

Description

KEY FRAME ACQUISITION METHOD FOR HUMAN IMAGE VIDEO SYSTEM

TECHNICAL FIELD

The present invention relates to the field of video processing, specifically relates to a key frame acquisition method for human image video system.

BACKGROUND

Face recognition is the core module of a human image video system. In most human image systems, corresponding tracks for a passing person are obtained. The tracks of a single person can be expressed as a set of rectangular boxes on K frames. The input of the face recognition module of the human image video system is a person's tracks, and the output is the identity of the person.

Relative to the recognition module of a human image system, the main difference of the human image video system is that it needs to pick faces from suitable frames for recognition. The easiest way is to recognize all K frames, however the system is to be slowed down because the time overhead of the recognition module is very large (1 frame for 1 second, and for a person there are 24 frames per second). Common practice is to choose P frames (P is approximately equal to 3) of the best front faces to ensure speed, but the recognition performance is affected.

Therefore, there is a need to develop a new method that can both ensure processing efficiency and affect recognition performance.

SUMMARY OF THE INVENTION

The purpose of the present invention is to overcome the above shortcomings of the prior art and to provide a key frame acquisition method for human image video system. By selecting the key frame and then performing face recognition, the method not only

ensures the speed but also has a higher recognition performance.

The object of the present invention can be achieved by the following technical solutions:

A key frame acquisition method for human image video system comprises the following steps:

- 1) acquiring a video image, and performing key point detection on each frame of the image;
- 2) performing attribute classification on each frame of the image according to detected key points, and calculating confidence under various attributes;
- 3) selecting P key frames according to the confidence of each frame of the image under different attributes as a basis of face recognition.

In the step 1), the key point detection specifically is:

Using a way of combining HoG and SVM to do a first round of the key point detection, and reserving K options for each key point. Then finding the most human face-like shape in K^N possibilities using global shape information as the final key points, wherein N is the number of key points.

The attributes include strong light, dim light, uneven light, blur classification, opened - mouth classification, closed - eyes classification, pose estimation and size.

Calculating the confidence under each attribute specifically is:

- a) Strong light: a strong light confidence S1 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;
- b) Dim light: a dim light confidence S2 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;
- c) Uneven light: an uneven light confidence S3 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;
- d) Blur classification: a blur confidence S4 is obtained by using Boosting classifier, based on LPSS, gradient histogram, color saturation distribution, and local autocorrelation;
- e) Opened-mouth classification: a confidence S5 is obtained by performing opened-mouth classification using a way combined LBP and SVM;

f) Closed-eyes classification: a confidence S6 is obtained by performing closed-eyes classification using a way combined LBP and SVM;

g) Pose estimation: The POSIT method is used to estimate rotation angles of the X axis, Y axis, and Z axis, and the rotation angles of the respective axis are converted to the confidences S7, S8, and S9 using Logistic Function;

h) Size: The result of dividing an actual face size by a standard size is converted into a confidence S10.

The standard size is taken $150 \text{ pixels} \times 150 \text{ pixels}$.

In the step 3), the process of selecting P key frames specifically is:

301) The confidences of each frame under different attributes are combined as the eigenvector (S1k, S2k, ..., S10k) of said frame, k denotes the kth frame;

302) The eigenvector is input into a first SVM classifier, and a Score of each frame is calculated in the first SVM classifier, and a frame with the largest Score is selected as the first key frame;

303) The eigenvector of the frame selected in the previous step and the unselected frame are combined as a new eigenvector of the unselected frame, and then the new eigenvector is input to a second SVM classifier, a Score of each unselected frame in the second SVM classifier is calculated, and the frame with the largest Score is selected as the second key frame;

304) The remaining P-2 frames are selected in the same way as in step 303).

The P is taken as three.

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

(1) In the case of having known key points, the present invention classifies attributes of images, obtains the confidence of each attribute, and selects the key frames according to the confidence, therefore improves the accuracy of the selected key frames;

(2) When selecting the key frames, the present invention not only considers the eigenvector of each frame itself, but also considers the eigenvectors of the selected frames comprehensively, and refers to the collaboration information between different frames, therefore further improves the accuracy of the selected key frames, which ensures the recognition performance;

(3) The present invention performs recognition by selecting the key frames, therefore the recognition speed is guaranteed.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a structure diagram of the present invention.

DETAILED DESCRIPTION OF EMBODIMENTS

The present invention is described in detail below with reference to the drawings and specific embodiments. This embodiment is implemented on the premise of the technical solutions of the present invention, and detailed description and the specific operation process are given. However, the protection scope of the present invention is not limited to the following embodiments.

As shown in FIG. 1, this embodiment provides a key frame acquisition method for human image video system. The method takes the influence of factors among multiple images into account, and also takes the information of collaboration between different frames into account. The specific steps are as follows:

Step S1, a video image is acquired, and a key point detection on each frame of the image is performed. The input of the key point detection process is the position of a face box and the image, and the output is the positions of the key points. A first round of the key point detection is did using HoG + SVM, and K options are reserved for each key point. Then the most human face-like shape in $K \wedge N$ possibilities is found using global shape information and the key points are acquired, wherein N is the number of key points.

Step S2, attribute classification is performed on each frame of the image according to the detected key points, and confidence under each attribute is calculated. The attributes here include strong light, dim light, uneven light, blur classification, opened-mouth classification, closed-eyes classification, pose estimation and size, the specific confidence is calculated as follows:

a) Strong light: a strong light confidence S1 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;

b) Dim light: a dim light confidence S2 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;

c) Uneven light: an uneven light confidence S3 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;

d) Blur classification: a blur confidence S4 is obtained by using Boosting classifier, and the main features used are:

d1) LPSS (Local Power Spectrum Slope): In a blurred image, information in the high frequency area is lost, so the energy occupied in the low frequency area is relatively high. Therefore, the image can be Fourier transformed to obtain the energy of different frequencies as the eigenvector;

d2) Gradient Histogram: There are very few points of strong edges in the blurred image, and the gradient intensity characterizes the edge intensity to some extent. Therefore, using the gradient histogram is also a way to describe blurriness;

d3) Color Saturation Distribution: The color saturation of a sharp image far exceeds that of a blurred image;

d4) Local Autocorrelation: In motion blur, objects and backgrounds are blurred except for the edges along the same direction as the motion. This feature can be described by describing the shape features of spectrum, or by Structural SIMilarity (SSIM).

e) Opened-month classification: a confidence S5 is obtained by performing opened-month classification using a way combined LBP and SVM;

f) Closed-eyes classification: a confidence S6 is obtained by performing closed-eyes classification using a way combined LBP and SVM;

g) Pose estimation: The POSIT method is used to estimate rotation angles of the X axis, Y axis, and Z axis, and the rotation angles of the respective axis are converted to the confidences S7, S8, and S9 using Logistic Function;

h) Size: The result of dividing an actual face size by a standard size is converted into a confidence S10. The standard size is taken 150 pixels $\times$ 150 pixels.

The values of the above confidences are (-1, 1).

Step S3, P key frames, as a basis of face recognition, are selected according to the

confidence of each frame of the image under different attributes.

Taking P as 3 as an example, the specific process of selecting key frames is:

301) The confidences of each frame under different attributes are combined as the eigenvector ($S1k, S2k, \dots, S10k$) of said frame, k denotes the k th frame;

302) The eigenvector is input into a first SVM classifier, and a score (Score) of each frame is calculated in the first SVM classifier, and a frame with the largest Score is selected as the first key frame. At this point, the dimension of the eigenvector is 10;

303) The eigenvector of the frame selected in the previous step and an unselected frame are combined as a new eigenvector of the unselected frame, and then the new eigenvector is input to a second SVM classifier. A score (Score) of each unselected frame in the second SVM classifier is calculated, and the frame with the largest Score is selected as the second key frame. At this point, the dimension of the eigenvector is 20;

304) The third key frame is selected in the same way as in step 303), in which case the dimension of the eigenvector is 30.

In the case of guaranteeing the minimum computation (ordinary 4-core in real-time), this method ensures the best performance.

What is claimed is:

1. A key frame acquisition method for human image video system, is characterized by, comprising the following steps:

1) acquiring a video image, and performing key point detection on each frame of the image;

2) performing attribute classification on each frame of the image according to detected key points, and calculating confidence under various attributes;

3) selecting P key frames according to the confidence of each frame of the image under different attributes as a basis of face recognition.

2. The key frame acquisition method for human image video system according to claim 1, wherein, in the step 1), the key points detection specifically is:

Using a way of combining HoG + SVM to do a first round of the key point detection, and reserving K options for each key point. Then finding the most human face-like shape in K^N possibilities using global shape information as the final key points, wherein N is the number of key points.

3. The key frame acquisition method for human image video system according to claim 1, wherein, the attributes include strong light, dim light, uneven light, blur classification, opened-mouth classification, closed-eyes classification, pose estimation and size.

4. The key frame acquisition method for human image video system according to claim 3, wherein, the calculating the confidence under various attributes specifically is:

a) Strong light: a strong light confidence S1 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;

b) Dim light: a dim light confidence S2 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;

c) Uneven light: an uneven light confidence S3 is obtained by using Boosting classifier, based on the LBP and Gray Grid eigenvectors;

d) Blur classification: a blur confidence S4 is obtained by using Boosting classifier, based on LPSS, gradient histogram, color saturation distribution, and local autocorrelation;

e) Opened-month classification: a confidence S5 is obtained by performing opened-month classification using a way combined LBP and SVM;

f) Closed-eyes classification: a confidence S6 is obtained by performing closed-eyes classification using a way combined LBP and SVM;

g) Pose estimation: The POSIT method is used to estimate rotation angles of the X axis, Y axis, and Z axis, and the rotation angles of the respective axis are converted to the confidences S7, S8, and S9 using Logistic Function;

h) Size: The result of dividing an actual face size by a standard size is converted into a confidence S10.

5. The key frame acquisition method for human image video system according to claim 4, wherein, the standard size is taken $150 \text{ pixels} \times 150 \text{ pixels}$.

6. The key frame acquisition method for human image video system according to claim 4, wherein, in the step 3), the process of selecting P key frames specifically is:

301) The confidences of each frame under different attributes are combined as the eigenvector (S1k, S2k, ..., S10k) of said frame, k denotes the kth frame;

302) the eigenvector is input into a first SVM classifier, and a Score of each frame is calculated in the first SVM classifier, and a frame with the largest Score is selected as the first key frame;

303) the eigenvector of the frame selected in the previous step and an unselected frame are combined as a new eigenvector of the unselected frame, and then the new eigenvector is input to a second SVM classifier. A Score of each unselected frame in the second SVM classifier is calculated, and the frame with the largest Score is selected as the second key frame;

304) the remaining P-2 frames are selected in the same way as in step 303).

7. The key frame acquisition method for human image video system according to claim 6, wherein the P is taken as three.