

What is claimed is:

1. A method for tongue image extraction, applicable to an electronic device, the method comprising:

S110, converting training images containing tongue into a matrix V , wherein all non-negative gray-values of one image correspond to one column of the matrix V , and training with a local non-negative matrix factorization (LNMF) algorithm, to factorize the matrix V into a product of a non-negative feature matrix W and a weight matrix H , that is, $V=WH$, wherein

the non-negative feature matrix W has a dimension of $n*r$, and r columns are a feature base image, wherein the feature base image refers to the non-negative feature matrix W representing tongue features and the non-negative feature matrix W constitutes a non-negative subspace, and

the weight matrix H has a dimension of $r*m$ and each column is encoded;

S120, using an embedded hidden Markov model (EHMM) to recognize whether a test image contains face, projecting the training images and the test image to the non-negative subspace respectively to obtain feature coefficients when the test image contains face, using a nearest neighbor criterion to obtain similarities among the feature coefficients corresponding to the training images and the test image, and extracting features of the test image representing tongue and having similarities higher than a similarity threshold to be the tongue features; and

S130, after projection, identifying, with different labels, a feature region with the tongue features and a non-feature region without the tongue features respectively, wherein a set of labels corresponds to boundary information of the feature region, extracting extreme values in up, down, left, and right directions of the boundary information, and determining a bounding box defining the feature region.

2. The method of claim 1, further comprising:

S150, using a linear regression model to calculate a translation value and a zoom value of the bounding box and using a loss function to constrain a position error of the bounding box, to adjust the bounding box to move to a suitable position.

3. The method of claim 1, further comprising:

before S110, binarizing the training images and the test image and setting gray-values of

pixels of the training images and the test image to be 0 or 255 according to a set gray-value threshold.

4. The method of claim 1, wherein using the EHMM to recognize that the test image contains face comprises:

selecting a plurality of feature points of the face to form a feature sequence;

inputting the test image into the EHMM to scan the test image from top to bottom and from left to right through a sliding window, to obtain a plurality of groups of eigenvectors, wherein the plurality of groups of eigenvectors form a sequence of observations; and

using a forward algorithm to derive a similar probability between the sequence of observations and the feature sequence containing the plurality of feature points of the face, and recognizing that the test image contains face when the similar probability is greater than a determination threshold, wherein

the EHMM comprises a set of super states, and the number of super states in the set of super states is the same as the number of slice images of the face in a vertical direction, and

each super state packages a set of embedded states, and the number of embedded states in the set of embedded states is the same as the number of slice images of the face in a horizontal direction.

5. The method of claim 1, wherein at S110, the training images are compressed, de-averaged, and normalized, and training with the LNMF algorithm to obtain the feature base image, wherein

subtract from each column element of the matrix V a mean value of the column element; and

normalize the training images according to a ratio of a difference between each column element of the matrix V and a minimum value in the column element to a difference between a maximum value and the minimum value in the column element.

6. The method of claim 4, wherein after recognizing that the test image contains face, classifying the test image into a corresponding age group-gender class and using a LNMF model

for respective age group-gender classes to extract tongue images in respective age group-gender classes comprise:

obtaining a convolutional neural network (CNN) model, wherein the CNN model is trained and used to determine gender and age group, and classifying the training images into respective age group-gender classes;

annotating the training images in respective age group-gender classes, wherein a label regarding age group, gender, and tongue is obtained for each training image;

training the LNMF model according to the label regarding age group, gender, and tongue to obtain a trained LNMF model;

using the trained CNN model to identify gender and age group of the test image and classifying the test image according to the gender and age group; and

extracting a tongue image by using the trained LNMF model trained with age group, gender, and tongue.

7. A device for tongue image extraction, comprising:

a matrix factorization module configured to convert training images containing tongue into a matrix V , wherein all non-negative gray-values of one image correspond to one column of the matrix V , and train with a local non-negative matrix factorization (LNMF) algorithm, to factorize the matrix V into a product of a non-negative feature matrix W and a weight matrix H , that is, $V=WH$, wherein

the non-negative feature matrix W has a dimension of $n*r$, and r columns are a feature base image, wherein the feature base image refers to the non-negative feature matrix W representing tongue features and the non-negative feature matrix W constitutes a non-negative subspace, and

the weight matrix H has a dimension of $r*m$ and each column is encoded;

a tongue features extracting module configured to use an embedded hidden Markov model (EHMM) to recognize whether a test image contains face, project the training images and the test image to the non-negative subspace respectively to obtain feature coefficients when the test image contains face, use a nearest neighbor criterion to obtain similarities among the feature coefficients corresponding to the training images and the test image, and extract features of the test image representing tongue and having similarities higher than a similarity threshold to be the

tongue features; and

a tongue image segmenting module configured to identify, with different labels, a feature region with the tongue features and a non-feature region without the tongue features respectively, wherein a set of labels corresponds to boundary information of the feature region, extract extreme values in up, down, left, and right directions of the boundary information, and determine a bounding box defining the feature region.

8. The device of claim 7, further comprising:

a bounding box adjusting module configured to use a linear regression model to calculate a translation value and a zoom value of the bounding box and use a loss function to constrain a position error of the bounding box, to adjust the bounding box to move to a suitable position.

9. The device of claim 7, further comprising a face recognition module configured to use the EHMM to recognize that the test image contains face, wherein the face recognition module configured to use the EHMM to recognize that the test image contains face is configured to:

select a plurality of feature points of the face to form a feature sequence;

input the test image into the EHMM to scan the test image from top to bottom and from left to right through a sliding window, to obtain a plurality of groups of eigenvectors, wherein the plurality of groups of eigenvectors form a sequence of observations; and

use a forward algorithm to derive a similar probability between the sequence of observations and the feature sequence containing the plurality of feature points of the face, and recognize that the test image contains face when the similar probability is greater than a determination threshold, wherein

the EHMM comprises a set of super states, and the number of super states in the set of super states is the same as the number of slice images of the face in a vertical direction, and

each super state packages a set of embedded states, and the number of embedded states in the set of embedded states is the same as the number of slice images of the face in a horizontal direction.

10. The device of claim 7, further comprising a binarizing module configured to binarize the

training images and the test image and set gray-values of pixels of the training images and the test image to be 0 or 255 according to a set gray-value threshold.

11. An electronic device, comprising a processor and a memory configured to store a program for tongue image extraction, which when executed by the processor, is configured to implement:

S110, converting training images containing tongue into a matrix V , wherein all non-negative gray-values of one image correspond to one column of the matrix V , and training with a local non-negative matrix factorization (LNMF) algorithm, to factorize the matrix V into a product of a non-negative feature matrix W and a weight matrix H , that is, $V=WH$, wherein

the non-negative feature matrix W has a dimension of $n*r$, and r columns are a feature base image, wherein the feature base image refers to the non-negative feature matrix W representing tongue features and the non-negative feature matrix W constitutes a non-negative subspace, and

the weight matrix H has a dimension of $r*m$ and each column is encoded;

S120, using an embedded hidden Markov model (EHMM) to recognize whether a test image contains face, projecting the training images and the test image to the non-negative subspace respectively to obtain feature coefficients when the test image contains face, using a nearest neighbor criterion to obtain similarities among the feature coefficients corresponding to the training images and the test image, and extracting features of the test image representing tongue and having similarities higher than a similarity threshold to be the tongue features; and

S130, after projection, identifying, with different labels, a feature region with the tongue features and a non-feature region without the tongue features respectively, wherein a set of labels corresponds to boundary information of the feature region, extracting extreme values in up, down, left, and right directions of the boundary information, and determining a bounding box defining the feature region.

12. The electronic device of claim 11, wherein the program for tongue image extraction, when executed by the processor, is further configured to implement:

before S110, binarizing the training images and the test image and setting gray-values of pixels of the training images and the test image to be 0 or 255 according to a set gray-value

threshold.

13. The electronic device of claim 11, wherein the program for tongue image extraction, when executed by the processor, is further configured to implement:

S150, using a linear regression model to calculate a translation value and a zoom value of the bounding box and using a loss function to constrain a position error of the bounding box, to adjust the bounding box to move to a suitable position.

14. The electronic device of claim 12, wherein the program for tongue image extraction, when executed by the processor, is further configured to implement: using the EHMM to recognize that the test image contains face, which comprises:

selecting a plurality of feature points of the face to form a feature sequence;

inputting the test image into the EHMM to scan the test image from top to bottom and from left to right through a sliding window, to obtain a plurality of groups of eigenvectors, wherein the plurality of groups of eigenvectors form a sequence of observations; and

using a forward algorithm to derive a similar probability between the sequence of observations and the feature sequence containing the plurality of feature points of the face, and recognizing that the test image contains face when the similar probability is greater than a determination threshold, wherein

the EHMM comprises a set of super states, and the number of super states in the set of super states is the same as the number of slice images of the face in a vertical direction, and

each super state packages a set of embedded states, and the number of embedded states in the set of embedded states is the same as the number of slice images of the face in a horizontal direction.

15. The electronic device of claim 12, wherein the program for tongue image extraction, when executed by the processor, is further configured to implement:

at S110, the training images are compressed, de-averaged, and normalized, and training with the LNMF algorithm to obtain the feature base image, wherein

subtract from each column element of the matrix V a mean value of the column

element; and

normalize the training images according to a ratio of a difference between each column element of the matrix V and a minimum value in the column element to a difference between a maximum value and the minimum value in the column element.

16. The electronic device of claim 12, wherein the program for tongue image extraction, when executed by the processor, is further configured to implement: after recognizing that the test image contains face, classifying the test image into a corresponding age group-gender class and using a LNMF model for respective age group-gender classes to extract tongue images in respective age group-gender classes, which comprise:

obtaining a convolutional neural network (CNN) model, wherein the CNN model is trained and used to determine gender and age group, and classifying the training images into respective age group-gender classes;

annotating the training images in respective age group-gender classes, wherein a label regarding age group, gender, and tongue is obtained for each training image;

training the LNMF model according to the label regarding age group, gender, and tongue to obtain a trained LNMF model;

using the trained CNN model to identify gender and age group of the test image and classifying the test image according to the gender and age group; and

extracting a tongue image by using the trained LNMF model trained with age group, gender, and tongue.

17. A non-transitory computer readable storage medium storing a computer program, the computer program comprising program instructions, which when executed by a processor, are configured to implement the method for tongue image extraction of claim 1.

18. The non-transitory computer readable storage medium of claim 17, wherein the program instructions, when executed by the processor, are further configured to implement: using the EHMM to recognize that the test image contains face, which comprises:

selecting a plurality of feature points of the face to form a feature sequence;

inputting the test image into the EHMM to scan the test image from top to bottom and

from left to right through a sliding window, to obtain a plurality of groups of eigenvectors, wherein the plurality of groups of eigenvectors form a sequence of observations; and

using a forward algorithm to derive a similar probability between the sequence of observations and the feature sequence containing the plurality of feature points of the face, and recognizing that the test image contains face when the similar probability is greater than a determination threshold, wherein

the EHMM comprises a set of super states, and the number of super states in the set of super states is the same as the number of slice images of the face in a vertical direction, and

each super state packages a set of embedded states, and the number of embedded states in the set of embedded states is the same as the number of slice images of the face in a horizontal direction.

19. The non-transitory computer readable storage medium of claim 17, wherein the program instructions, when executed by the processor, are further configured to implement:

at S110, the training images are compressed, de-averaged, and normalized, and training with the LNMF algorithm to obtain the feature base image, wherein

subtract from each column element of the matrix V a mean value of the column element; and

normalize the training images according to a ratio of a difference between each column element of the matrix V and a minimum value in the column element to a difference between a maximum value and the minimum value in the column element.

20. The non-transitory computer readable storage medium of claim 17, wherein the program instructions, when executed by the processor, are further configured to implement: after recognizing that the test image contains face, classifying the test image into a corresponding age group-gender class and using a LNMF model for respective age group-gender classes to extract tongue images in respective age group-gender classes, which comprise:

obtaining a convolutional neural network (CNN) model, wherein the CNN model is trained and used to determine gender and age group, and classifying the training images into respective age group-gender classes;

annotating the training images in respective age group-gender classes, wherein a label regarding age group, gender, and tongue is obtained for each training image;

training the LNMF model according to the label regarding age group, gender, and tongue to obtain a trained LNMF model;

using the trained CNN model to identify gender and age group of the test image and classifying the test image according to the gender and age group; and

extracting a tongue image by using the trained LNMF model trained with age group, gender, and tongue.