

Method for face recognition-based video human image tracking under complex scenes

Technical Field

[0001] The present invention relates to the field of video processing technology, especially to a method for face recognition-based video human image tracking under complex scenes.

Background Art

[0002] Face tracking is a process of determining the movement trajectory and size change of a certain face among video or image sequences. Face tracking has always been of great significance in fields of image analysis and recognition as well as image monitoring and retrieving, etc. such as video multimedia message (MMS) in cellphone multimedia messages, human-machine interface, privilege control and smart monitoring system, etc. The accuracy, precision and robustness of tracking have always been major concerns of the industry, and many effective algorithms have appeared successively. Currently, the main video human image tracking algorithms are as follows:

a) face detection-based method, in which the face trajectory is obtained by performing face detection to each frame image of video and then associating faces in two continuous frames according to the relationship in human image positions between the two continuous frames.

b) method based on face detection and object tracking, as an improvement on the previous method, in which face detection is performed to every fixed frame of video and face positions in frames between two continuous fixed frames are obtained through object tracking methods such as optical flow and mean shift, etc.

[0003] Above schemes can deliver good effects under simple scenes, in which faces are clear, mainly frontal, and unshielded. However, they fail to make full use of video information, and face trajectory interruption and mixing-up, etc. are prone to occur under

complex scenes, in which faces are frequently shielded and turned on a large scale, leading to false positive and poor performance.

BRIEF SUMMARY OF THE INVENTION

[0004] The present invention aims to overcome the defects of above prior arts and provide a method for face recognition-based video human image tracking under complex scenes with fast processing speed, high detection precision and low false positive rate.

[0005] The aims of the present invention can be achieved via the following technical solutions:

[0006] A method for face recognition-based video human image tracking under complex scenes, comprising following steps:

- 1) decoding acquired video data to obtain a series of video frame images;
- 2) utilizing a face detection and object tracking method to process the obtained series of video frame images, so as to obtain all possible face trajectories;
- 3) judging similarities between the face trajectories obtained in the step 2) and subsequent adjacent face trajectories, and combining face trajectories having higher similarities than a threshold value;
- 4) outputting the combined face trajectories.

[0007] Step 2) is specifically:

201) grouping the series of video frame images obtained in the step 1) and choosing video frame image every m frames as detection frame F_{km} ;

202) performing face detection to the chosen multiple detection frames F_{km} and marking a face box on each detection frame;

203) choosing $(m-1)$ frames behind the detection frame F_{km} as tracking frames and obtaining subsequent trajectories on the tracking frames relative to a face on the detection frame through an object tracking and local face detection manner, and thus obtaining all

possible face trajectories.

[0008] In the step 203), obtaining the subsequent trajectories on the tracking frames relative to a face on the detection frame through an object tracking and local face detection manner is specifically:

2a) assuming position of a face on the detection frame F_{km} is D_{km} and searching tracking stable points meeting a set condition within region D'_{km} which is 1.5 times bigger than the face box on the detection frame F_{km} ;

2b) obtaining corresponding positions of the tracking stable points within the region D'_{km} on tracking frame F_{km+1} utilizing an optical flow algorithm;

2c) taking average displacement of all the tracking stable points as movement displacement of the face from the detection frame F_{km} to the tracking frame F_{km+1} , and thus obtaining estimated position D'_{km+1} of the face on the tracking frame F_{km+1} ;

2d) performing local face detection on the tracking frame F_{km+1} based on the estimated position D'_{km+1} to obtain accurate position D_{km+1} of the face;

2e) repeating the steps 2a) ~2d) until the face trajectories interrupt or all the detection frames have been detected, and thus obtaining all the possible face trajectories.

[0009] The conditions for the face trajectory interrupting comprise:

I) face position is located on edge of the video frame image;

II) the tracking stable points meeting the set conditions fail to be found;

III) the face fails to be detected at the estimated position of next frame.

[0010] The tracking stable points meeting the set conditions refer to that number of the tracking stable points is greater than 100.

[0011] The subsequent adjacent face trajectory in the step 3) refers to another face trajectory adjacent to current face trajectory in both time and space, thereinto, adjacence in time refers to that start frame of the another face trajectory is within S frames behind end frame of the current face trajectory; adjacence in space refers to that distance between face

on the start frame of the another face trajectory and face on the end frame of the current face trajectory is less than L face heights.

[0012] Step 3) is specifically:

301) extracting LBP characteristics from all the face trajectories obtained in the step 2), and reducing dimension of the LBP characteristics to obtain final recognition characteristics;

302) sequencing all the face trajectories according to sequence of their end frames and calculating the similarities between each of face trajectories and corresponding subsequent adjacent face trajectory in turn:

choosing n frames of faces respectively from the current face trajectory and the subsequent adjacent face trajectory, and cross-contrasting them to obtain n^2 groups of similarities, and choosing the highest group as similarities between the two trajectories, thereinto, the similarities obtained by cross-contrast is obtained by calculating cosine distance between recognition characteristics respectively corresponding to two frame images;

303) combining two groups of trajectories having higher similarities than the set threshold value.

[0013] In the step 301), extracting the recognition characteristics from the face trajectory is specifically:

3a) selecting N frames of faces from the face trajectory randomly;

3b) extracting the LBP characteristics of each frame of face;

3c) utilizing a dimension reduction matrix trained beforehand to reduce the dimension of the LBP characteristics to obtain the final recognition characteristics.

[0014] Compared with the prior arts, the present invention possesses the following advantages:

1) the present invention provides an effective solution for face interruption under

complex scenes resulting from shielding and turning, etc. by calculating the similarities between the current face trajectories and the subsequent adjacent face trajectories and combining two trajectories with a high similarities;

2) when conducting tracking frame face detection, the present invention improves the processing speed of the system and avoids the mixing-up and false positive of face trajectory due to shielding via object tracking and local detection solution .

DETAILED DESCRIPTION OF THE INVENTION

[0015] The present invention will now be described in detail with reference to specific embodiments. The embodiments are implemented premised on the technical solution of the present invention, and although detailed description of the implementation mode and the specific operational process are presented, the present invention is not limited to the embodiment described below.

[0016] The embodiment provides a face recognition-based video human image tracking method under complex scenes. The input of this method is a segment of human image video and the output is several segments of face trajectories with each face trajectory comprising all the frame numbers where the face appears and the face position on the corresponding frames. The specific steps of the method are as follows:

- 1) decoding acquired video data to obtain a series of video frame images;
- 2) utilizing a face detection and object tracking method to process the obtained series of video frame images, so as to obtain all possible face trajectories $\{G_i\}$, specifically:
 - 201) grouping the series of video frame images obtained in the step 1) and choosing video frame image every m frames as detection frame F_{km} , thereinto, m may be 3~5;
 - 202) performing face detection to the chosen multiple detection frames F_{km} and marking a face box on each detection frame;
 - 203) choosing $(m-1)$ frames behind the detection frame F_{km} as tracking frames, i.e.

taking F_{k+m+1} , F_{k+m+2} , ... , F_{k+m-1} as the tracking frames, and obtaining subsequent trajectories on the tracking frames relative to a face on the detection frame through an object tracking and local face detection manner, and thus obtaining all possible face trajectories.

In the step 203), obtaining the subsequent trajectories on each of the tracking frames relative to a face on the detection frame through an object tracking and local face detection manner is specifically:

2a) assuming position of a face on the detection frame F_{km} is D_{km} and searching tracking stable points meeting a set condition within region D'_{km} which is 1.5 times bigger than the face box on the detection frame F_{km} ;

2b) obtaining corresponding positions of the tracking stable points on the tracking frame F_{k+m+1} utilizing an optical flow algorithm;

2c) taking average displacement of all the tracking stable points as movement displacement of the face from the detection frame F_{km} to the tracking frame F_{k+m+1} , and thus obtaining estimated position D'_{k+m+1} of the face on the tracking frame F_{k+m+1} ;

2d) performing local face detection on the tracking frame F_{k+m+1} based on the estimated position D'_{k+m+1} , i.e. within the region near the estimated position D'_{k+m+1} which is generally 3 times bigger than the face box, to obtain accurate position D_{k+m+1} of the face;

2e) repeating the steps 2a) ~2d) until the face trajectories interrupt or all the detection frames have been detected, and thus obtaining all the possible face trajectories.

The conditions for the face trajectory interrupting herein comprise:

I) the face position is located on edge of the video frame image;

II) the tracking stable points meeting the set conditions fail to be found, thereinto, the tracking stable points meeting the set conditions refer to that number of the tracking stable points is greater than 100.

III) the face fails to be detected at the estimated position of next frame;

3) judging similarities between the face trajectories obtained in the step 2) and subsequent adjacent face trajectories, and combining face trajectories having higher similarities than a threshold value;

The subsequent adjacent face trajectory in the step 3) refers to another face trajectory adjacent to current face trajectory in both time and space, thereinto, adjacence in time refers to that start frame of the another face trajectory is within S frames behind end frame of the current face trajectory, and the S frames are 3~6s; adjacence in space refers to that distance between face on the start frame of the another face trajectory and face on the end frame of the current face trajectory is less than L face heights, and L is greater than or equal to 5.

The Step 3) is specifically:

301) extracting LBP characteristics from all the face trajectories obtained in the step 2), and reducing dimension of the LBP characteristics to obtain final recognition characteristics:

3a) selecting N frames of faces from the face trajectory randomly, and N herein may be 3~5;

3b) extracting the LBP characteristics of each frame of face, which is a high-dimensional array;

3c) utilizing a dimension reduction matrix (PCA+LDA) trained beforehand to reduce the dimension of the LBP characteristics to obtain the final recognition characteristics;

302) sequencing all the face trajectories according to sequence of their end frames and calculating the similarities between each of face trajectories and corresponding subsequent adjacent face trajectory in turn:

choosing n frames of faces respectively from the current face trajectory and the subsequent adjacent face trajectory, and cross-contrasting them to obtain n^2 groups of

similarities, and choosing the highest group as similarities between the two trajectories, thereinto, the similarities obtained by cross-contrast is obtained by calculating cosine distance between recognition characteristics respectively corresponding to two frame images, and n may be 5;

303) combining two groups of trajectories having higher similarities than the set threshold value.

[0017] The face trajectory obtained via the above method has a false positive rate of one ten-thousandth and a confidence of 95%.

What is claimed is:

1. A method for face recognition-based video human image tracking under complex scenes, comprising following steps:

- 1) decoding acquired video data to obtain a series of video frame images;
- 2) utilizing a face detection and object tracking method to process the obtained series of video frame images, so as to obtain all possible face trajectories;
- 3) judging similarities between the face trajectories obtained in the step 2) and subsequent adjacent face trajectories, and combining face trajectories having higher similarities than a set threshold value;
- 4) outputting the combined face trajectories.

2. The method for face recognition-based video human image tracking under complex scenes of claim 1, wherein the Step 2) is specifically:

201) grouping the series of video frame images obtained in the step 1) and choosing video frame image every m frames as detection frame F_{km} ;

202) performing face detection to the chosen multiple detection frames F_{km} and marking a face box on each detection frame;

203) choosing $(m-1)$ frames behind the detection frame F_{km} as tracking frames and obtaining subsequent trajectories on the tracking frames relative to a face on the detection frame through an object tracking and local face detection manner, and thus obtaining all the possible face trajectories.

3. The method for face recognition-based video human image tracking under complex scenes of claim 2, wherein in the step 203), obtaining the subsequent trajectories on the tracking frames relative to a face on the detection frame through an object tracking and local face detection manner is specifically:

2a) assuming position of a face on the detection frame F_{km} is D_{km} and searching tracking stable points meeting a set condition within region D'_{km} which is 1.5 times bigger than the face box on the detection frame F_{km} ;

2b) obtaining corresponding positions of the tracking stable points within the region D'_{km} on tracking frame F_{km+1} utilizing an optical flow algorithm;

2c) taking average displacement of all the tracking stable points as movement displacement of the face from the detection frame F_{km} to the tracking frame F_{km+1} , and thus obtaining estimated position D'_{km+1} of the face on the tracking frame F_{km+1} ;

2d) performing local face detection on the tracking frame F_{km+1} based on the estimated position D'_{km+1} to obtain accurate position D_{km+1} of the face;

2e) repeating the steps 2a) ~2d) until the face trajectories interrupt or all the detection frames have been detected, and thus obtaining all the possible face trajectories.

4. The method for face recognition-based video human image tracking under complex scenes of claim 3, wherein the conditions for the face trajectory interrupting comprise:

I) face position is located on edge of the video frame image;

II) the tracking stable points meeting the set conditions fail to be found;

III) the face fails to be detected at the estimated position of next frame.

5. The method for face recognition-based video human image tracking under complex scenes of claim 3 or 4, wherein the tracking stable points meeting the set conditions refer to that number of the tracking stable points is greater than 100.

6. The method for face recognition-based video human image tracking under complex scenes of claim 2, wherein the subsequent adjacent face trajectory in the step 3) refers to another face trajectory adjacent to current face trajectory in both time and space, thereinto, adjacence in time

refers to that start frame of the another face trajectory is within S frames behind end frame of the current face trajectory; adjacence in space refers to that distance between face on the start frame of the another face trajectory and face on the end frame of the current face trajectory is less than L face heights.

7. The method for face recognition-based video human image tracking under complex scenes of claim 6, wherein the Step 3) is specifically:

301) extracting recognition characteristics from all the face trajectories obtained in the step 2);

302) sequencing all the face trajectories according to sequence of their end frames and calculating the similarities between each of face trajectories and corresponding subsequent adjacent face trajectory in turn:

choosing n frames of faces respectively from the current face trajectory and the subsequent adjacent face trajectory, and cross-contrasting them to obtain n^2 groups of similarities, and choosing the highest group as similarities between the two trajectories, thereinto, the similarities obtained by cross-contrast is obtained by calculating cosine distance between recognition characteristics respectively corresponding to two frame images;

303) combining two groups of trajectories having higher similarities than the set threshold value.

8. The method for face recognition-based video human image tracking under complex scenes of claim 7, wherein in the step 301), extracting the recognition characteristics from the face trajectory is specifically:

3a) selecting N frames of faces from the face trajectory randomly;

3b) extracting LBP characteristics of each frame of face;

3c) utilizing a dimension reduction matrix trained beforehand to reduce the dimension of

the LBP characteristics to obtain the final recognition characteristics.