoof samples, and the detection score set includes liveness scores and spoof scores;

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; and

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 some embodiments, 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, includes:

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 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 k th modality, $k = 1, 2, \dots, m$, m is a total number of modalities, $\max_{rs}(k)$ is a maximum value of liveness scores in a detection score set of the k th modality, $\mu_{rs}(k)$ is an average value of the liveness scores in the detection score set of the k th modality, $\min_{as}(k)$ is a minimum value of spoof scores in the detection score set of the k th modality, and $\mu_{as}(k)$ is an average value of the spoof scores in the detection score set of the k th modality;

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

$\hat{s}_{c_i}^j(k) = \frac{|s_{c_i}^j(k) - q_s(k)|}{q_s(k)}$, wherein $s_{c_i}^j(k)$ and $\hat{s}_{c_i}^j(k)$ are respectively i th detection scores in the detection score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, M_k$, M_k is a total number of detection scores in the detection score set of the k th 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.

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, includes:

the weight of each modality is calculated as:

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

wherein $w(k)$ is the weight of the k th modality;

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

In some embodiments, 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 k th 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;

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

In some embodiments, it is judged 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, includes:

if the fusion score is greater than a set first threshold, it is judged that the facial data belongs to a live person, otherwise, it is judged that the facial data belongs to a spoof; or

if the liveness detection decision fusion result is greater than a set second threshold, it is judged that the facial data belongs to a live person, otherwise, it is judged that the facial data belongs to a spoof; or

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, it is judged that the facial data belongs to a live person, otherwise, it is judged that the facial data belongs to a spoof.

In some embodiments, the multi-modal facial data includes face images under different types of light sources.

In some embodiments, the multi-modal facial data includes multiple different regions of a face image.

In some embodiments, the multiple different regions of each face image include 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.

According to a sixth aspect, the present disclosure provide a multimodal liveness detection apparatus, the apparatus includes a data acquisition module, a detection score calculation module, a fusion score calculation module, a decision fusion result calculation module and a liveness judgment module.

The data acquisition module is configured to acquire multi-modal facial data of a same user.

The detection score calculation module is 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;

The fusion score calculation module is 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.

The decision fusion result calculation module is 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.

The liveness judgment module is 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.

In some embodiments, the weight of each modality is determined by a training data preparation module, a normalization module and a weight calculation module.

The training data preparation module is configured to perform 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 includes liveness samples and spoof samples, and the detection score set includes liveness scores and spoof scores.

The normalization module is configured to normalize 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.

The weight calculation module is configured to calculate the weight of each modality according to the normalized detection score set of each modality and the liveness rejection rate of the liveness detection training sample set of each modality.

In some embodiments, the normalization module may include a normalized-dimension calculation unit and a normalization unit.

The normalized-dimension calculation unit is configured to calculate a non-overlapping region of a distribution of the liveness scores of each modality and a distribution of the spoof scores of each modality by the following formula, as a normalized dimension of each modality:

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

The normalization unit is configured to normalize the detection score set of each modality according to the normalized dimension of each modality:

$\hat{s}_{c_i}^j(k) = \frac{|s_{c_i}^j(k) - q_s(k)|}{q_s(k)}$, wherein $s_{c_i}^j(k)$ and $\hat{s}_{c_i}^j(k)$ are respectively i th detection scores in the detection score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, M_k$, M_k is a total number of detection scores in the detection score set of the k th 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.

The weight calculation module is configured to calculate the weight of each modality by the following formula:

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

wherein $w(k)$ is the weight of the k th modality;

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

In some embodiments, the fusion score is calculated by the following formula:

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

The decision fusion result is calculated by the following formula:

$\delta_f = \sum_{i=1}^m w(k) \cdot \delta_d(k)$, wherein 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 k th modality.

In some embodiments, the liveness judgment module is configured to

judge that the facial data belongs to a live person if the fusion score is greater than a set first threshold, and otherwise, judge that the facial data belongs to a spoof; or

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

judge 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, and otherwise, judge that the facial data belongs to a spoof.

In some embodiments, the multi-modal facial data includes face images under different types of light sources.

In some embodiments, the multi-modal facial data includes multiple different regions of each face image.

In some embodiments, the multiple different regions of each face image include 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.

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

According to an eighth aspect, the present disclosure provide a device for multimodal liveness detection, including a 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 in the fifth aspect.

The multimodal liveness detection method and apparatus, the storage medium, and the device for multimodal liveness detection in the present disclosure have the following beneficial effects:

1. In the related art, a single modality is used for face liveness detection; whereas in the present disclosure, multi-modal facial data is used for multimodal face liveness detection, thereby increasing the amount of information available for face liveness detection, and achieving a better liveness detection effect.

2. Detection scores of multiple modalities are fused by weight, and weighted voting is performed

on liveness detection results of the multiple modalities, that is, not only detection scores are fused, but also multimodal decision fusion is performed. Furthermore, the fused weight is related to the distribution of the detection scores and the liveness rejection rate of each modality, and the detection scores and decisions are subjected to multimodal fusion by using the distribution of the detection scores and the liveness rejection rate, and the dimensions and distribution of the detection scores are unified by weights, thereby reducing the adverse effects of differences in the performance of liveness detection of different modalities on liveness detection after fusion, such that the multimodal liveness detection method is more effective and feasible, and a better multimodal liveness detection effect can be obtained.

3. The detection score fusion and decision fusion performed in the present disclosure are both simple numerical calculation problems, do not introduce additional time overhead, and do not affect the overall speed of a liveness detection algorithm.

Brief Description of the Drawings

Fig. 1 is a flowchart of a multimodal identity authentication method according to some embodiments of the present disclosure;

Fig. 2 is a schematic diagram of distribution of comparison scores of same person and comparison scores of different persons in a k th modality;

Fig. 3 is a schematic diagram of a multimodal identity authentication apparatus according to some embodiments of the present disclosure;

Fig. 4 is a flowchart of a multimodal liveness detection method according to some embodiments of the present disclosure;

Fig. 5 is a schematic diagram of distribution of spoof scores and liveness scores in a k th modality;

Fig. 6 is a schematic diagram of a multimodal liveness detection apparatus according to some embodiments of the present disclosure; and

Fig. 7 is an internal structural diagram of a computer device according to some embodiments of the present disclosure.

Detailed Description of the Embodiments

In order to make the technical problems to be solved, the technical solutions and advantages of some embodiments of the present disclosure clearer, hereinafter, the technical solutions in some embodiments of the present disclosure will be described clearly and thoroughly in combination with

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 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 k th modality, $k = 1, 2, \dots, m$, and m is the total number of modalities.

S300: a comparison result of each modality is determined according to the inter-group comparison score of each modality and a comparison threshold set for each modality, and a weighted voting is performed on the comparison result of each modality according to a weight set for each modality to obtain a decision fusion result.

The decision fusion result is a result of performing weighted voting on comparison results of the multiple modalities, and the weight of each modality reflects the effect achieved by the comparison result of each modality in the final authentication, and the proportion of the comparison result of each modality in the final authentication.

Exemplarily, the decision fusion result may be calculated by the following formula:

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

where δ is the decision fusion result, $w(k)$ is the weight of the k th modality, and $\delta(k)$ is a comparison result of the k th modality:

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

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

Taking each group of biometric data among the two groups, i.e. A and B of biometric data including two modalities of a face feature and an iris feature as an example, similarity calculation is performed on the iris features in the A and B groups, to obtain an iris inter-group comparison score. If the iris inter-group comparison score is greater than a given iris comparison threshold, then it is considered that iris features belong to same person, and the comparison result $\delta(k)$ is assigned with a label 1; and if the iris inter-group comparison score is less than or equal to the given iris comparison threshold, then it is considered that the iris features belong to different persons, and the comparison result $\delta(k)$ is assigned with a label 0.

The same applies to the face feature, similarity calculation is performed on face features of groups A and B, to obtain a face inter-group comparison score. If the face inter-group comparison score is greater than a given face comparison threshold, then it is considered that face features belong to same person, and the comparison result $\delta(k)$ is assigned with a label 1; and if the face inter-group comparison score is less than or equal to the given face comparison threshold, then it is considered that the face features belong to different persons, and the comparison result $\delta(k)$ is assigned with a label 0.

Finally, decision fusion is performed on all assigned comparison results $\delta(1)$, $\delta(2)$, ..., $\delta(k)$, ..., $\delta(m)$, and this is actually weighted voting, so as to obtain a final decision fusion result δ .

During weighted voting, different weights are set for each modality; and 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.

The dimensions and distributions of comparison scores obtained in different modalities are not unified, and the biometric authentication performance in different modalities is also different, such that the comparison results of different modalities have different contributions to the final decision fusion result. In the related art, when weights of different modalities are set according to experience, the settings of the weights are inaccurate, thereby reducing the accuracy of multimodal identity authentication.

In the present disclosure, 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 distribution of the comparison scores of the training sample set of the modality may reflect the performance of the modality authentication method, and the comparison scores can also be subjected to dimensional unification according to the distribution. Moreover, 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 is an index for evaluating the performance of the authentication method, and indicates the rate of samples of same person being judged as samples of different persons. The false rejection rate introduces the determination of weights, and can be used for generation of weights, such that optimization is performed towards a target direction beneficial to decreasing the false rejection rate.

Therefore, in the present disclosure, according to the weight of each modality determined on the basis of 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 dimensions and distribution of the comparison scores can be unified, and the adverse effects of differences in the performance of authentication methods in different modalities on identity authentication after fusion can also be reduced, and then a better identity authentication accuracy can be obtained.

S400: whether the two groups of multi-modal biometric data belong to same person is judged according to the decision fusion result.

Generally, if the decision fusion result is greater than a set decision threshold T_δ , it is judged that the two groups of multi-modal biometric data belong to same person; otherwise, it is judged that the two groups of multi-modal biometric data do not belong to same person.

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 the present disclosure, the decision fusion includes a training process and an inference process. The ultimate object of the training process is to obtain a fused weight $w(k)$ of each modality by a labeled training sample set; and the ultimate object of the inference process is to fuse obtained comparison results of multiple modalities by weights obtained by training so as to obtain a decision fusion result, and to perform authentication according to the decision fusion result.

The inference process includes step S100 to step S400 above. A training process of determining the weight of each modality includes S10 to S30.

S10: pairwise comparison is performed 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 includes comparison scores of same person and comparison scores of different persons.

A training sample set for one modality, such as an iris sample set, includes several iris feature samples, and the iris feature samples are from multiple persons. Iris feature samples of same person or different persons are compared to obtain comparison scores of same person and comparison scores of different persons, thereby forming a comparison score set.

Exemplarily, taking $s_i^j(k)$ representing an i th comparison score in a comparison score set of a k th modality as an example, $i = 1, 2, 3, \dots, N_k$, N_k is the total number of comparison scores in the comparison score set of the k th modality, $j = 0, 1$, $j = 0$ 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.

The comparison scores of same person obtained by comparison regarding the k th modality are denoted as $a(k)$; and

$$a(k) = \{s_i^1(k), i = 1, 2, 3, \dots, N_{k1}\}$$

the comparison scores of different persons obtained by comparison regarding the k th modality are denoted as $r(k)$.

$$r(k) = \{s_i^0(k), i = 1, 2, 3, \dots, N_{k2}\}$$

The comparison scores of same person and the comparison scores of different persons substantially satisfy Gaussian distribution, and the comparison scores of same person and the comparison scores of different persons in the k th modality are counted, and the distributions thereof are as shown in Fig. 2:

$\max_r(k)$ is a maximum value of the comparison scores of different persons in the comparison score set of the k th modality;

$\min_r(k)$ is a minimum value of the comparison scores of different persons in the comparison score set of the k th modality;

$\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the k th modality;

$\max_a(k)$ is a maximum value of the comparison scores of same person in the comparison score set of the k th modality;

$\min_a(k)$ is a minimum value of the comparison scores of same person in the comparison score set of the k th modality; and

$\mu_a(k)$ is the average value of the comparison scores of same person in the comparison score set of the k th modality.

Obviously, an overlapping portion of distributions of the comparison scores of same person and the comparison scores of different persons may reflect the performance of a comparison algorithm. The larger the overlapping portion, the less obvious the difference between the comparison scores of same person and the comparison scores of different persons, and the poorer the performance; on the contrary, the smaller the overlapping portion, the higher the distinction degree between the comparison scores of same person and the comparison scores of different persons, 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.

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 k th 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 k th 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:

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

where $\hat{s}_i^j(k)$ is an i th comparison score in the comparison score set of the k th modality after normalization.

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 k th modality;

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

FRR_k is the false rejection rate of the training sample set of the k th 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

direction beneficial to decreasing the false rejection rate.

The present disclosure do not limit the form of multi-modal biometric data. In one example, the multi-modal biometric data includes 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.

In the present disclosure, two biological modalities, i.e. face recognition and iris recognition are fused by a multimodal decision fusion method, so as to achieve more accurate identity authentication.

Most of existing multimodal fusion is decision fusion. In the present disclosure, feature fusion is performed on the left-eye iris feature and the right-eye iris feature, such that the authentication effect is better. Since the left and right irises of a person are different, and belong to data of different modalities. By fusing the left-eye iris feature and the right-eye iris feature, the identity authentication effect can be improved; in cooperation with multimodal decision fusion of iris and face, feature fusion and decision fusion are used in combination; and a better identity authentication effect is achieved by multiple fusion manners, such that the identity authentication is more secure and reliable.

In one example, the left-eye iris feature and the right-eye iris feature are fused to obtain the iris feature, includes: steps S1 and S2.

S1: principal component analysis is performed on the left-eye iris feature and the right-eye iris feature by the following formula.

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

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^{lT} 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.

S2: the features obtained by performing principal component analysis on the left-eye iris feature and the right-eye iris feature are fused by the following formula, to obtain the iris feature:

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

where Z is the iris feature, P^0 and P^1 are projection matrices, and P^0 and P^1 are obtained by

training the left-eye iris features training set and the right-eye iris features training set.

In the present disclosure, the left-eye iris feature and the right-eye iris feature are both used, which has a high data dimension, and introduces an additional calculation amount and storage space. Therefore, the principal component analysis method is used to perform dimension reduction processing on the left-eye iris feature and the right-eye iris feature, so as to extract more prominent features from original features, reduce the data amount during classification, and increase the classification capability for the features.

In the principal component analysis method, high-dimensional data is projected onto a low-dimensional space via linear mapping, such that the variance of the projected data is the maximum, and thus the characteristics of the original data are retained.

After the iris feature is obtained by fusing the left-eye iris feature and the right-eye iris feature, subsequent multimodal identity authentication is performed by combining the iris feature with a face feature.

For example, with regard to two groups, i.e. groups A and B of biometric data, a left-eye iris feature and a right-eye iris feature of each group can be fused by the described process to obtain one iris feature for each group, and then an inter-group comparison score (i.e. similarity) between iris features of the two groups is calculated. If the inter-group comparison score is greater than a given iris comparison threshold, then it is judged that the iris features of the two groups are from same person, and the comparison result $\delta(k)$ is assigned with a label 1; and if the inter-group comparison score is less than or equal to the iris comparison threshold, then it is judged that the iris features of the two groups are from different persons, and the comparison result $\delta(k)$ is assigned with a label 0.

Since there is only one face feature in each group, face features of the two groups of can be directly compared. If an inter-group comparison score is greater than a given face comparison threshold, then it is judged that the face features of the two groups are from same person, and a comparison result $\delta(k)$ is assigned with a label 1; and if the inter-group comparison score is less than or equal to the face comparison threshold, then it is judged that the face features of the two groups are from different persons, and the comparison result $\delta(k)$ is assigned with a label 0.

Finally, decision fusion is performed on the assigned comparison result $\delta(k)$ by the weight $w(k)$.

In some embodiments of the present disclosure, the μ^l , W^l , P^o and P^1 are obtained by training, includes: the following steps S11 to S16.

S11: the average value of the left-eye iris features and the average value of the right-eye iris features are calculated by the following formula:

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

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

S12: a covariance matrix of the left-eye iris features samples and the right-eye iris feature samples is calculated by the following formula:

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

wherein S^l is the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples, and the size thereof is $d \times d$, and $l = 0, 1$ denotes a left eye and a right eye, respectively.

S13: eigenvalues of the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples are decomposed by the following formula:

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

wherein λ_t^l and p_t^l are respectively eigenvalues and eigenvectors of a covariance matrix S^l , $t = 0, 1, 2, \dots, d$, d is a number of rows and a number of columns of the covariance matrix S^l .

S14: the first r eigenvectors accounting for 95% are selected according to the percentage φ^l of principal components to constitute the principal component analysis matrix W^l ;

$$\text{where } \varphi^l = \frac{\sum_{i=1}^r \lambda_i^l}{\sum_{i=1}^d \lambda_i^l}, \quad W^l = [p_0^l, p_1^l, \dots, p_r^l].$$

S15: principal component analysis is performed on the left-eye iris feature samples and the right-eye iris feature samples by the following formula.

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

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.

S16: the projection matrices P^0 and P^1 are obtained by optimization by the following formula:

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

wherein $Z^0 = [z_0^0, z_1^0, \dots, z_n^0]$, $Z^1 = [z_0^1, z_1^1, \dots, z_n^1]$, $S_{Z^0Z^0}$ 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 .

To more clearly illustrate the effects of the present disclosure, the multimodal identity authentication method (the total number of modalities $m = 2$) in the present disclosure is tested on a test set, and the identity authentication effect thereof is as shown in the following table.

Identity authentication method	Iris recognition method	Face recognition method	Multimodal identity authentication method in the present disclosure
Identity authentication effect	98.9%	98.1%	99.87%

From the content above, it can be determined that compared with a single modality, the multimodal identity authentication method in the present disclosure has a better identity authentication effect.

In some embodiments of the present disclosure, provided is a multimodal identity authentication apparatus. As shown in Fig. 3, the multimodal identity authentication apparatus includes a data acquisition module 1, a comparison score calculation module 2, a decision fusion module 3 and a judgment module 4.

The data acquisition module 1 is configured to acquire two groups of multi-modal biometric data.

The comparison score calculation module 2 is 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.

The decision fusion module 3 is 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.

The judgment module 4 is configured to judge whether the two groups of multi-modal biometric data belong to same person according to the decision fusion result.

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 some embodiments of the present disclosure, the weight of each modality can be determined by a training data preparation module, a normalization module and a weight calculation module.

The training data preparation module is configured to perform 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 includes comparison scores of same person and comparison scores of different persons.

The normalization module is configured to normalize 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.

The weight calculation module is configured to calculate 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.

The normalization module includes a normalized-dimension calculation unit and a normalization unit.

The normalized-dimension calculation unit is configured to calculate, by the following formula, the 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 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 k th modality, $k = 1, 2, \dots, m$, 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 k th modality, $\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the k th modality, $\min_a(k)$ is a minimum value of comparison scores of same person in the comparison score set of the k th modality, and $\mu_a(k)$ is an average value of the comparison scores of same person in the comparison score set of the k th modality.

The normalization unit is configured to normalize the comparison score set of each modality according to the normalized dimension of each modality:

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

wherein $s_i^j(k)$ and $\hat{s}_i^j(k)$ are respectively i th comparison scores in the comparison score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, N_k$, N_k is a total number of comparison scores in the comparison score set of the k th 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.

The weight calculation module is configured to calculate the weight of each modality by the following formula:

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

wherein $w(k)$ is the weight of the k th modality;

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

FRR_k is the false rejection rate of the training sample set of the k th modality.

In some embodiments, the decision fusion result may be calculated by the following formula:

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

wherein δ is the decision fusion result, $w(k)$ is the weight of the k th modality, and $\delta(k)$ is a comparison result of the k th modality:

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

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

The judgment module is further configured to judge that the two groups of multi-modal biometric data belong to same person if the decision fusion result is greater than a set decision threshold T_δ ; otherwise, judge that the two groups of multi-modal biometric data do not belong to same person.

In some embodiments of the present disclosure, the multi-modal biometric data includes 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.

The left-eye iris feature and the right-eye iris feature can be fused to obtain the iris feature by a first principal component analysis module and a feature fusion module.

The first principal component analysis module is configured to perform principal component analysis on the left-eye iris feature and the right-eye iris feature by the following formula.

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

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^{lT} 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.

The feature fusion module is configured to fuse the features obtained by performing principal component analysis on the left-eye iris feature and the right-eye iris feature by the following formula, to obtain the iris feature:

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

wherein Z is the iris feature, P^0 and P^1 are projection matrices, and P^0 and P^1 are obtained by training the left-eye iris features training set and the right-eye iris features training set.

The μ^l , W^l , P^0 and P^1 as described above can be obtained by training by an average value calculation module, a covariance matrix calculation module, an eigenvalue decomposition module, a principal component analysis matrix acquisition module, a second principal component analysis module and a projection matrix calculation module.

The average value calculation module is configured to calculate the average value of the left-

eye iris features and the average value of the right-eye iris features by the following formula:

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

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

The covariance matrix calculation module is configured to calculate a covariance matrix of the left-eye iris features samples and the right-eye iris feature samples by the following formula:

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

wherein S^l is the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples, and the size thereof is $d \times d$, and $l = 0, 1$ denotes a left eye and a right eye, respectively.

The eigenvalue decomposition module is configured to decompose eigenvalues of the covariance matrix of the left-eye iris features samples and the right-eye iris feature samples by the following formula:

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

wherein λ_t^l and p_t^l are respectively eigenvalues and eigenvectors of a covariance matrix S^l , $t = 0, 1, 2, \dots, d$, d is a number of rows and a number of columns of the covariance matrix S^l .

The principal component analysis matrix acquisition module is configured to select 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 ,

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

The second principal component analysis module is configured to perform principal component analysis on the left-eye iris feature samples and the right-eye iris feature samples by the following formula, to obtain:

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

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.

The projection matrix calculation module is configured to optimize by the following formula, to obtain the projection matrices P^0 and P^1 :

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

wherein $Z^0 = [z_0^0, z_1^0, \dots, z_n^0]$, $Z^1 = [z_0^1, z_1^1, \dots, z_n^1]$, $S_{Z^0 Z^0}$ 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 .

The implementation principle and technical effects of the multimodal identity authentication apparatus provided in the embodiments of the present disclosure are the same as those of the embodiments of the multimodal identity authentication method. For the sake of brief illustration, for parts not mentioned in the embodiments of the multimodal identity authentication apparatus, reference can be made to the corresponding content in the embodiments of the multimodal identity authentication method. It can be clearly understood by a person skilled in the art, for convenience and conciseness of illustration, for specific working processes of the described multimodal identity authentication apparatus and units, reference can all be made to the corresponding process in the embodiments of the multimodal identity authentication method, and details will not be repeated herein again.

In the embodiments of the multimodal identity authentication method provided in the present disclosure, the multimodal identity authentication method may implement service logic by a computer program and record same on a storage medium; and the storage medium may be read and executed by a computer, thereby implementing the effects of the solutions described in the embodiments of the multimodal identity authentication method in the present description. Therefore, the present disclosure further 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 in the embodiments of the multimodal identity authentication method.

The storage medium may include a physical apparatus for storing information, and generally, the information is digitalized and then stored in a medium such as an electric medium, a magnetic medium, or an optical medium. The storage medium may include: an apparatus that stores information in an electric energy manner, e.g. various memories, such as an RAM and an ROM; an apparatus that stores information in a magnetic energy manner, such as a hard disk, a floppy disk, 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 identity authentication 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 identity authentication method, and for details, reference may be made to the illustration of the embodiments of the multimodal identity authentication method, and details will not be repeated herein again.

In the present disclosure, further provided is a device for multimodal identity authentication. The device may be a separate computer, and may also include an actual operation apparatus or the like using one or more multimodal identity authentication methods of the present description or the multimodal identity authentication apparatus of one or more embodiments. The device for multimodal identity authentication 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 identity authentication method.

According to the illustration of the embodiments of the multimodal identity authentication method, the device for multimodal identity authentication 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 identity authentication method, and for details, reference may be made to the illustration of the embodiments of the multimodal identity authentication method, and details will not be repeated herein again.

The current silent face liveness detection method uses a single data source for liveness detection (namely, single-modality liveness detection), and the amount of information available in the single data source is fixed, such that a good face liveness detection effect cannot be achieved.

In some embodiments of the present disclosure, a multimodal liveness detection method is provided. As shown in Fig. 4, the method includes steps S2100 to S2500.

S2100: multi-modal facial data of a same user is acquired.

Since the amount of information available in a single data source is fixed, using a single data source for liveness detection cannot achieve a good face liveness detection effect. Therefore, in the present disclosure, multiple data sources are used for liveness detection, i.e. multimodal liveness detection.

In this step, multi-modal facial data is acquired, and the modality type and acquisition manner

of the multi-modal facial data are not limited. For example, the multi-modal facial data may be face images collected under different types of light sources, and the like.

S2200: liveness detection is performed on the multi-modal facial data of the same user, so as to obtain detection scores of the multiple modalities.

In the present disclosure, the manner of obtaining the detection scores is not limited. For example, a convolutional neural network may be used to respectively process multi-modal facial data, so as to obtain detection scores of the multiple modalities.

The detection scores of the multiple modalities may be denoted as $s_c(k)$, where $s_c(k)$ is a detection score of a k th modality, $k = 1, 2, \dots, m$, and m is the total number of modalities.

S2300: weighted summation is performed on the detection scores of the multiple modalities according to a weight set for each modality, to obtain a fusion score.

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

Exemplarily, the fusion score may be calculated by the following formula:

where s_c is the fusion score, and $w(k)$ is a weight set for the k th modality.

The dimensions and distributions of detection scores obtained in different modalities are not unified, and the liveness detection performance in different modalities is also different, and thus directly fusing the detection scores obtained in different modalities may reduce the liveness detection performance.

In the present disclosure, weighted summation is performed on the detection scores of the multiple modalities by weights, to obtain a fusion score. The weight of each modality is determined according to a distribution of detection scores of a training sample set of the modality and a liveness rejection rate of the training sample set of the modality. The distribution of the detection scores of the training sample set of the modality may reflect the performance of a liveness detection algorithm, and the detection scores can also be subjected to dimensional unification according to the distribution. The liveness rejection rate is an index for evaluating liveness detection, indicates the rate of a liveness sample being judged as a spoof sample, and is an evaluation for misjudging the liveness sample, and the liveness rejection rate can also reflect the performance of the liveness detection algorithm. The liveness rejection rate introduces the determination of weights, and can be used for generation of weights, such that optimization is performed towards a target direction beneficial to decreasing the liveness rejection rate.

Therefore, according to the weight of each modality determined on the basis of the distribution of detection scores of a training sample set of the modality and the liveness rejection rate of the training sample set of the modality, the dimensions and distribution of the detection scores can be unified, and the adverse effects of differences in the performance of liveness detection in different modalities on liveness detection after fusion can also be reduced, and then a better liveness detection effect can be obtained.

S2400: a liveness detection result of each modality is determined according to the detection scores of each modality and a liveness detection threshold set for each modality, and weighted voting is performed 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.

The liveness detection decision fusion result is a result of performing weighted voting on the liveness detection results of the multiple modalities, and reflects the effects achieved by the liveness detection result of each modality in the final liveness detection, and the proportion of the liveness detection result of each modality in the final liveness detection.

Exemplarily, the liveness detection decision fusion result may be calculated by the following formula:

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

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 k th modality.

S2500: it is judged whether the facial data belongs to a live person or a spoof according to the fusion score and/or the liveness detection decision fusion result.

Two judgment factors may be obtained from the multi-modal facial data through the processing above, which are the fusion score and the liveness detection decision fusion result. The fusion score and the liveness detection decision fusion result are both important indexes during liveness detection, and can be judged respectively, and can also be judged jointly.

For example, if the fusion score s_c is greater than a set first threshold T_s , it is judged that the facial data belongs to a live person, otherwise, it is judged that the facial data belongs to a spoof.

Or, if the liveness detection decision fusion result δ_f is greater than a set second threshold δ_s , it is judged that the facial data belongs to a live person, otherwise, it is judged that the facial data belongs to a spoof.

Or, if the fusion score s_c is greater than the set first threshold T_s and the liveness detection decision fusion result δ_f is greater than the set second threshold δ_s , it is judged that the facial data belongs to a live person, otherwise, it is judged that the facial data belongs to a spoof.

The present disclosure have the following beneficial effects:

1. In the related art, a single modality is used for face liveness detection; whereas in the present disclosure, multi-modal facial data is used for multimodal face liveness detection, thereby increasing the amount of information available for face liveness detection, and achieving a better liveness detection effect.

2. Detection scores of multiple modalities are fused by weight, and weighted voting is performed on liveness detection results of the multiple modalities, that is, not only detection scores are fused, but also multimodal decision fusion is performed. Furthermore, the fused weight is related to the distribution of the detection scores and the liveness rejection rate of each modality, and the detection scores and decisions are subjected to multimodal fusion by using the distribution of the detection scores and the liveness rejection rate, and the dimensions and distribution of the detection scores are unified by weights, thereby reducing the adverse effects of differences in the performance of liveness detection of different modalities on liveness detection after fusion, such that the multimodal liveness detection method is more effective and feasible, and a better multimodal liveness detection effect can be obtained.

3. The detection score fusion and decision fusion performed in the present disclosure are both simple numerical calculation problems, do not introduce additional time overhead, and do not affect the overall speed of a liveness detection algorithm.

The present disclosure include a training process and an inference process. The ultimate object of the training process is to obtain a fused weight $w(k)$ of each modality by a labeled training sample set; and the ultimate object of the inference process is to judge, regarding the obtained multi-modal facial data, whether the multi-modal facial data belongs to a liveness or spoof category by weights obtained by training.

The inference process includes steps S2100-S2500 above; and a training process of determining the weight of each modality includes steps S210 to S230.

S210: liveness detection is performed 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 includes liveness samples and spoof samples, and the detection score set includes liveness scores and spoof scores. A detection score obtained by performing liveness detection on the liveness sample is a liveness score, and a detection score obtained by performing liveness detection on the spoof sample is a spoof score.

Exemplarily, $s_{c_i}^j(k)$ denotes an i th detection score in a detection score set of a k th modality, $i = 1, 2, 3, \dots, M_k$, M_k is the total number of detection scores in the detection score set of the k th modality, $j = 0, 1$, $j = 0$ 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.

A liveness score obtained by performing liveness detection on the liveness sample in the k th modality is denoted as $r_s(k)$; and

$$r_s(k) = \{s_{c_i}^1(k), i = 1, 2, 3, \dots, M_{k1}\}$$

a spoof score obtained by performing liveness detection on the spoof sample in the k th modality is denoted as $a_s(k)$;

$$a_s(k) = \{s_{c_i}^0(k), i = 1, 2, 3, \dots, M_{k2}\}$$

The distribution of the liveness scores and the distribution of the spoof scores substantially satisfy Gaussian distribution, and the spoof scores and the liveness scores of the k th modality are counted, and the distributions thereof are as shown in Fig. 5.

$\max_{r_s}(k)$ is a maximum value of the liveness scores in the detection score set of the k th modality.

$\min_{r_s}(k)$ is a minimum value of the liveness scores in the detection score set of the k th modality.

$\mu_{r_s}(k)$ is an average value of the liveness scores in the detection score set of the k th modality.

$\max_{a_s}(k)$ is a maximum value of the spoof scores in the detection score set of the k th modality.

$\min_{a_s}(k)$ is a minimum value of the spoof scores in the detection score set of the k th modality.

$\mu_{a_s}(k)$ is the average value of the spoof scores in the detection score set of the k th modality.

Obviously, the overlapping portion of the distributions of the liveness scores and the spoof scores can reflect the performance of a liveness detection algorithm of the modality, and the larger

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 k th 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 k th 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:

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

where $\hat{s}_{c_i}^j(k)$ is an i th detection score in the detection score set of the k th modality after normalization.

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 k th modality;

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

LRR_k is a liveness rejection rate of a liveness detection training sample set of the k th 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.

Example I:

An existing face liveness detection method uses a deep learning method to extract features from an entire face image to perform a binary classification task, so as to judge whether the face belongs to a live person or a spoof. The method only uses texture information of the entire face, and has a poor detection effect on a face in which a local small range is modified. For example, if only an eye region is replaced with a spoof eye region, this replacement can break through a silent liveness detection method based on an entire face.

In order to solve the described problem, in the present example, multi-modal facial data includes a plurality of different regions of a face image, wherein the multiple different regions of each face image include 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.

Taking a visible-light face image as an example, a method for performing multimodal liveness detection thereon is:

1. Facial key points of the visible-light face image are positioned.
2. A plurality of different regions of the face image are obtained according to the positioning result of the facial key points, and are used as the multi-modal facial data. The plurality of different regions may be the whole region, an eye region, and a mouth region of the face image, and an upper-half face region and a lower-half face region, etc.
3. A detection score for each region (modality) can be obtained by a liveness detection algorithm, and then score fusion and decision fusion are performed by the method in the present disclosure.

The present example uses the whole region and the local region of the visible-light face image for multimodal liveness detection, and has a good liveness detection effect on the modification of a local small range. Liveness detection for the near-infrared face image is the same as that of the visible-light face image, and details will not be repeated herein again.

Example II:

The multi-modal facial data includes face images under different types of light sources; and the face images under different types of light sources may include visible-light face images and near-infrared face images.

The present example is applied to a dual-mode device capable of obtaining both visible-light face images and near-infrared face images, and the visible-light face images and the near-infrared

face images collected thereby are used as multi-modal facial data, detection scores thereof are respectively obtained, and then score fusion and decision fusion are performed by using the method of the present disclosure.

To more clearly illustrate the effects of the present disclosure, the multimodal liveness detection method (the total number of modalities $m = 3$) in the present disclosure is tested on a test set, and the liveness detection effect thereof is as shown in the following table.

Liveness detection method	Modality 1	Modality 2	Modality 3	Multimodal fusion
Liveness detection effect	97.3%	96.1%	97.5%	99.71%

From the content above, it can be determined that compared with a single modality, the multimodal liveness detection method in the present disclosure has a better liveness detection effect.

It should be understood that, although the steps in the flowcharts involved in the embodiments as described above are displayed in sequence according to the indication of arrows, these steps are not necessarily executed in sequence according to the indication of the arrows. Unless explicitly specified herein, these steps are executed in no strict order, and these steps may be performed in other orders. Furthermore, at least some of the steps in the flowcharts involved in the embodiments as stated above may include a plurality of steps or a plurality of stages, these steps or stages are not necessarily executed at the same time, but may be performed at different times; and the order in which these steps or stages are executed is also not necessarily sequential, but may be executed in turn or alternately with at least some of the steps or stages in other steps or other stages.

Some embodiments of the present disclosure provide a multimodal liveness detection apparatus. As shown in Fig. 6, the multimodal liveness detection apparatus includes a data acquisition module 10, a detection score calculation module 20, a fusion score calculation module 30, a decision fusion result calculation module 40 and a liveness judgment module 50.

The data acquisition module 10 is configured to acquire multi-modal facial data of a same user.

The detection score calculation module 20 is 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.

The fusion score calculation module 30 is 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.

The decision fusion result calculation module 40 is 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.

The liveness judgment module 50 is configured to judge whether the facial data belongs to a live person or a spoof according to the fusion score and/or the liveness detection decision fusion result.

In some embodiments of the present disclosure, the weight of each modality is determined by a training data preparation module, a normalization module and a weight calculation module.

The training data preparation module is configured to perform 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 includes liveness samples and spoof samples, and the detection score set includes liveness scores and spoof scores.

The normalization module is configured to normalize 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.

The weight calculation module is configured to calculate the weight of each modality according to the normalized detection score set of each modality and the liveness rejection rate of the liveness detection training sample set of each modality.

The normalization module as described above may include a normalized-dimension calculation unit and a normalization unit.

The normalized-dimension calculation unit is configured to calculate a non-overlapping region of a distribution of the liveness scores of each modality and a distribution of the spoof scores of each modality by the following formula, as a normalized dimension of each modality:

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

scores in the detection score set of the k th modality, $\min_{as}(k)$ is a minimum value of spoof scores in the detection score set of the k th modality, and $\mu_{as}(k)$ is an average value of the spoof scores in the detection score set of the k th modality.

The normalization unit is configured to normalize the detection score set of each modality according to the normalized dimension of each modality.

$\hat{s}_{c_i}^j(k) = \frac{|s_{c_i}^j(k) - q_s(k)|}{q_s(k)}$, wherein $s_{c_i}^j(k)$ and $\hat{s}_{c_i}^j(k)$ are respectively i th detection scores in the detection score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, M_k$, M_k is a total number of detection scores in the detection score set of the k th 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.

The weight calculation module above is configured to calculate the weight of each modality by the following formula:

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

wherein $w(k)$ is the weight of the k th modality;

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

As another improvement of the embodiments of the present disclosure, the fusion score is calculated by the following formula:

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

The decision fusion result is calculated by the following formula:

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

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

In the present disclosure, the liveness judgment module is configured to: judge that the facial data belongs to a live person if the fusion score is greater than a set first threshold, and otherwise, judge that the facial data belongs to a spoof; or judge that the facial data belongs to a live person if

the liveness detection decision fusion result is greater than a set second threshold, and otherwise, judge that the facial data belongs to a spoof; or judge 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, and otherwise, judge that the facial data belongs to a spoof.

The multi-modal facial data as described above may include face images under different types of light sources.

Or the multi-modal facial data as described above may further include multiple different regions of a face image. The multiple different regions of a face image include 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.

The implementation principle and technical effects of the multimodal liveness detection apparatus provided in the embodiments of the present disclosure are the same as those of the embodiments of the multimodal liveness detection method. For the sake of brief illustration, for parts not mentioned in the embodiments of the multimodal liveness detection apparatus, reference can be made to the corresponding content in the embodiments of the multimodal liveness detection method. It can be clearly understood by a person skilled in the art, for convenience and conciseness of illustration, for specific working processes of the described multimodal liveness detection apparatus and units, reference can all be made to the corresponding process in the embodiments of the multimodal liveness detection method, and details will not be repeated herein again.

The multimodal liveness detection method of the described embodiments provided in some embodiments of the present disclosure may implement service logic by a computer program and record same on a storage medium; and the storage medium may be read and executed by a computer, thereby implementing the effects of the solutions described in the embodiments of the multimodal liveness detection method in the present description. Therefore, the present disclosure further provide a non-transitory computer-readable storage medium for multimodal liveness detection, including a memory for storing instructions executable by a processor, which when executed by the processor, implement the steps in the embodiments of the multimodal liveness detection method.

The storage medium may include a physical apparatus for storing information, and generally, the information is digitalized and then stored in a medium such as an electric medium, a magnetic medium, or an optical medium. The storage medium may include: an apparatus that stores information in an electric energy manner, e.g. various memories, such as an RAM and an ROM; an apparatus that stores information in a magnetic energy manner, such as a hard disk, a floppy disk,

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 medical image processing method when executed by the processor. The display screen of the computer device may

be a liquid crystal display screen or an electronic ink display screen; and the input apparatus of the computer device may be a touch layer covered on the display screen, may also be a key, a trackball or a touch panel arranged on a housing of the computer device, and may also be an external keyboard, touch panel or mouse, etc.

A person skilled in the art may understand that the structure shown in Fig. 7 is only a block diagram of some structures related to the solutions of some embodiments of the present disclosure, and does not limit a computer device to which the solutions of some embodiments of the present disclosure are applied. A specific computer device may include more or fewer components than those shown in the figure, or combine some components, or have different component arrangements.

Finally, it should be noted that the embodiments above are merely specific embodiments of the present disclosure, and are used to describe technical solutions of some embodiments of the present disclosure, rather than limit same, and the scope of protection of the present disclosure is not limited thereto. Although the present disclosure has been explained in detail with reference to the embodiments above, a person of ordinary skill in the art would understand: any person skilled in the art who is familiar with the present technical field can still modify or easily conceive of changes to the technical solutions disclosed in the embodiments above within the technical scope disclosed in some embodiments of the present disclosure, or make equivalent replacements for some of the technical features therein; moreover, these modifications, changes or replacements shall not render that the nature of the corresponding technical solutions departs from the spirit and scope of the technical solutions in embodiments of the present disclosure. All these shall be covered in the scope of protection of the present disclosure. Therefore, the scope of protection of the present disclosure shall be subject to the scope of protection of the appended claims.

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 k th modality, $k = 1, 2, \dots, m$, 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 k th modality, $\mu_r(k)$ is an average value of the comparison scores of different persons in the comparison score set of the k th modality, $\min_a(k)$ is a minimum value of comparison scores of same person in the comparison score set of the k th modality, and $\mu_a(k)$ is an average value of the comparison scores of same person in the comparison score set of the k th modality;

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

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

wherein $s_i^j(k)$ and $\hat{s}_i^j(k)$ are respectively i th comparison scores in the comparison score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, N_k$, N_k is a total number of comparison scores in the comparison score set of the k th 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 k th modality;

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

FRR_k is the false rejection rate of the training sample set of the k th 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 k th modality, and $T(k)$ is a comparison threshold of the k th 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:

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

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^{lT} 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:

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

wherein Z is the iris feature, P^0 and P^1 are projection matrices, and P^0 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^0 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:

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

wherein x_i^l is an i th left-eye iris feature sample in the left-eye iris features training set or an i th right-eye iris feature sample in the right-eye iris features training set, $i = 1, 2, 3, \dots, n$, 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:

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

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, $t = 0, 1, 2, \dots, d$, 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 $\phi^l = \frac{\sum_{i=1}^r \lambda_i^l}{\sum_{i=1}^d \lambda_i^l}$, $W^l = [p_0^l, p_1^l, \dots, p_r^l]$;

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

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

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^0 and P^1 as:

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

wherein $Z^0 = [z_0^0, z_1^0, \dots, z_n^0]$, $Z^1 = [z_0^1, z_1^1, \dots, z_n^1]$, $S_{Z^0 Z^0}$ 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 .

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 k th modality, $k = 1, 2, \dots, m$, m is a total number of modalities, $\max_{rs}(k)$ is a maximum value of liveness scores in a detection score set of the k th modality, $\mu_{rs}(k)$ is an average value of the liveness scores in the detection score set of the k th modality, $\min_{as}(k)$ is a minimum value of spoof scores in the detection score set of the k th modality, and $\mu_{as}(k)$ is an average value of the

spoof scores in the detection score set of the k th modality;

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

$\hat{s}_{c_i}^j(k) = \frac{|s_{c_i}^j(k) - q_s(k)|}{q_s(k)}$, wherein $s_{c_i}^j(k)$ and $\hat{s}_{c_i}^j(k)$ are respectively i th detection scores in the detection score set of the k th modality before normalization and after normalization, $i = 1, 2, 3, \dots, M_k$, M_k is a total number of detection scores in the detection score set of the k th 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 k th modality;

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