DESCRIPTION

METHOD FOR TRACKING PINCHED FINGERTIPS BASED ON VIDEO

FIELD OF THE INVENTION

The present invention relates to the field of image processing and analysis technologies, and more particularly to a method for tracking pinched fingertips based on a video.

BACKGROUND OF THE INVENTION

In recent years, with the rapid development of virtual reality and augmented reality, human-computer interaction technology has been greatly developed and attracted the attention of more and more researchers. Currently, as one of the core technologies of human-computer interaction, finger detection and tracking is widely used in various fields, such as gesture recognition, in-air writing, and the like. Among them, video-based in-air writing has been widely used in the field of character recognition, locus authentication and the like in a in-air writing system.

In a traditional in air-writing system, the situation of human hand writing is simulated mainly by extracting the movement locus of a single fingertip in front of the camera. However, this way of in-air writing does not really simulate the actual situation of human hand writing, because when a person writes with a pen, he/she usually uses three fingers to pinch the pen instead of only a single finger. Therefore, the use of a single finger to simulate the situation of human hand writing not only fails to correctly reflect the habit of writing by a human hand, but also greatly limits the flexibility of human hand writing, resulting in a large difference between the in-air writing locus and the actual handwriting.

Compared with the in-air writing by a single finger, the in-air writing by pinched fingertips obviously more closely simulates the actual gesture of a person writing with a pen, and more realistically reflects the actual situation of hand writing. However, the

method for tracking recognition of the pinched fingertips is still in the blank, the shape of the pinched fingertips is complicated, and the tracking is more difficult. Therefore, it is necessary to design a method for tracking pinched fingertips, so that the method of writing with pinched fingertips can be applied to the in-air writing system.

CONTENTS OF THE INVENTION

An object of the present invention is to overcome the shortcomings and deficiencies in the prior art, and to provide a method for tracking pinched fingertips based on a video, which can extract effective skin colour and texture features to achieve effective tracking recognition of the pinched fingertips.

In order to achieve the above object, the present invention adopts the following technical solution: A method for tracking pinched fingertips based on a video is provided for continuously tracking pinched fingers in in-air writing, and characterized by comprising the following steps:

step 1: a video frame image is captured and converted from an RGB image to a YCbCr image; the YCbCr image is segmented by a YCbCr skin colour detector to obtain a skin colour area, which is segmented by a fist detector to obtain a fist area; a pinched fingertip cascade classifier is used in the fist area to position a pinched fingertip area, so as to determine the initial position of the pinched fingertips;

step 2: a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image; a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; an LK optical flow method is used to match the contour points of the previously and currently processed video frame images to determine the movement

distance of the skin colour target search area;

step 3: a pinched fingertip skin colour mapping table is constructed; with the pinched fingertip area of the previous video frame image divided into sixteen area units, the closest values of the CbCr colour band of the respective area units are respectively searched in the pinched fingertip skin colour mapping table, and a CbCr-Gray colour feature is obtained in combination with a gray value of the respective area units; the HOG feature of the pinched fingertip area of the previous video frame image is calculated, thereby obtaining a CbCr-Gray-HOG feature of the pinched fingertip area;

step 4: a filter tracking method based on a Gaussian kernel function is used in combination with the CbCr-Gray-HOG feature of the pinched fingertip area to determine a central point of the pinched fingertip area of a currently processed video frame image; and

step 5: it is judged whether the condition for ending the tracking of the pinched fingertips is established: if it is established, the tracking ends; otherwise, the process skips to step 2.

Step 1 (i.e. a video frame image is captured and converted from an RGB image to a YCbCr image; the YCbCr image is segmented by a YCbCr skin colour detector to obtain a skin colour area, which is segmented by a fist detector to obtain a fist area; a pinched fingertip cascade classifier is used in the fist area to position a pinched fingertip area, so as to determine the initial position of the pinched fingertips) comprises the following steps:

(1.1) a plurality of consecutive video frame images are captured;

(1.2) the video frame images are converted from an RGB image to a YCbCr image, respectively;

(1.3) a YCbCr skin colour filter is constructed according to the clustering characteristics of human skin colour and used to determine the skin colour area in the respective YCbCr images, and a non-skin colour area is eliminated to eliminate the interference of illumination and the non-skin colour area;

(1.4) the fist classifier based on the Haar-like feature is trained offline, and used

to position the fist area in the respective skin colour areas;

(1.5) the pinched fingertip cascade classifier based on the HOG feature is trained offline, and used to position the pinched fingertip area in the respective fist areas; and

(1.6) thresholds T1 and T2 are set, and the maximum changes in the abscissa and ordinate values of the central point of the respective pinched fingertip areas are judged: if the maximum change in the abscissa value is less than or equal to T1, and the maximum change in the ordinate value is less than or equal to T2, the central point of the pinched fingertip area of the last video frame image in the plurality of consecutive video frame images is set as the initial position of the pinched fingertips, and the process skips to step 2; otherwise, a plurality of subsequent consecutive video frame images are captured and the process skips to step (1.2).

Preferably, constructing a YCbCr skin colour filter according to the clustering characteristics of human skin colour as described in step (1.3) means that the YCbCr image is separated into a Y channel, a Cb channel and a Cr channel, and an area where the Cb channel value is within the colour band range of [77, 127] and the Cr channel value is within the colour band range of [133, 173] is selected as the skin colour area.

In step (1.6), the threshold T1 is set to be one tenth of the width of the pinched fingertip area, and the threshold T2 is set to be one tenth of the height of the pinched fingertip area.

Step 2 (i.e. a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image; a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; an LK optical flow method is used to match the contour points of the previously and currently processed video frame images to determine the movement

distance of the skin colour target search area) comprises the following steps:

(2.1) a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image;

(2.2) a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; and

(2.3) an LK optical flow method is used to match the contour points of the previously and currently processed video frame images, and the movement distances are calculated for the respective corresponding contour points in the previously and currently processed video frame images, with the average or intermediate value of the movement distances defined as the movement distance of the skin colour target search area.

Step 3 (i.e. a pinched fingertip skin colour mapping table is constructed; with the pinched fingertip area of the previous video frame image divided into sixteen area units, the closest values of the CbCr colour band of the respective area units are respectively searched in the pinched fingertip skin colour mapping table, and a CbCr-Gray colour feature is obtained in combination with a gray value of the respective area units; the HOG feature of the pinched fingertip area of the previous video frame image is calculated, thereby obtaining a CbCr-Gray-HOG feature of the pinched fingertip area) comprises the following steps:

(3.1) a pinched fingertip skin colour mapping table is formed by the skin colours with the selected sixteen different colour bands based on the clustering characteristics of human skin colour;

(3.2) the pinched fingertip area of the previous video frame image is divided into sixteen area units, which are mapped into the skin colour mapping table to find the

closest value of the respective area unit CbCr colour bands in the pinched fingertip skin colour mapping table, and the gray value of the respective area units is calculated to obtain the CbCr-Gray colour feature of the sixteen-dimensional pinched fingertip area;

(3.3) the HOG feature of the pinched fingertip area of the previous video frame image in nine gradient directions is calculated, with each of the gradient directions including three dimensions, thereby obtaining the HOG feature of twenty-seven dimensions in total; in combination with the HOG feature based on five dimensions, with the one-dimensional redundant HOG feature eliminated whose gradient values are all zero, the HOG feature of the thirty-one-dimensional pinched fingertip area is constituted; and

(3.4) the CbCr-Gray colour feature of the sixteen-dimensional pinched fingertip area is combined with the HOG feature of the thirty-one-dimensional pinched fingertip area to obtain the CbCr-Gray-HOG feature of the forty-seven-dimensional pinched fingertip area.

Step 4 (i.e. a filter tracking method based on a Gaussian kernel function is used in combination with the CbCr-Gray-HOG feature of the pinched fingertip area to determine a central point of the pinched fingertip area of a currently processed video frame image) comprises the following steps:

(4.1) a sampling area is delimited on the currently processed video frame image by using a point obtained by shifting the central point of the pinched fingertip area of the previous video frame image by the movement distance of the skin colour target search area as a center; in the sampling area, a pinched fingertip search area having a size of $M \times N$ is established centering on the central point of the pinched fingertip area of the previous video frame image; the pinched fingertip search area is cyclically sampled by a cyclic matrix to obtain a plurality of pinched fingertip candidate samples $p_{m,n}$, wherein m and n respectively refer to the abscissa and ordinate values $m \in \{0,...,M-1\}$ and $n \in \{0,...,N-1\}$ of the pinched fingertip candidate sample $p_{m,n}$; and

(4.2) a Gaussian function is used to mark the respective pinched fingertip

candidate samples $p_{m,n}$; and equations (1), (2), (3), and (4) are used to calculate the optimal loss function for the respective pinched fingertip candidate samples $p_{m,n}$:

$$\min_{w} \sum_{m,n} \left| \left\langle \phi\left(p_{m,n}\right), w \right\rangle - Gaussian\left(m,n\right) \right|^2 + \lambda \left\| w \right\|^2 \quad (1)$$

$$w = \sum\nolimits_{m,n} \alpha\left(m,n\right) \phi\left(p_{m,n}\right) \quad (2)$$

$$\alpha = F^{-1}\left(\frac{F\left(y\right)}{F\left(\phi\left(p\right) \cdot \phi\left(p\right)\right) + \lambda} \right) \quad (3)$$

$$\left\langle \phi\left(p\right), \phi\left(p'\right) \right\rangle = \exp\left(-\frac{1}{\sigma^2} \left(\left\| p \right\|^2 + \left\| p' \right\|^2 - 2\mathcal{F}^{-1}\left(\sum_{c} \left(\mathcal{F}\left(p_c\right) \right)^{*} \odot \mathcal{F}\left(p'_c\right) \right) \right) \right) \quad (4)$$

Where ϕ is a Hilbert space mapping function, λ is a regularization parameter, w is a solution in closed form, α is a coefficient, F and F^{-1} are discrete Fourier transform and inverse discrete Fourier transform, respectively, y is a regression function labeling image block matrix, * is a conjugate complex number, p is an image block matrix expressed by the CbCr-Gray-HOG feature of the pinched fingertip area, p' is the transpose of the feature image block matrix p, p_c is a feature image block of the c-th dimension in the feature image block matrix p, σ^2 is the variance in the Gaussian function, and $\odot$ is the multiplication between the elements of the Fourier domain.

The minimum value calculated by equation (1) is used to obtain the minimum value of the loss, and then the coefficient α corresponding to the minimum value of the loss is obtained and substituted into equation (5):

$$\hat{y} = \mathcal{F}^{-1}\left(\mathcal{F}\left(\alpha\right) \odot \mathcal{F}\left(\phi\left(z\right)\right) \right) \quad (5)$$

Where z is an image block of the pinched fingertip search area, and $\hat{y}$ is a response value after the kernel-related convolution;

The respective pinched fingertip candidate samples $p_{m,n}$ are substituted into equation (5) to calculate the corresponding $\hat{y}$ thereof, and the pinched fingertip

candidate sample $p_{m,n}$ corresponding to the minimum value of $\hat{y}$ is used as the pinched fingertip area of the currently processed video frame image, with the point (m,n) corresponding to the pinched fingertip candidate sample as the central point of the pinched fingertip area of the currently processed video frame image.

Judging whether the condition for ending the tracking of the pinched fingertips is established as described in step 5 means that the threshold values T3 and T4 are set, and it is judged whether changes in the abscissa and ordinate values of the central point of the pinched fingertip area of adjacent two video frame images in the plurality of consecutive video frame images are smaller than T3 and T4, respectively: if so, the condition for ending the tracking of the pinched fingertips is established, otherwise it is not established.

The threshold T3 is set to be one tenth of the width of the pinched fingertip area, and the threshold T4 is set to be one tenth of the height of the pinched fingertip area.

Compared with the prior art, the present invention has the following advantages and benefits:

1. The method for tracking pinched fingertips of the present invention is applied to an in-air writing system and can accurately recognize a movement locus of pinched fingertips, thereby realizing the in-air writing system by means of the pinched fingertips; different from the single fingertip in-air writing, the pinched fingertip in-air writing more closely simulates the actual gesture of a person writing with a pen, retaining the flexibility of the hand writing locus, so that the video-based in-air writing system truly reflects the actual hand writing situation.

2. The method for tracking pinched fingertips of the present invention uses a YCbCr skin colour detector, a fist detector based on a Haar-like feature and a pinched fingertip cascade classifier based on an HOG feature in the initialization stage to form a multi-component related detector, accurately positioning the pinched fingertips from complex scenes; then the YCbCr skin colour detector is used to extract the skin colour target search area, a Canny edge detector is used to acquire the contour and sample the contour points, and an LK optical flow method is used to match previous and next

frames of the contour points, thereby determining a movement distance of the skin colour target search area; afterwards, the pinched fingertip skin colour mapping table is constructed, the pinched fingertip area is mapped into the pinched fingertip skin colour mapping table, the CbCr-Gray colour feature is obtained in combination with a gray value, and the HOG feature of the pinched fingertip area is calculated, thereby obtaining the CbCr-Gray-HOG feature of the pinched fingertip area; finally, based on the CbCr-Gray-HOG feature of the pinched fingertip area, a filter tracking method based on a Gaussian kernel function is used for positioning to determine a central point of the pinched fingertip area of a currently processed video frame image, with effective tracking achievable by extracting effective skin colour and texture features.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram showing the simulation of the actual hand writing situation by the method for tracking pinched fingertips of the present invention;

Fig. 2 is a flow chart showing the method for tracking pinched fingertips of the present invention and its application in a in-air signature identity authentication system;

Fig. 3 is a flow chart of the method for detecting the initial position of the pinched fingertips in the initialization stage of the first step of the method for tracking pinched fingertips of the present invention; and

Fig. 4 is a schematic diagram showing the process of constructing a skin colour mapping table and extracting multidimensional features in the third step of the method for tracking pinched fingertips of the present invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

The present invention will be further described below in detail with reference to drawings and specific embodiments.

Example

As shown in Figs. 1-4, a method for tracking pinched fingertips based on a video of the present invention is used for continuously tracking pinched fingers in in-air

writing, and comprises the following five steps:

step 1: a video frame image is captured and converted from an RGB image to a YCbCr image; the YCbCr image is segmented by a YCbCr skin colour detector to obtain a skin colour area, which is segmented by a fist detector to obtain a fist area; a pinched fingertip cascade classifier is used in the fist area to position a pinched fingertip area, so as to determine the initial position of the pinched fingertips;

step 2: a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image; a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; an LK optical flow method is used to match the contour points of the previously and currently processed video frame images to determine the movement distance of the skin colour target search area;

step 3: a pinched fingertip skin colour mapping table is constructed; with the pinched fingertip area of the previous video frame image divided into sixteen area units, the closest values of the CbCr colour band of the respective area units are respectively searched in the pinched fingertip skin colour mapping table, and a CbCr-Gray colour feature is obtained in combination with a gray value of the respective area units; the HOG feature of the pinched fingertip area of the previous video frame image is calculated, thereby obtaining a CbCr-Gray-HOG feature of the pinched fingertip area;

step 4: a filter tracking method based on a Gaussian kernel function is used in combination with the CbCr-Gray-HOG feature of the pinched fingertip area to determine a central point of the pinched fingertip area of a currently processed video frame image; and

step 5: it is judged whether the condition for ending the tracking of the pinched

fingertips is established: if it is established, the tracking ends; otherwise, the process skips to step 2.

Step 1 (i.e. a video frame image is captured and converted from an RGB image to a YCbCr image; the YCbCr image is segmented by a YCbCr skin colour detector to obtain a skin colour area, which is segmented by a fist detector to obtain a fist area; a pinched fingertip cascade classifier is used in the fist area to position a pinched fingertip area, so as to determine the initial position of the pinched fingertips) comprises the following steps:

(1.1) a plurality of consecutive video frame images are captured, with the number of frames set by the designer;

(1.2) the video frame images are converted from an RGB image to a YCbCr image, respectively;

(1.3) a YCbCr skin colour filter is constructed according to the clustering characteristics of human skin colour and used to determine the skin colour area in the respective YCbCr images, and a non-skin colour area is eliminated to eliminate the interference of illumination and the non-skin colour area;

(1.4) the fist classifier based on the Haar-like feature is trained offline, and used to position the fist area in the respective skin colour areas;

(1.5) the pinched fingertip cascade classifier based on the HOG feature is trained offline, and used to position the pinched fingertip area in the respective fist areas; and

(1.6) thresholds T1 and T2 are set, and the maximum changes in the abscissa and ordinate values of the central point of the respective pinched fingertip areas are judged: if the maximum change in the abscissa value is less than or equal to T1, and the maximum change in the ordinate value is less than or equal to T2, the central point of the pinched fingertip area of the last video frame image in the plurality of consecutive video frame images is set as the initial position of the pinched fingertips, and the process skips to step 2; otherwise, a plurality of subsequent consecutive video frame images are captured and the process skips to step (1.2).

The first step is an initialization stage, and the method for tracking pinched

fingertips based on a video of the present invention takes whether the pinched fingertips stay at the same position for a certain period of time as a judgment condition for the start of the in-air writing tracking; since there may be slight jitter in the hand of the writer and there may be an error in image recognition, whether the pinched fingertips are at the same position is determined according to whether the changes in the abscissa and ordinate values of the central point of the pinched fingertip area are within a certain range.

Preferably, constructing a YCbCr skin colour filter according to the clustering characteristics of human skin colour as described in step (1.3) means that the YCbCr image is separated into a Y channel, a Cb channel and a Cr channel, and an area where the Cb channel value is within the colour band range of [77, 127] and the Cr channel value is within the colour band range of [133, 173] is selected as the skin colour area.

Training offline the fist classifier based on the Haar-like feature as described in step (1.4) means that an image including a fist is used as a positive sample and an image not including the fist is used as a negative sample, and the Haar-like feature is used to study the positive and negative samples to generate a fist classifier, with the recognition target of the fist classifier being a fist.

Training offline the pinched fingertip cascade classifier based on the HOG feature as described in step (1.5) means that an image including pinched fingertips is used as a positive sample and an image not including the pinched fingertips is used as a negative sample, and the HOG feature is used to study the positive and negative samples to generate a pinched fingertip cascade classifier, with the recognition target of the pinched fingertip cascade classifier being the pinched fingertips.

In step (1.6), the threshold T1 is set to be one tenth of the width of the pinched fingertip area, and the threshold T2 is set to be one tenth of the height of the pinched fingertip area.

Step 2 (i.e. a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous

video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image; a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; an LK optical flow method is used to match the contour points of the previously and currently processed video frame images to determine the movement distance of the skin colour target search area) comprises the following steps:

(2.1) a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image;

(2.2) a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; and

(2.3) an LK optical flow method is used to match the contour points of the previously and currently processed video frame images, and the movement distances are calculated for the respective corresponding contour points in the previously and currently processed video frame images, with the average or intermediate value of the movement distances defined as the movement distance of the skin colour target search area.

Step 3 (i.e. a pinched fingertip skin colour mapping table is constructed; with the pinched fingertip area of the previous video frame image divided into sixteen area units, the closest values of the CbCr colour band of the respective area units are respectively searched in the pinched fingertip skin colour mapping table, and a CbCr-Gray colour feature is obtained in combination with a gray value of the respective area units; the HOG feature of the pinched fingertip area of the previous video frame image is

calculated, thereby obtaining a CbCr-Gray-HOG feature of the pinched fingertip area) comprises the following steps:

(3.1) a pinched fingertip skin colour mapping table is formed by the skin colours with the selected sixteen different colour bands based on the clustering characteristics of human skin colour;

(3.2) the pinched fingertip area of the previous video frame image is divided into sixteen area units, which are mapped into the skin colour mapping table to find the closest value of the respective area unit CbCr colour bands in the pinched fingertip skin colour mapping table, and the gray value of the respective area units is calculated to obtain the CbCr-Gray colour feature of the sixteen-dimensional pinched fingertip area;

(3.3) the HOG feature of the pinched fingertip area of the previous video frame image in nine gradient directions is calculated, with each of the gradient directions including three dimensions, thereby obtaining the HOG feature of twenty-seven dimensions in total; in combination with the HOG feature based on five dimensions, with the one-dimensional redundant HOG feature eliminated whose gradient values are all zero, the HOG feature of the thirty-one-dimensional pinched fingertip area is constituted; and

(3.4) the CbCr-Gray colour feature of the sixteen-dimensional pinched fingertip area is combined with the HOG feature of the thirty-one-dimensional pinched fingertip area to obtain the CbCr-Gray-HOG feature of the forty-seven-dimensional pinched fingertip area.

Forming a pinched fingertip skin colour mapping table by the skin colours with the selected sixteen different colour bands as described in step (3.1) means that the colours of ivory white, linen, lemon yellow and brown are selected as the four basic skin colour values, and three skin colour values in the vicinity of the respective skin colour values are selected to form skin colours with totally sixteen different colour bands to constitute a pinched fingertip skin colour mapping table.

Step 4 (i.e. a filter tracking method based on a Gaussian kernel function is used in combination with the CbCr-Gray-HOG feature of the pinched fingertip area to

determine a central point of the pinched fingertip area of a currently processed video frame image) comprises the following steps:

(4.1) a sampling area is delimited on the currently processed video frame image by using a point obtained by shifting the central point of the pinched fingertip area of the previous video frame image by the movement distance of the skin colour target search area as a center; in the sampling area, a pinched fingertip search area having a size of $M \times N$ is established centering on the central point of the pinched fingertip area of the previous video frame image; the pinched fingertip search area is cyclically sampled by a cyclic matrix to obtain a plurality of pinched fingertip candidate samples $p_{m,n}$, wherein m and n respectively refer to the abscissa and ordinate values $m \in \{0,...,M-1\}$ and $n \in \{0,...,N-1\}$ of the pinched fingertip candidate sample $p_{m,n}$; and

(4.2) a Gaussian function is used to mark the respective pinched fingertip candidate samples $p_{m,n}$; and equations (1), (2), (3), and (4) are used to calculate the optimal loss function for the respective pinched fingertip candidate samples $p_{m,n}$:

$$\min_{w} \sum_{m,n} \left| \left\langle \phi\left(p_{m,n}\right), w \right\rangle - Gaussian(m,n) \right|^2 + \lambda \|w\|^2 \quad (1)$$

$$w = \sum_{m,n} \alpha(m,n) \phi\left(p_{m,n}\right) \quad (2)$$

$$\alpha = F^{-1}\left(\frac{F(y)}{F(\phi(p) \cdot \phi(p)) + \lambda} \right) \quad (3)$$

$$\left\langle \phi(p), \phi(p') \right\rangle = \exp\left(-\frac{1}{\sigma^2} \left(\|p\|^2 + \|p'\|^2 - 2\mathcal{F}^{-1}\left(\sum_{c} \left(\mathcal{F}(p_c) \right)^* \odot \mathcal{F}(p'_c) \right) \right) \right) \quad (4)$$

Where ϕ is a Hilbert space mapping function, λ is a regularization parameter, w is a solution in closed form, α is a coefficient, F and F^{-1} are discrete Fourier transform and inverse discrete Fourier transform, respectively, y is a regression function labeling image block matrix, * is a conjugate complex number, p is an image block matrix expressed by the CbCr-Gray-HOG feature of the pinched fingertip area, p' is the transpose of the feature image block matrix p, p_c is a feature image block of

the c-th dimension in the feature image block matrix p, σ^2 is the variance in the Gaussian function, and $\odot$ is the multiplication between the elements of the Fourier domain.

The minimum value calculated by equation (1) is used to obtain the minimum value of the loss, and then the coefficient α corresponding to the minimum value of the loss is obtained and substituted into equation (5):

$$\hat{y} = \mathcal{F}^{-1}\left(\mathcal{F}(\alpha) \odot \mathcal{F}(\phi(z))\right) \quad (5)$$

Where z is an image block of the pinched fingertip search area, and $\hat{y}$ is a response value after the kernel-related convolution;

The respective pinched fingertip candidate samples $p_{m,n}$ are substituted into equation (5) to calculate the corresponding $\hat{y}$ thereof, and the pinched fingertip candidate sample $p_{m,n}$ corresponding to the minimum value of $\hat{y}$ is used as the pinched fingertip area of the currently processed video frame image, with the point (*m, n*) corresponding to the pinched fingertip candidate sample as the central point of the pinched fingertip area of the currently processed video frame image.

Judging whether the condition for ending the tracking of the pinched fingertips is established as described in step 5 means that the threshold values T3 and T4 are set, and it is judged whether changes in the abscissa and ordinate values of the central point of the pinched fingertip area of adjacent two video frame images in the plurality of consecutive video frame images are smaller than T3 and T4, respectively: if so, the condition for ending the tracking of the pinched fingertips is established, otherwise it is not established.

The method for tracking pinched fingertips based on a video of the present invention takes the pinched fingertips staying in a small area of the same position for a certain period of time as a criterion for judging the end of tracking.

The threshold T3 is set to be one tenth of the width of the pinched fingertip area, and the threshold T4 is set to be one tenth of the height of the pinched fingertip area.

The method for tracking pinched fingertips of the present invention is applied to

an in-air writing system and can accurately recognize a movement locus of pinched fingertips, thereby realizing the in-air writing system by means of the pinched fingertips; different from the single fingertip in-air writing, the pinched fingertip in-air writing more closely simulates the actual gesture of a person writing with a pen, retaining the flexibility of the hand writing locus, so that the video-based in-air writing system truly reflects the actual hand writing situation.

The above examples are preferred embodiments of the present invention, but the embodiments of the present invention are not limited thereto, and any other alterations, modifications, substitutions, combinations and simplifications should be equivalent replacements and included in the scope of protection of the present invention.

CLAIMS

1. A method for tracking pinched fingertips based on a video, used for continuously tracking pinched fingers in in-air writing, characterized in that: this method comprises the following five steps:

step 1: a video frame image is captured and converted from an RGB image to a YCbCr image; the YCbCr image is segmented by a YCbCr skin colour detector to obtain a skin colour area, which is segmented by a fist detector to obtain a fist area; a pinched fingertip cascade classifier is used in the fist area to position a pinched fingertip area, so as to determine the initial position of the pinched fingertips;

step 2: a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image; a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; an LK optical flow method is used to match the contour points of the previously and currently processed video frame images to determine the movement distance of the skin colour target search area;

step 3: a pinched fingertip skin colour mapping table is constructed; with the pinched fingertip area of the previous video frame image divided into sixteen area units, the closest values of the CbCr colour band of the respective area units are respectively searched in the pinched fingertip skin colour mapping table, and a CbCr-Gray colour feature is obtained in combination with a gray value of the respective area units; the HOG feature of the pinched fingertip area of the

previous video frame image is calculated, thereby obtaining a CbCr-Gray-HOG feature of the pinched fingertip area;

step 4: a filter tracking method based on a Gaussian kernel function is used in combination with the CbCr-Gray-HOG feature of the pinched fingertip area to determine a central point of the pinched fingertip area of a currently processed video frame image; and

step 5: it is judged whether the condition for ending the tracking of the pinched fingertips is established: if it is established, the tracking ends; otherwise, the process skips to step 2.

2. The method for tracking pinched fingertips based on a video according to claim 1, characterized in that: step 1 (i.e. a video frame image is captured and converted from an RGB image to a YCbCr image; the YCbCr image is segmented by a YCbCr skin colour detector to obtain a skin colour area, which is segmented by a fist detector to obtain a fist area; a pinched fingertip cascade classifier is used in the fist area to position a pinched fingertip area, so as to determine the initial position of the pinched fingertips) comprises the following steps:

(1.1) a plurality of consecutive video frame images are captured;

(1.2) the video frame images are converted from an RGB image to a YCbCr image, respectively;

(1.3) a YCbCr skin colour filter is constructed according to the clustering characteristics of human skin colour and used to determine the skin colour area in the respective YCbCr images, and a non-skin colour area is eliminated to eliminate the interference of illumination and the non-skin colour area;

(1.4) the fist classifier based on the Haar-like feature is trained offline, and used to position the fist area in the respective skin colour areas;

(1.5) the pinched fingertip cascade classifier based on the HOG feature is trained offline, and used to position the pinched fingertip area in the respective fist areas; and

(1.6) thresholds T1 and T2 are set, and the maximum changes in the abscissa

and ordinate values of the central point of the respective pinched fingertip areas are judged: if the maximum change in the abscissa value is less than or equal to T1, and the maximum change in the ordinate value is less than or equal to T2, the central point of the pinched fingertip area of the last video frame image in the plurality of consecutive video frame images is set as the initial position of the pinched fingertips, and the process skips to step 2; otherwise, a plurality of subsequent consecutive video frame images are captured and the process skips to step (1.2).

3. The method for tracking pinched fingertips based on a video according to claim 2, characterized in that: constructing a YCbCr skin colour filter according to the clustering characteristics of human skin colour as described in step (1.3) means that the YCbCr image is separated into a Y channel, a Cb channel and a Cr channel, and an area where the Cb channel value is within the colour band range of [77, 127] and the Cr channel value is within the colour band range of [133, 173] is selected as the skin colour area.

4. The method for tracking pinched fingertips based on a video according to claim 2, characterized in that: in step (1.6), the threshold T1 is set to be one tenth of the width of the pinched fingertip area, and the threshold T2 is set to be one tenth of the height of the pinched fingertip area.

5. The method for tracking pinched fingertips based on a video according to claim 1, characterized in that: step 2 (i.e. a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image; a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point;

an LK optical flow method is used to match the contour points of the previously and currently processed video frame images to determine the movement distance of the skin colour target search area) comprises the following steps:

(2.1) a next video frame image is captured as the currently processed one; a target search area is established on the currently processed video frame image by centering on the central point of the pinched fingertip area obtained from the previous video frame image; the YCbCr skin colour detector is used to extract a skin colour area of the target search area, thereby obtaining a skin colour target search area of the currently processed video frame image;

(2.2) a Canny edge detector is used to respectively process the skin colour target search areas of the previously and currently processed video frame images to obtain a contour, and a dense sampling method is used to obtain a contour point; and

(2.3) an LK optical flow method is used to match the contour points of the previously and currently processed video frame images, and the movement distances are calculated for the respective corresponding contour points in the previously and currently processed video frame images, with the average or intermediate value of the movement distances defined as the movement distance of the skin colour target search area.

6. The method for tracking pinched fingertips based on a video according to claim 1, characterized in that: step 3 (i.e. a pinched fingertip skin colour mapping table is constructed; with the pinched fingertip area of the previous video frame image divided into sixteen area units, the closest values of the CbCr colour band of the respective area units are respectively searched in the pinched fingertip skin colour mapping table, and a CbCr-Gray colour feature is obtained in combination with a gray value of the respective area units; the HOG feature of the pinched fingertip area of the previous video frame image is calculated, thereby obtaining a CbCr-Gray-HOG feature of the pinched fingertip area) comprises the following steps:

(3.1) a pinched fingertip skin colour mapping table is formed by the skin colour of the selected sixteen different colour bands based on the clustering characteristics of human skin colour;

(3.2) the pinched fingertip area of the previous video frame image is divided into sixteen area units, which are mapped into the skin colour mapping table to find the closest value of the respective area unit CbCr colour bands in the pinched fingertip skin colour mapping table, and the gray value of the respective area units is calculated to obtain the CbCr-Gray colour feature of the sixteen-dimensional pinched fingertip area;

(3.3) the HOG feature of the pinched fingertip area of the previous video frame image in nine gradient directions is calculated, with each of the gradient directions including three dimensions, thereby obtaining the HOG feature of twenty-seven dimensions in total; in combination with the HOG feature based on five dimensions, with the one-dimensional redundant HOG feature eliminated whose gradient values are all zero, the HOG feature of the thirty-one-dimensional pinched fingertip area is constituted; and

(3.4) the CbCr-Gray colour feature of the sixteen-dimensional pinched fingertip area is combined with the HOG feature of the thirty-one-dimensional pinched fingertip area to obtain the CbCr-Gray-HOG feature of the forty-seven-dimensional pinched fingertip area.

7. The method for tracking pinched fingertips based on a video according to claim 6, characterized in that: step 4 (i.e. a filter tracking method based on a Gaussian kernel function is used in combination with the CbCr-Gray-HOG feature of the pinched fingertip area to determine a central point of the pinched fingertip area of a currently processed video frame image) comprises the following steps:

(4.1) a sampling area is delimited on the currently processed video frame image by using a point obtained by shifting the central point of the pinched fingertip area of the previous video frame image by the movement distance of the

skin colour target search area as a center; in the sampling area, a pinched fingertip search area having a size of $M \times N$ is established centering on the central point of the pinched fingertip area of the previous video frame image; the pinched fingertip search area is cyclically sampled by a cyclic matrix to obtain a plurality of pinched fingertip candidate samples $p_{m,n}$, wherein m and n respectively refer to the abscissa and ordinate values $m \in \{0,...,M-1\}$ and $n \in \{0,...,N-1\}$ of the pinched fingertip candidate sample $p_{m,n}$; and

(4.2) a Gaussian function is used to mark the respective pinched fingertip candidate samples $p_{m,n}$; and equations (1), (2), (3), and (4) are used to calculate the optimal loss function for the respective pinched fingertip candidate samples $p_{m,n}$:

$$\min_{w} \sum_{m,n} \left| \left\langle \phi\left(p_{m,n}\right), w \right\rangle - Gaussian(m,n) \right|^2 + \lambda \|w\|^2 \quad (1)$$

$$w = \sum\nolimits_{m,n} \alpha(m,n) \phi\left(p_{m,n}\right) \quad (2)$$

$$\alpha = F^{-1}\left(\frac{F(y)}{F\left(\phi(p) \cdot \phi(p)\right) + \lambda} \right) \quad (3)$$

$$\left\langle \phi(p), \phi(p') \right\rangle = \exp\left(-\frac{1}{\sigma^2} \left(\|p\|^2 + \|p'\|^2 - 2\mathcal{F}^{-1}\left(\sum_c \left(\mathcal{F}(p_c)\right)^* \odot \mathcal{F}(p'_c) \right) \right) \right) \quad (4)$$

where ϕ is a Hilbert space mapping function, λ is a regularization parameter, w is a solution in closed form, α is a coefficient, F and F^{-1} are discrete Fourier transform and inverse discrete Fourier transform, respectively, y is a regression function labeling image block matrix, * is a conjugate complex number, p is an image block matrix expressed by the CbCr-Gray-HOG feature of the pinched fingertip area, p' is the transpose of the feature image block matrix p, p_c is a feature image block of the c-th dimension in the feature image block matrix p, σ^2 is the variance in the Gaussian function, and $\odot$ is the multiplication between the elements of the Fourier domain;

the minimum value calculated by equation (1) is used to obtain the minimum value of the loss, and then the coefficient α corresponding to the minimum value of the loss is obtained and substituted into equation (5):

$$\hat{y} = \mathcal{F}^{-1}\left(\mathcal{F}(\alpha) \odot \mathcal{F}(\phi(z))\right) \quad (5)$$

where z is an image block of the pinched fingertip search area, and $\hat{y}$ is a response value after the kernel-related convolution;

the respective pinched fingertip candidate samples $p_{m,n}$ are substituted into equation (5) to calculate the corresponding $\hat{y}$ thereof, and the pinched fingertip candidate sample $p_{m,n}$ corresponding to the minimum value of $\hat{y}$ is used as the pinched fingertip area of the currently processed video frame image, with the point (*m, n*) corresponding to the pinched fingertip candidate sample as a central point of the pinched fingertip area of the currently processed video frame image.

8. The method for tracking pinched fingertips based on a video according to claim 1, characterized in that: judging whether the condition for ending the tracking of the pinched fingertips is established as described in step 5 means that the threshold values T3 and T4 are set, and it is judged whether changes in the abscissa and ordinate values of the central point of the pinched fingertip area of adjacent two video frame images in the plurality of consecutive video frame images are smaller than T3 and T4, respectively: if so, the condition for ending the tracking of the pinched fingertips is established, otherwise it is not established.

9. The method for tracking pinched fingertips based on a video according to claim 8, characterized in that: The threshold T3 is set to be one tenth of the width of the pinched fingertip area, and the threshold T4 is set to be one tenth of the height of the pinched fingertip area.