(12) STANDARD PATENT (11) Application No. AU 2016352215 B2
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Method and device for tracking location of human face, and electronic equipment

(51) International Patent Classification(s)
G06T 7/20 (2006.01) ***G06K 9/62*** (2006.01)

(21) Application No: **2016352215** (22) Date of Filing: **2016.11.04**

(87) WIPO No: **WO17/080399**

(30) Priority Data

(31) Number	(32) Date	(33) Country
201510772348.7	**2015.11.12**	**CN**

(43) Publication Date: **2017.05.18**
(44) Accepted Journal Date: **2020.10.01**

(71) Applicant(s)
Alibaba Group Holding Limited

(72) Inventor(s)
Wang, Nan;Du, Zhijun;Zhang, Yu

(74) Agent / Attorney
Spruson & Ferguson, Level 24 Tower 2 Darling Park 201 Sussex Street, Sydney, NSW, 2000, AU

(56) Related Art
AREE ALI MOHAMMED ET AL, "Efficient Face Tracking and Detection in Video: Based on Template Matching", PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, COMPUTER VISION, AND PATTERN RECOGNITION (IPCV), Athens, (2012-01-01)
CN 101527838 A

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2017 年 5 月 18 日 (18.05.2017)

WIPO | PCT

(10) 国际公布号
WO 2017/080399 A1

(51) 国际专利分类号:
G06T 7/20 (2017.01) *G06K 9/62* (2006.01)

(21) 国际申请号: PCT/CN2016/104491

(22) 国际申请日: 2016 年 11 月 4 日 (04.11.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510772348.7 2015 年 11 月 12 日 (12.11.2015) CN

(71) 申请人: **阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED)** [—/CN]; 英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: **王楠 (WANG, Nan)**; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 **杜志军 (DU, Zhijun)**; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 **张宇 (ZHANG, Yu)**; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: **北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.)**; 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (CN)。

(81) **指定国** (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) **指定国** (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

[见续页]

(54) Title: METHOD AND DEVICE FOR TRACKING LOCATION OF HUMAN FACE, AND ELECTRONIC EQUIPMENT

(54) 发明名称：一种人脸位置跟踪方法、装置和电子设备

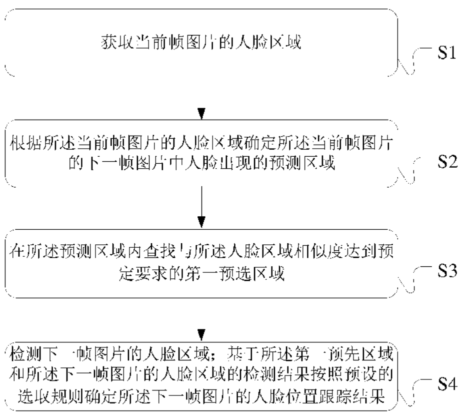

图 1

S1 Acquire a human face region of a current frame of image
S2 Determine, according to the human face region of the current frame of image, a prediction region, in which a human face appears, in a next frame of image of the current frame of image
S3 Find a first preselected region, which reaches a preset requirement in the aspect of similarity with the human face region, in the prediction region
S4 Detect a human face region of the next frame of image; and determine a human face location tracking result of the next frame of image based on the first preselected region and a detection result of the human face region of the next frame of image and according to a preset selection rule

(57) Abstract: A method and device for tracking the location of a human face, and electronic equipment. The method comprises: acquiring a human face region of a current frame of image (S1); determining, according to the human face region of the current frame of image, a prediction region, in which a human face appears, in a next frame of image of the current frame of image (S2); finding a first preselected region, which reaches a preset requirement in the aspect of similarity with the human face region, in the prediction region (S3); detecting a human face region of the next frame of image; and determining a human face location tracking result of the next frame of image based on the first preselected region and a detection result of the human face region of the next frame of image and according to a preset selection rule (S4). The method can accurately locate and track human face regions in frames of image in complex scenes such as strong light interference and fast face movement, thereby improving a human face tracking effect and improving user experience.

(57) 摘要：一种人脸位置跟踪方法、装置和电子设备。所述方法包括：获取当前帧图片的人脸区域（S1）；根据所述当前帧图片的人脸区域确定所述当前帧图片的下一帧图片中人脸出现的预测区域（S2）；在所述预测区域内查找与所述人脸区域相似度达到预定要求的第一预选区域（S3）；检测所述下一帧图片的人脸区域；基于所述第一预选区域和所述下一帧图片的人脸区域的检测结果，按照预设的选取规则确定所述下一帧图片的人脸位置跟踪结果（S4）。该方法可以在强光线干扰、人脸快速移动等复杂场景中准确定位跟踪出帧图片中的人脸区域，提高人脸跟踪效果，提高用户使用体验。

WO 2017/080399 A1

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

28 Aug 2020 2016352215

FACE LOCATION TRACKING METHOD, APPARATUS, AND ELECTRONIC DEVICE

[0001] The present application claims priority to Chinese Patent Application No. 201510772348.7, filed on November 12, 2015 and entitled "FACE LOCATION TRACKING METHOD, APPARATUS, AND ELECTRONIC DEVICE", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present application relates to the field of image information data processing, and in particular, to a face location tracking method, apparatus, and electronic device.

BACKGROUND

[0003] Face tracking is usually a process of determining the face's movement and change of size in a video or an image sequence. Face tracking always plays an important role and is widely used in image analysis and recognition, image monitoring and retrieval, instant video communication, etc.

[0004] Generally, the face tracking processing can mainly include identifying a face location in a video. During video recording, when a face moves, a specific algorithm such as a particle shift or a mean shift can be used to track the specific location of the face in the video. Currently, face tracking method processing used in the existing technology mainly include: performing face detection on each image frame, that is, each frame can be considered as a separate image, and then face detection is performed on each image frame, so as to obtain a face location in each image frame by calculation. However, in actual implementations, for example, in a process in which a user uses a front-facing camera of a mobile phone to take a selfie, a face tracking loss or a detection error is usually caused due to a sudden light or scene change, strong light or metering interference, rapid face movement, etc. Consequently, tracked face images are usually discontinuous in a user video monitoring process or a video call process, and a real-time smooth tracking effect cannot be achieved, thereby greatly compromising user experience, especially in a terminal device with relatively poor processing performance. Certainly, the face tracking method in the existing technology cannot meet a relatively high face-tracking requirement of a

2016352215 28 Aug 2020

user.

[0005] For the face tracking methods in the existing technology, a face tracking loss or an error occurs especially in a complex scenario such as a sudden light change, light interference, or rapid face movement, which can result in a blurred face image or discontinuous tracked face images in a video. The effectiveness of face detection and tracking, and user experience may be compromised.

SUMMARY

[0005a] It is an object of the present invention to substantially overcome or at least ameliorate one or more disadvantages of existing arrangements.

[0006] The present application provides a face location tracking method, apparatus, and electronic device, so that a face region in an image frame can be accurately located in a complex scenario such as a sudden light change, light interference, or rapid face movement, to improve effectiveness of face tracking. In addition, face tracking loss can be avoided, thereby improving face location tracking efficiency and user experience.

[0007] A face location tracking method, apparatus, and electronic device provided in the present application are implemented as follows:

[0008] A face location tracking method is provided, and the method includes:

obtaining a face region in a current image frame;

determining a predicted region that includes a face in a next image frame of the current image frame based on the face region in the current image frame;

searching the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement; and

detecting a face region in the next image frame, and determining a face location tracking result of the next image frame based on the first preselected region, a result of detecting the face region in the next image frame, and a predetermined selection rule.

[0009] A face location tracking apparatus is provided, and the apparatus includes:

a detection module, configured to detect a face region in a current image frame;

2016352215 28 Aug 2020

a predicted region calculation module, configured to calculate a predicted region that includes a face in a next image frame of the current image frame based on the face region that is in the current image frame and that is detected by the detection module;

a preselected region calculation module, configured to search the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement; and

a tracking result selection module, configured to determine a face location tracking result of the next image frame based on the first preselected region, a result of detecting a face region in the next image frame of the current image frame by the detection module, and a predetermined selection rule.

[0010] A face location tracking electronic device is provided, and the electronic device includes:

an information acquisition unit, configured to obtain a current image frame waiting to be processed;

a processing unit, configured to: detect a face region in the current image frame; calculate a predicted region that includes a face in a next image frame of the current image frame based on the detected face region in the current image frame, and search the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement; and determine a face location tracking result of the next image frame based on the first preselected region, a result of detecting a face region in the next image frame of the current image frame by the processing unit, and a predetermined selection rule; and

a display unit, configured to display the face location tracking result obtained by the processing unit.

[0011] In the face location tracking method, apparatus, and electronic device provided in the present application, a range of the predicted region that includes a face in the next image frame can be predicted based on the face region in the current image frame. Then, a preselected region of a face region with similarity to a face region in a previous image frame that meets a predetermined requirement (for example, the highest similarity) can be found within the range of the predicted region. As such, information about a preselected face can be obtained from the next image frame of the current frame based on the previous image frame. Further, in the solutions in the present application, the face region in the next image frame can be detected. If no face region is detected from the next image frame, a preselected region obtained by means of

2016352215 28 Aug 2020

calculation based on the previous image frame can be used as the face region in the next image frame. If the face region can be detected, the next image frame includes at least two face regions, and a face region that meets the requirement can be selected as the final face location tracking result of the next image frame based on the predetermined selection rule. Therefore, in the present application, even if no face region is detected from the current image frame due to a complex scenario such as a sudden light change, a face location can be located and tracked by using a preselected region obtained by prediction based on a face region in the previous image frame. As such, continuity of face detection and tracking can be ensured, face detection and tracking can be more effective, and user experience of face tracking can be improved.

[0011a] In one aspect there is provided a computer-implemented method for tracking locations of a face across a plurality of images comprising a first image and a second image, the method comprising: determining a first face region within the first image, the first face region including the location of the face within the first image; based on the determined first face region within the first image, determining a predicted face region within the second image; determining a first region of similarity within the predicted face region, the first region of similarity having at least a predetermined degree of similarity to the first face region within the first image; determining whether a second face region is present within the second image; and determining the location of the face within the second image based on the first region of similarity, the determination of whether the second face region is present within the second image, and a face region selection rule comprising: when the second face region is not present within the second image, the location of the face is the first region of similarity; when the second face region is present within the second image, and a spatial overlap coefficient between the second face region and the first region of similarity is 0, the location of the face is the first region of similarity; when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is less than a preset threshold, the location of the face is the first region of similarity; and when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is greater than or equal to the preset threshold, the location of the face is the second face region.

2016352215 28 Aug 2020

[0011b] In one aspect there is provided a non-transitory, computer-readable medium storing one or more instructions executable by a computer system to perform operations comprising: determining a first face region within a first image, the first face region including a location of a face within the first image; based on the determined first face region within the first image, determining a predicted face region within a second image; determining a first region of similarity within the predicted face region, the first region of similarity having at least a predetermined degree of similarity to the first face region within the first image; determining whether a second face region is present within the second image; and determining the location of the face within the second image based on the first region of similarity, the determination of whether the second face region is present within the second image, and a face region selection rule comprising: when the second face region is not present within the second image, the location of the face is the first region of similarity; when the second face region is present within the second image, and a spatial overlap coefficient between the second face region and the first region of similarity is 0, the location of the face is the first region of similarity; when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is less than a preset threshold, the location of the face is the first region of similarity; and when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is greater than or equal to the preset threshold, the location of the face is the second face region.

[0011c] In one aspect there is provided a computer-implemented system, comprising: one or more computers; and one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising: determining a first face region within a first image, the first face region including a location of a face within the first image; based on the determined first face region within the first image, determining a predicted face region within a second image; determining a first region of similarity within the predicted face region, the first region of similarity having at least a predetermined degree of similarity to the first face region within the first image; determining whether a second face region is present within the second image; and determining the location of the face within the second

2016352215 28 Aug 2020

image based on the first region of similarity, the determination of whether the second face region is present within the second image, and a face region selection rule comprising: when the second face region is not present within the second image, the location of the face is the first region of similarity; when the second face region is present within the second image, and a spatial overlap coefficient between the second face region and the first region of similarity is 0, the location of the face is the first region of similarity; when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is less than a preset threshold, the location of the face is the first region of similarity; and when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is greater than or equal to the preset threshold, the location of the face is the second face region.

BRIEF DESCRIPTION OF DRAWINGS

[0012] To describe the technical solutions in the implementations of the present application or in the prior art more clearly, the following briefly introduces the accompanying drawings used for describing the implementations or existing technology. Apparently, the accompanying drawings in the following description merely show some implementations of the present application, and a person of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

[0013] FIG. 1 is a method flowchart illustrating a face location tracking method, according to an implementation of the present application;

[0014] FIG. 2 is a schematic diagram illustrating determining a predicted region that includes a face in a next image frame based on a face region in a current image frame;

[0015] FIG. 3 is a schematic diagram illustrating searching a predicted region for a first preselected region, according to the present application;

[0016] FIG. 4 is a schematic diagram illustrating a selection scenario of determining a face location tracking result, according to the present application;

[0017] FIG. 5 is a schematic diagram illustrating further searching for a second preselected region, according to an implementation of the present application;

2016352215 28 Aug 2020

[0018] FIG. 6 is a schematic diagram illustrating a module structure of a face location tracking apparatus, according to an implementation of the present application;

[0019] FIG. 7 is a schematic diagram illustrating a module structure of a preselected region calculation module, according to an implementation of the present application;

[0020] FIG. 8 is a schematic diagram illustrating a module structure of a preselected region calculation module, according to another implementation of the present application;

[0021] FIG. 9 is a schematic diagram illustrating a module structure of a tracking result selection module, according to an implementation of the present application; and

[0022] FIG. 10 is a schematic structural diagram illustrating a face location tracking electronic apparatus, according to an implementation of the present application.

DESCRIPTION OF EMBODIMENTS

[0023] To make a person skilled in the art understand the technical solutions in the present application better, the following clearly and completely describes the technical solutions in the implementations of the present application with reference to the accompanying drawings in the implementations of the present application. Apparently, the described implementations are merely some but not all of the implementations of the present application. All other implementations obtained by a person of ordinary skill in the art based on the implementations of the present application without creative efforts shall fall within the protection scope of the present application.

[0024] The following describes in detail a basic data processing method in the present application with reference to the accompanying drawings. FIG. 1 is a method flowchart illustrating a face location tracking method, according to an implementation of the present application. Although the present application provides the operation steps of the method in the following implementations or accompanying drawings, the method can include more or fewer operation steps based on a conventional or non-creative effort. In a step in which there is no necessary causal relationship in logic, an execution sequence of these steps is not limited to an execution sequence provided in the implementations of the present application. When the method is executed by an apparatus or a terminal product in actual application, the method can be executed based on the sequence of the method in the implementations or accompanying drawings or can be executed in parallel (for example, a parallel processor or a multi-thread processing environment).

[0025] Specifically, as shown in FIG. 1, the present application provides an implementation of a face location tracking method, and the method can include the following steps.

[0026] S1: Obtain a face region in a current image frame.

[0027] Generally, face location tracking is mainly used in video stream information processing recorded by a camera apparatus, for example, a monitoring device, or a camera of a user's mobile phone. The face location tracking method provided in the present application can include but is not limited to video stream information processing, and the solution in the present application can still be used for face tracking of consecutive images or cine film digital information in other application scenarios. To clearly describe the solution in the present application, this implementation can be described by using an application scenario in which a user uses a front-facing camera of a mobile phone to take a selfie. In this implementation, a face region in a current image frame of a current video can be obtained first. Specifically, face detection can be performed on a video image frame at the beginning of a video stream. Generally, face detection needs to be performed on each image frame until a face is detected. Generally, when a face is detected from the current image frame, information related to the face image can be obtained, such as an image color parameter, an image size, and a frame distance. Face location can usually be represented by a specific region. For example, a rectangular box can usually be used to represent a region in which a face is detected.

[0028] In a terminal product application scenario where data processing capability of a mobile communications terminal or a vehicle dashboard video recorder is lower than a data processing capability of a personal computer (PC), the present application can provide an implementation of detecting a face region. In an implementation, the obtaining a face region in a current image frame can include:

detecting and obtaining the face region in the current image frame by using an Adaboost method for reducing a quantity of classification levels.

[0029] Adaboost is an iterative algorithm. The primary process of Adaboost includes training different classifiers (weak classifiers) for a same training set, and combining the weak classifiers into a stronger final classifier (strong classifier). The Adaboost algorithm can usually be used to determine a weighted value of each sample based on whether classification of each sample in each training set is correct and the accuracy of a previous overall classification. Then, a new data set whose weighted value has been modified can be sent to a next-layer classifier for training, and classifiers obtained during each training are combined into a final decision classifier. A classifier in an Adaboost cascade structure is usually composed of a series of serial classifiers. When determining samples to be identified, only a sample determined as positive by a classifier at a previous level is sent to a next classifier for further processing. Otherwise, a negative sample is directly rejected. In the cascade structure, previous classifiers are simpler in structure, use less feature data, but have a higher detection rate. Negative samples greatly different from a target object can be filtered out as much as possible. Subsequent classifiers use more feature data and have a more complex structure, so that negative samples similar to the target object can be distinguished from the target object.

[0030] It should be noted that in this implementation, in a process that the face region in the current image frame is detected by using the Adaboost method, a quantity of classification levels can be appropriately reduced as needed, so as to reduce calculation complexity in face detection and perform rapid face detection.

[0031] In an implementation of the face location tracking method in the present application, when a user performs active photographing by using a front-facing camera or a rear-facing camera of a mobile phone, if more than one person is photographed, a proportion of a main photographed face on a screen or a display is usually large. In this case, a face closest to the camera can be a main target. Therefore, to more accurately track a face location and meet a user's face location tracking expectation, in another implementation of the present application, only the face closest to the camera may be tracked during face location tracking. In an implementation process, the largest face region in the current image frame can be selected as a face tracking object. Therefore, in another implementation of the present application, the obtaining a face region in a current image frame can include:

when at least two faces are detected from the current image frame, selecting a region corresponding to a face with the largest area in the current image frame as the face region in the current image frame.

[0032] In this implementation, the face region in the current image frame of the video can be obtained.

[0033] S2: Determine a predicted region that includes a face in a next image frame of the current image frame based on the face region in the current image frame.

[0034] After the face region in the current image frame is obtained, the predicted region that includes a face in the next image frame can be determined based on the face region in the current image frame. The predicted region can include a range of a face region determined in the next image frame of the current image frame based on a specific algorithm or rule. FIG. 2 is a schematic diagram illustrating determining a predicted region that includes a face in a next image frame based on a face region in a current image frame. As shown in FIG. 2, in the current image frame *N*, a rectangular box *A* is a detected face region in the current image frame. In a next image frame *N*+1, a rectangular box *B* obtained after a length and a width of the rectangular box *A* of the face region in the previous image frame *N* (that is, the current image frame *N*) are separately expanded by *K* pixels is used as a predicted region that includes a face in the next image frame *N*+1. Certainly, in another implementation, the predicted region can be determined by using another method. For example, a rectangular box obtained after the length and the width of the rectangular box *A* of the face region in the current image frame are separately expanded by 1.5 times can be used as the predicted region.

[0035] The predicted region that includes a face in the next image frame of the current image frame is determined based on the face region in the current image frame and a specific rule.

[0036] S3: Search the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement.

[0037] After the predicted region that may include a face in the next image frame is determined, the first preselected region with high similarity to the face region can be searched and matched within the range of the predicted region. The face region can be obtained from the previous image frame (that is, the current image frame *N* in S2).

[0038] In this implementation, a template matching method can be provided to calculate similarity between the face region and the first preselected region in the next image frame. In an implementation process, the face region in the current image frame can be used as an original template, and a boundary of the face region can be set as a moving window. A new matching template can be obtained during each window movement. Within the range of the predicted region, each time the predicted region moves by one step, data of a new moving window region can be obtained, and similarity between the moving window region and the face region is calculated. In the present application, calculating or determining similarity between two regions is not limited to a specific method, and other methods that can implement the same or similar function can be applied to the present application. Specifically, the determining a predicted region that includes a face in a next image frame of the current image frame based on the face region in the current image frame can include:

[0039] S301: Traverse the predicted region based on a first step size to obtain a comparison region of the face region.

[0040] The first step size can be set based on a processing speed or a processing precision requirement in actual face location tracking. In an optional implementation, the data calculation amount can be reduced to improve the data processing speed, and a face location tracking accuracy requirement can be ensured. In this implementation, a value range of the first moving step can be greater than or equal to two pixels.

[0041] S302: Calculate similarity between the face region and the comparison region.

[0042] A corresponding calculation method and a corresponding calculation parameter can be selected to calculate similarity between different image regions based on different application scenarios or different data processing requirements. For example, the similarity between the face region and the comparison region can be calculated based on an image color, an image texture, or an image gradient of different regions. The present application provides an implementation of calculating the similarity. Specifically, in an implementation of the present application, the similarity, denoted as *dis*, between the face region and the comparison region can be calculated by using the following equations:

$$\min X = \max\left(-left_{ori}, -left_{des}\right)$$

$$\max X = \max\left(width - left_{ori}, width - left_{des}\right)$$

$$\min Y = \max\left(-top_{ori}, -top_{des}\right)$$

$$\max Y = \max\left(height - top_{ori}, height - top_{des}\right)$$

$$sumDis = \left\{ \sum_{i=\max(1,\min X)}^{\min(width,\max X)} \sum_{j=\max(1,\min Y)}^{\min(height,\max X)} \min\left\{\left|f(i,j) - g(i,j)\right|, x\right\} \right\}$$

$$effctiveNum = \left[\min\left(width, \max X\right) - \max\left(1, \min X\right)\right] * \left[\min\left(height, \max X\right) - \max\left(1, \min Y\right)\right]$$

$$dis = sumDis * \left(width * height\right) / effctiveNum$$

[0043] In the above equations, $left_{ori}$, $left_{des}$, top_{ori}, and top_{des} can respectively represent a left boundary location of the face region, a left boundary location of a current comparison region, an upper boundary location of the face region, and an upper boundary location of the current comparison region. The variable *width* can represent a width of the face region, *height* can represent a height of the face region, $f(i,j)$ can represent a grayscale value of a pixel whose coordinates are (i,j) in the face region in the current image frame (that is, a frame K preceding a next frame K+1), $g(i,j)$ can represent a grayscale value of a pixel whose coordinates are (i,j) in a comparison region in the next image frame, x can represent a specified empirical threshold, and *dis* is the similarity between the face region and the comparison region. In the above equations, max (a, b) can represent that a larger value between a and b is selected, and min (a, b) can represent that a smaller value between a and b is selected.

[0044] S303: Use a comparison region with similarity that meets the predetermined requirement in the next image frame as the first preselected region in the next image frame.

[0045] In this implementation, the predetermined requirement may be set and can be used to select a comparison region that meets a prediction requirement in the predicted region. For example, the predetermined requirement may be set as follows: the similarity between the comparison region and the face region is more than 90% or a comparison region within a percentage is specified after similarity sorting. For example, the first three highest-similarity comparison regions. In an optional implementation of the present application, the comparison region with similarity that meets the predetermined requirement can include:

a comparison region with the largest similarity among comparison regions in the next image frame.

[0046] FIG. 3 is a schematic diagram illustrating searching a predicted region for a first preselected region, according to the present application. As shown in FIG. 3, in a predicted region B of the next image frame N + 1, a first preselected region C with similarity to the face region A in the current image frame N that meets the requirement can be found. In this implementation, only a region that is in a next-frame comparison region and with the highest similarity to the face region in the current image frame can be selected as the first preselected region. In this case, in comparison with selecting multiple preselected regions, a data processing amount can be reduced, a preselected region can be selected faster, and a face location processing speed can be improved.

[0047] In this implementation, when the template matching method is used to calculate the first preselected region, addition and subtraction can be performed on a grayscale value of a pixel within a specific region. In comparison with another existing tracking algorithm, there is no need to perform massive processing and storage, and time complexity and space complexity are low. The application of this implementation is broader, especially for mid-low end mobile phone and monitoring device with weak information data processing capability. Therefore, the calculation amount can be effectively reduced and face tracking precision can be improved. In addition, in a face tracking environment of short-distance video recording, for example, in an application scenario of a front-facing camera of a mobile phone, a proportion of the face in a screen is usually large when a user takes a selfie. In the template matching method in this implementation, effective information of a face region in a larger proportion of the video screen can be obtained, and a tracking result can be more reliable in comparison with other tracking algorithms.

[0048] In this implementation, the first preselected region with similarity to the face region that meets the predetermined requirement can be searched for in the predicted region by using a specific calculation method.

[0049] S4: Detect a face region in the next image frame, and determine a face location tracking result of the next image frame based on the first preselected region, a result of detecting the face region in the next image frame, and a predetermined selection rule.

[0050] When the current image frame is switched to the next image frame, it can be detected whether a face region exists in the next image frame. If a face region is detected from the next image frame, at least two face regions are obtained from the next-image frame, that is, the detected face region and the first preselected region obtained based on face tracking prediction. In the present application, the final face tracking result of the next image frame can be obtained by particular collaborative calculation and analysis based on the at least two face regions.

[0051] In the present application, the predetermined selection rule can be used to determine which face region is used as the final face location tracking result. The selection rule in this implementation can include a selection rule for selecting the face region in the next image frame or the first preselected region according to a percentage of an overlap area between the face region in the next image frame and the first preselected region in either the face region in the next image frame or the first preselected region. In this implementation, the percentage of the overlap area between the face region and the first preselected region in the face region of the next image frame or the first preselected region can be defined as an overlap coefficient Q.

[0052] FIG. 4 is a schematic diagram illustrating a selection scenario of determining a face location tracking result, according to the present application. As shown in FIG. 4, a rectangular box D can represent the detected face region in the next image frame, and is referred to as a detection result here. A rectangular box C can represent the first preselected region in the next image frame and obtained by tracking calculation in steps S1 to S3 or another implementation, and is referred to as a tracking result here. The shaded area is the final determined face location tracking result of the next image frame. When both the detection result and the tracking result exist in the next image frame, if the detection result does not overlap with the tracking result, that is, the overlap coefficient Q is 0, the tracking result can be used as the face location tracking result, as shown in 4-1 in FIG. 4. If there is an overlap region between the detection result and the tracking result, but an overlap area is relatively small and does not meet a specified overlap requirement, the tracking result can be used as the face location tracking result, as shown in 4-2 in FIG. 4. In another case, if there is an overlap region between the detection result and the tracking result, and an overlap area is large and meets a specified overlap requirement, for example, 95% of the regions are overlapped, the detection result can be selected as the face location tracking result, as shown in 4-3 in FIG. 4. Certainly, if no face region is detected from the next image frame, the tracking result can be directly used as the face location tracking result, as shown in 4-4 in FIG. 4.

[0053] Therefore, in an optional implementation, determining a face location tracking result of the next image frame based on detecting the face region in the next image frame and a predetermined selection rule can include:

when the result of detecting the face region in the next image frame is that no face region is detected, using the first preselected region as the face location tracking result of the next image frame;

when a coefficient of an overlap between the detected face region in the next image frame and the first preselected region is 0, use the first preselected region as the face location tracking result of the next image frame;

when the coefficient of the overlap between the detected face region in the next image frame and the first preselected region is less than a predetermined threshold, use the first preselected region as the face location tracking result of the next image frame; or

when the coefficient of the overlap between the detected face region in the next image frame and the first preselected region is greater than or equal to the predetermined threshold, use the detected face region in the next image frame as the face location tracking result of the next image frame.

[0054] This implementation provides an implementation method for selecting the final face location tracking result from the detection result and the tracking result. In this implementation of the present application, a face location can be accurately and rapidly tracked in a complex environment such as rapid face movement, a sudden light change, or strong light interference. In the implementation, when a frame loss occurs in an image frame and no face is detected, a face location can still be tracked and determined, so that a continuous tracking effect of the face location can be implemented, and face tracking can be smooth. Even if a frame loss does not occur, a more suitable region can be selected between a detection result and a tracking result in an image frame based on a predetermined selection rule as a face location tracking result, so as to improve effectiveness of face tracking effect and user experience.

[0055] It should be noted that a current frame, a next frame, a previous frame, a frame preceding the previous frame in the present application can be considered as relative concepts for describing an image frame information processing object in actual applications. If an image frame at a moment in a video stream can be marked as a current image frame N, a corresponding next frame may be an $(N+1)$th image frame, and a previous frame may be an $(N-1)$th image frame. After tracking on the $(N+1)$th image frame is completed, a face location in an $(N+2)$th image frame can be further tracked and processed. In this case, the current image frame is the $(N+1)$th image frame, and correspondingly, a next image frame of the current image frame $N+1$ may be the $(N+2)$th image frame.

[0056] During continuous image frame processing, after tracking processing on a face location in a current image frame is completed, a face location in a next image frame can be further tracked by using a processing result of the current image frame as reference information or initialization information for tracking the face location in the next image frame. In some application scenarios, a quantity of image frames that need to be processed per second in the video stream may usually be more than a dozen frames or even dozens of frames. During face location tracking, if no face is detected because face tracking loss in frame *N* caused by a sudden light change or rapid face movement, a face region obtained from the result of detecting or processing a previous frame *N*–1 can be used as a face region tracked from frame *N* where the face tracking loss occurs. In a related implementation, if in addition, no face is detected or tracked from the previous frame *N*–1, a face region result obtained by detecting or processing a frame *N*–2 can still be used, and so on. Certainly, if no face is detected in multiple consecutive frames based on a predetermined determining rule, it can be determined that the face is not captured in the video recording range.

[0057] In the above implementation, when the predicted region that includes a face is searched for in the next image frame, a value of the first step size can be determined as needed. For example, the predicted region can be moved by two pixels or five pixels each time. Generally, a larger step size can indicate a faster speed of searching for a region that is similar to a previous image frame face region, and less data needs to be processed. A smaller step size can indicate higher search accuracy. In an implementation in which the value range of the first moving step is greater than or equal to two pixels, to further improve accuracy of searching the first preselected region, in another preferred implementation provided in the present application, the method can further include:

[0058] S304: Search for a second preselected region with the highest similarity to the face region within a range of a second step size surrounding the first preselected region, where the second step size is less than the first step size. In this case, the second preselected region obtained by means of precise searching can be used as the face location region tracking result of the next image frame. Either of the detection result and the second preselected region may be subsequently determined as the final face location tracking result.

[0059] Correspondingly, the determining a face location tracking result of the next image frame based on the first preselected region, a result of detecting the face region in the next image frame, and a predetermined selection rule includes: determining the face location tracking result of the next image frame based on the second preselected region, the result of detecting the face region in the next image frame, and the predetermined selection rule.

[0060] In specific application, for example, the first step is two pixels, similarity between a face region within a range of a pixel surrounding the first preselected region and the face region in the previous image frame may be calculated in this implementation, so as to obtain a region with the highest similarity. Specifically, in this implementation, the second preselected region can be calculated by using the similarity calculation method in step S302. Certainly, other calculation methods for determining similarity between two regions are not excluded. Details for those calculation methods are not described here. FIG. 5 is a schematic diagram illustrating further searching for a second preselected region, according to an implementation of the present application. As shown in FIG. 5, a rectangular box *C* is a first preselected region of a face region that is determined by using two pixels as the step size of a predicted region, and a rectangular box *D* is a comparison region *C_ru* of an upper right pixel of the first preselected region *C*. Certainly, a range of a pixel surrounding the first preselected region can include a comparison region *C_d* formed by moving the first preselected region downward by one pixel, a comparison region *C_u* formed by moving the first preselected region upward by one pixel, a comparison region *C_ld* formed by moving the first preselected region towards the lower left corner by one pixel, and so on. Then, the similarity between the comparison region of a pixel surrounding the first preselected region and the face region can be calculated, and a comparison region with the highest similarity can be selected as the second preselected region.

[0061] In this implementation, the first preselected region is calculated by setting a relatively large first step size, so as to effectively reduce the calculation amount in image comparison and searching to improve data processing speed of face location tracking. In this implementation, based on the result of the first preselected region, more accurate search can be performed nearby using the second step size smaller than the first step size, to obtain the second preselected region that has more accurate tracking result. As such, rapid searching can be implemented, and accuracy of face tracking can be improved, thereby improving effectiveness of face tracking.

[0062] Based on the face location tracking method in the present application, the present application provides a face location tracking apparatus. FIG. 6 is a schematic diagram illustrating a module structure of a face location tracking apparatus, according to an implementation of the present application. As shown in FIG. 6, the apparatus can include:

a detection module 101, configured to detect a face region in a current image frame;

a predicted region calculation module 102, configured to calculate a predicted region that includes a face in a next image frame of the current image frame based on the face region that is in the current image frame and that is detected by the detection module 101;

a preselected region calculation module 103, configured to search the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement; and

a tracking result selection module 104, configured to determine a face location tracking result of the next image frame based on the first preselected region, a result of detecting a face region in the next image frame of the current image frame by the detection module 101, and a predetermined selection rule.

[0063] In the implementation of the face location tracking apparatus in the present application, the detection module 101 can continuously detect, over time, a face region in an image frame obtained by a camera apparatus. For example, 15 frames of video images are shot per second in a video stream, and a face region in a current frame (an *N*th frame) of image can be detected during face location tracking. After detection and tracking processing on information data of the current frame (the *N*th frame) of image is completed, a face region in a next frame (an (*N*+*1*)th frame) of image can be further detected.

[0064] In an implementation of the detection module 101 of the apparatus in the present application, the face region in the current image frame can be detected and obtained by using an Adaboost method for reducing a quantity of classification levels. As such, a data calculation amount during face detection can be reduced, and a positioning and processing speed of face location tracking can be improved.

[0065] In another implementation of the apparatus in the present application, only a face closest to a camera can be tracked during face location tracking processing. In an implementation process, the largest face region in the current image frame can be selected as a face tracking object. Therefore, in another implementation of the apparatus in the present application, that the detection module 101 detects a face region in a current image frame includes:

when at least two faces are detected from the current image frame, selecting a region corresponding to a face with the largest area in the current image frame as the face region in the current image frame.

[0066] The present application is not limited to a method of searching for the first preselected region by the preselected region calculation module 103 based on calculation. FIG. 7 is a schematic diagram illustrating a module structure of a preselected region calculation module 103 of the apparatus, according to an implementation of the present application. As shown in FIG. 7, the preselected region calculation module 103 can include a comparison region module 1031, a similarity calculation module 1032, and a first preselection module 1033.

[0067] The comparison region module 1031 is configured to traverse the predicted region based on a specified first step to obtain a comparison region of the face region.

[0068] The similarity calculation module 1032 is configured to calculate similarity between the face region and the comparison region.

[0069] In an implementation of the apparatus in the present application, the similarity calculation module 1032 can calculate the similarity *dis* between the face region and the comparison region by using the following equations:

$$\min X = \max\left(-left_{ori}, -left_{des}\right)$$

$$\max X = \max\left(width - left_{ori}, width - left_{des}\right)$$

$$\min Y = \max\left(-top_{ori}, -top_{des}\right)$$

$$\max Y = \max\left(height - top_{ori}, height - top_{des}\right)$$

$$sumDis = \left\{ \sum_{i=\max(1,\min X)}^{\min(width,\max X)} \sum_{j=\max(1,\min Y)}^{\min(height,\max X)} \min\left\{\left|f(i,j) - g(i,j)\right|, x\right\} \right\}$$

$$effctiveNum = \left[\min\left(width, \max X\right) - \max\left(1, \min X\right)\right] * \left[\min\left(height, \max X\right) - \max\left(1, \min Y\right)\right]$$

$$dis = sumDis * \left(width * height\right) / effctiveNum$$

[0070] In the above equations, $left_{ori}$, $left_{des}$, top_{ori}, and top_{des} respectively represent a left boundary location of the face region, a left boundary location of a current comparison region, an upper boundary location of the face region, and an upper boundary location of the current comparison region. The variable *width* represents a width of the face region, *height* represents a height of the face region, $f(i,j)$ represents a grayscale value of a pixel whose coordinates are (i,j) in the face region in the current image frame, and $g(i,j)$ represents a grayscale value of a pixel whose coordinates are (i,j) in a comparison region in the next image frame; and x represents a specified empirical threshold, and *dis* is the similarity between the face region and the comparison region.

[0071] The first preselection module 1033 is configured to use a comparison region with similarity that meets the predetermined requirement in the next image frame as the first preselected region in the next image frame.

[0072] The equations used by the similarity calculation module 1032 included in the apparatus can be implemented in a specific implementation process on the apparatus/module by using a computer readable program language, for example, a C language, or can be implemented in a form of hardware and software using certain hardware structure as needed.

[0073] In an optional implementation, the predetermined requirement specified by the first preselection module 1033 may be set with the highest similarity to the face region. Therefore, in another implementation, the comparison region with similarity that meets the predetermined requirement in the first preselection module 1033 can include:

a comparison region with the largest similarity among comparison regions in the next image frame.

[0074] In the above implementation, the first step specified by the comparison region module 1031 can be set according to a requirement for a processing speed or processing accuracy of the face location tracking apparatus in the present application. In an implementation of the apparatus in the present application, a value range of the first step size can be set to be greater than or equal to two pixels.

[0075] The present application can further provide a preferred implementation of the face location tracking apparatus. FIG. 8 is a schematic diagram illustrating a module structure of a preselected region calculation module, according to another implementation of the present application. As shown in FIG. 8, the preselected region calculation module 103 can further include:

a second preselection module 1034, configured to search for a second preselected region with the highest similarity to the face region within a range of a second step size surrounding the first preselected region, where the second step size is less than the first step size.

[0076] Correspondingly, that the tracking result selection module 104 determines a face location tracking result of the next image frame based on the first preselected region, a result of detecting a face region in the next image frame of the current image frame by the detection module 101, and a predetermined selection rule includes: the tracking result selection module 104 determines the face location tracking result of the next image frame based on the second preselected region, the result of detecting the face region in the next image frame of the current image frame by the detection module 101, and the predetermined selection rule.

[0077] In the face location tracking apparatus in this implementation, based on the result of the first preselected region, more precise searching can be performed within the range of the second step size and smaller than the first step size, and the second preselected region with a more accurate tracking result is obtained. As such, rapid searching processing can be implemented, and accuracy of face location tracking can be improved, thereby improving the effectiveness of face tracking.

[0078] FIG. 9 is a schematic diagram illustrating a module structure of a tracking result selection module 104, according to an implementation of the present application. As shown in FIG. 9, the tracking result selection module 104 can include a detection and calculation module 1041 and a selection module 1042.

[0079] The detection and calculation module 1041 is configured to: detect the face region in the next image frame, and calculate a coefficient Q of an overlap between the face region of the next image frame and the first preselected region when the face region in the next image frame is detected. In this implementation, the overlap coefficient Q can be represented as a percentage of an overlap area between the face region in the next image frame and the first preselected region in either the face region or the first preselected region.

[0080] The selection module 1042 is configured to use the first preselected region as the face location tracking result of the next image frame, when at least one of the following conditions is met: no face region is detected by the detection and calculation module 1041 from the next image frame, the overlap coefficient calculated by the detection and calculation module 1041 is 0, or the overlap coefficient calculated by the detection and calculation module 1041 is less than a predetermined threshold. Or the selection module 1042 is configured to use the face region that is in the next image frame and that is detected by the detection module 101 as the face location tracking result of the next image frame, when the overlap coefficient calculated by the detection and calculation module 1041 is greater than or equal to the predetermined threshold.

[0081] This implementation provides a solution for selecting a final face location tracking result from a detection result and a tracking result. In this implementation of the present application, a face location can be accurately and rapidly tracked in a complex environment such as rapid face movement, a sudden light change, or strong light interference. In this implementation, when a frame loss occurs in an image frame and no face is detected, a face location can still be tracked and determined, so that a continuous tracking effect of the face location can be implemented to ensure smooth face tracking. Even if a frame loss does not occur, a more suitable region can be selected between a detection result and a tracking result in an image frame as a face location tracking result based on a predetermined selection rule, so as to improve effectiveness of face tracking and user experience.

[0082] The face location tracking method or apparatus in the present application can be applied to multiple terminal devices to perform more rapid, accurate, and fluent face location tracking. For example, such devices can include a video camera device, a monitoring device, and a face location tracking device for consecutive image frames of a mobile communications terminal based on an Android system or an iOS system. Therefore, the present application further provides a face location tracking electronic device, and the electronic device can include a camera apparatus for detecting and obtaining a video image frame, a display for video playback, a processing unit for information data processing, etc. Specifically, FIG. 10 is a schematic structural diagram illustrating a face location tracking electronic apparatus, according to an implementation of the present application. As shown in FIG. 10, the electronic device can include:

an information acquisition unit 1, configured to obtain a current image frame waiting to be processed;

a processing unit 2, configured to: detect a face region in the current image frame; calculate a predicted region that includes a face in a next image frame of the current image frame based on the detected face region in the current image frame, and search the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement; and determine a face location tracking result of the next image frame based on the first preselected region, a result of detecting a face region in the next image frame of the current image frame by the processing unit, and a predetermined selection rule; and

a display unit 3, configured to display the face location tracking result obtained by the processing unit 2.

[0083] The information acquisition unit 1 in the electronic device in the present application can include a front-facing camera, a rear-facing camera, or a monitoring camera apparatus of a mobile terminal. In another application scenario, an implementation in which a computer processes image information data that is obtained in real time or has been previously obtained is also included, for example, the computer performs face location tracking processing on video information. The processing unit 2 can include a central processing unit (CPU), and certainly, can further include a single-chip microcomputer with a logic processing capability, a logic gate circuit, an integrated circuit, etc. The display unit 3 can generally include a display, a mobile terminal display screen, a projection device, etc.

[0084] Although the face detection method, data processing such as data exchange between module units, and information display are described in the content of the present application, the present application is not limited to the data processing and the information display mentioned in the standard or the implementations. The above description in the implementations of the present application is merely application of some implementations of the present application, and a processing method slightly modified based on some standards and methods can also be used to implement the solutions in the implementations of the present application. Certainly, other non-creative changes in accordance with steps of the processing method in the implementations of the present application can still be used to implement the same application. Details are not described here.

[0085] Although the present application provides the operation steps of the method in the implementations or flowcharts, the method can include more or fewer operation steps based on a conventional or non-creative effort. A sequence of the steps enumerated in the implementations is merely one of execution sequences of the steps, and does not represent a unique execution sequence. When the method is executed by an apparatus or a client product in actual application, the method can be executed based on the sequence of the method in the implementations or accompanying drawings or can be executed in parallel (for example, a parallel processor or a multi-thread processing environment).

[0086] The units or modules described in the above implementations can be specifically implemented by a computer chip or an entity, or implemented by a product with a certain function. For ease of description, the above apparatus and device are described by using various modules and various units. Certainly, during implementation of the present application, functions of multiple modules can be implemented in one or more pieces of software and/or hardware, for example, a first preselection module and a second preselection module, or modules that implement a same function can be implemented by using a combination of multiple submodules or subunits.

[0087] A person skilled in the art also knows that, in addition to implementing a controller by using a computer readable program code, logic programming can be performed on the method steps to enable the controller to implement a same function in forms of a logic gate, a switch, a dedicated integrated circuit, a programmable logic controller, and an embedded microcontroller. Therefore, the controller can be considered as a hardware component, and an apparatus that is included in the controller and that is used to implement various functions can also be considered as a structure in the hardware component. Or, an apparatus for implementing various functions can be even considered as both a software module for implementing the method and the structure in the hardware component.

[0088] The present application can be described in a general context of a computer executable instruction executed by a computer, such as a C language, or a program module based on an Android design platform or an iOS design platform. Generally, the program module includes a routine, a program, an object, a component, a data structure, a type, or the like that executes a specific task or implements a specific abstract data type. The present application can also be applied to a distributed computing environment in which a task is executed by a remote processing device that is connected by using a communications network. In the distributed computing environment, the program module can be located in local and remote computer storage media including a storage device.

[0089] It can be learned from description of the above implementations that, a person skilled in the art can clearly understand that the present application can be implemented by using software in addition to a necessary universal hardware platform. Based on such an understanding, the technical solutions in the present application essentially or the part contributing to the prior art can be implemented in a form of a software product. The software product can be stored in a storage medium, such as a ROM/RAM, a magnetic disk, or an optical disc, and includes several instructions for instructing a computer device (which can be a personal computer, a mobile terminal, a server, or a network device) to perform the methods described in the implementations or in some parts of the implementations of the present application.

[0090] The implementations in this specification are all described in a progressive manner. For same or similar parts in the implementations, reference can be made to these implementations, and each implementation focuses on a difference from other implementations. The present application can be applied to many general-purpose or dedicated computer system environments or configurations, for example, a personal computer, a server computer, a handheld device or a portable device, a tablet device, a mobile communications terminal, a multiprocessor system, a microprocessor system, a programmable electronic device, a network PC, a small computer, a mainframe computer, and a distributed computing environment including any of the above systems or devices.

[0091] Although the present application is depicted by using the implementations, a person of ordinary skill in the art knows that the present application has many variations and changes without departing from the spirit of the present application, and the appended claims include these variations and changes without departing from the spirit of the present application.

2016352215 28 Aug 2020

CLAIMS:

1. A computer-implemented method for tracking locations of a face across a plurality of images comprising a first image and a second image, the method comprising:

determining a first face region within the first image, the first face region including the location of the face within the first image;

based on the determined first face region within the first image, determining a predicted face region within the second image;

determining a first region of similarity within the predicted face region, the first region of similarity having at least a predetermined degree of similarity to the first face region within the first image;

determining whether a second face region is present within the second image; and

determining the location of the face within the second image based on the first region of similarity, the determination of whether the second face region is present within the second image, and a face region selection rule comprising:

when the second face region is not present within the second image, the location of the face is the first region of similarity;

when the second face region is present within the second image, and a spatial overlap coefficient between the second face region and the first region of similarity is 0, the location of the face is the first region of similarity;

when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is less than a preset threshold, the location of the face is the first region of similarity; and

when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is greater than or equal to the preset threshold, the location of the face is the second face region.

2. The computer-implemented method of claim 1, wherein determining the first region of similarity within the predicted face region comprises:

setting a first comparison region within the predicted face region;

determining a first degree of similarity between the first comparison region and the first face region;

determining that the first degree of similarity satisfies the predetermined degree of

2016352215 28 Aug 2020

similarity; and

based on the determination that the first degree of similarity satisfies the predetermined degree of similarity, determining the first comparison region to be the first region of similarity.

3. The computer-implemented method of claim 2, wherein determining the first degree of similarity between the first comparison region and the first face region comprises performing calculations according to the following equations:

$$\min X = \max\left(-left_{ori}, -left_{des}\right)$$

$$\max X = \max\left(width - left_{ori}, width - left_{des}\right)$$

$$\min Y = \max\left(-top_{ori}, -top_{des}\right)$$

$$\max Y = \max\left(height - top_{ori}, height - top_{des}\right)$$

$$sumDis = \left\{ \sum_{i=\max(1,\min X)}^{\min(width,\max X)} \sum_{j=\max(1,\min Y)}^{\min(height,\max X)} \min\left\{\left|f(i,j) - g(i,j)\right|, x\right\} \right\}$$

$$effctiveNum = \left[\min\left(width, \max X\right) - \max\left(1, \min X\right)\right] * \left[\min\left(height, \max X\right) - \max\left(1, \min Y\right)\right]$$

$$dis = sumDis * \left(width * height\right) / effctiveNum$$

wherein $left_{ori}$, $left_{des}$, top_{ori}, and top_{des} respectively represent a left boundary location of the first face region, a left boundary location of the first comparison region, an upper boundary location of the first face region, and an upper boundary location of the first comparison region,

wherein *width* represents a width of the first face region and *height* represents a height of the first face region,

wherein *f(i,j)* represents a grayscale value of a pixel of the first image having coordinates *(i,j)* in the first face region of the first image and *g(i,j)* represents a grayscale value of a pixel of the second image having coordinates *(i,j)* in the first comparison region of the second image, and

wherein *x* represents a preset threshold and *dis* is the first degree of similarity between the first face region and the first comparison region.

4. The computer-implemented method of claim 2, wherein determining the first region of similarity within the predicted face region further comprises:

2016352215 28 Aug 2020

setting a second comparison region within the predicted face region, the second comparison region being different from the first comparison region;

determining a second degree of similarity between the second comparison region and the first face region; and

determining that the second degree of similarity satisfies the predetermined degree of similarity, and

wherein determining the first comparison region to be the first region of similarity comprises:

determining that the first degree of similarity is greater than the second degree of similarity; and

based on the determination that the first degree of similarity is greater than the second degree of similarity, determining the first comparison region to be the first region of similarity.

5. The computer-implemented method of claim 2, wherein determining the first region of similarity within the predicted face region further comprises:

setting a second comparison region within the predicted face region by translating the first comparison region by a first step, the first step being two or more pixels of the second image.

6. The computer-implemented method of claim 5, further comprising:

based on the determination of the first region of similarity within the predicted face region, determining a second region of similarity within the predicted face region, comprising:

setting a third comparison region within the predicted face region by translating the first region of similarity by a second step smaller than the first step;

determining a third degree of similarity between the third comparison region and the first face region;

determining that the third degree of similarity is greater than the first degree of similarity; and

based on the determination that the third degree of similarity is greater than the first degree of similarity, determining the third comparison region to be the second region of similarity,

wherein the determination of the location of the face within the second image is further based on the second region of similarity.

2016352215 28 Aug 2020

7. The computer-implemented method of claim 1, wherein determining a first face region within the first image comprises:

determining that the first image comprises a plurality of faces;

determining a plurality of face region areas; and

determining a region comprising a face of the plurality of faces having the largest face region area to be the first face region.

8. A non-transitory, computer-readable medium storing one or more instructions executable by a computer system to perform operations comprising:

determining a first face region within a first image, the first face region including a location of a face within the first image;

based on the determined first face region within the first image, determining a predicted face region within a second image;

determining a first region of similarity within the predicted face region, the first region of similarity having at least a predetermined degree of similarity to the first face region within the first image;

determining whether a second face region is present within the second image; and

determining the location of the face within the second image based on the first region of similarity, the determination of whether the second face region is present within the second image, and a face region selection rule comprising:

when the second face region is not present within the second image, the location of the face is the first region of similarity;

when the second face region is present within the second image, and a spatial overlap coefficient between the second face region and the first region of similarity is 0, the location of the face is the first region of similarity;

when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is less than a preset threshold, the location of the face is the first region of similarity; and

when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is greater than or equal to the preset threshold, the location of the face is the second face region.

9. The computer-readable medium of claim 8, wherein determining the first region of similarity within the predicted face region comprises:

setting a first comparison region within the predicted face region;

determining a first degree of similarity between the first comparison region and the first face region;

determining that the first degree of similarity satisfies the predetermined degree of similarity; and

based on the determination that the first degree of similarity satisfies the predetermined degree of similarity, determining the first comparison region to be the first region of similarity.

10. The computer-readable medium of claim 9, wherein determining the first degree of similarity between the first comparison region and the first face region comprises performing calculations according to the following equations:

$$\min X = \max\left(-left_{ori}, -left_{des}\right)$$

$$\max X = \max\left(width - left_{ori}, width - left_{des}\right)$$

$$\min Y = \max\left(-top_{ori}, -top_{des}\right)$$

$$\max Y = \max\left(height - top_{ori}, height - top_{des}\right)$$

$$sumDis = \left\{ \sum_{i=\max(1,\min X)}^{\min(width,\max X)} \sum_{j=\max(1,\min Y)}^{\min(height,\max X)} \min\left\{ \left| f(i,j) - g(i,j) \right|, x \right\} \right\}$$

$$effctiveNum = \left[\min(width, \max X) - \max(1, \min X)\right] * \left[\min(height, \max X) - \max(1, \min Y)\right]$$

$$dis = sumDis * \left(width * height\right) / effctiveNum$$

wherein $left_{ori}$, $left_{des}$, top_{ori}, and top_{des} respectively represent a left boundary location of the first face region, a left boundary location of the first comparison region, an upper boundary location of the first face region, and an upper boundary location of the first comparison region,

wherein *width* represents a width of the first face region and *height* represents a height of the first face region,

wherein *f(i,j)* represents a grayscale value of a pixel of the first image having coordinates *(i,j)* in the first face region of the first image and *g(i,j)* represents a grayscale value of a pixel of the second image having coordinates *(i,j)* in the first comparison region of the second image, and

wherein x represents a preset threshold and *dis* is the first degree of similarity between the first face region and the first comparison region.

11. The computer-readable medium of claim 9, wherein determining the first region of similarity within the predicted face region further comprises:

setting a second comparison region within the predicted face region, the second comparison region being different from the first comparison region;

determining a second degree of similarity between the second comparison region and the first face region; and

determining that the second degree of similarity satisfies the predetermined degree of similarity, and

wherein determining the first comparison region to be the first region of similarity comprises:

determining that the first degree of similarity is greater than the second degree of similarity; and

based on the determination that the first degree of similarity is greater than the second degree of similarity, determining the first comparison region to be the first region of similarity.

12. The computer-readable medium of claim 9, wherein determining the first region of similarity within the predicted face region further comprises:

setting a second comparison region within the predicted face region by translating the first comparison region by a first step, the first step being two or more pixels of the second image.

13. The computer-readable medium of claim 12, wherein the operations further comprise:

based on the determination of the first region of similarity within the predicted face region, determining a second region of similarity within the predicted face region, comprising:

setting a third comparison region within the predicted face region by translating the first region of similarity by a second step smaller than the first step;

determining a third degree of similarity between the third comparison region and the first face region;

determining that the third degree of similarity is greater than the first degree of

2016352215 28 Aug 2020

similarity; and

based on the determination that the third degree of similarity is greater than the first degree of similarity, determining the third comparison region to be the second region of similarity,

wherein the determination of the location of the face within the second image is further based on the second region of similarity.

14. The computer-readable medium of claim 8, wherein determining a first face region within the first image comprises:

determining that the first image comprises a plurality of faces;

determining a plurality of face region areas; and

determining a region comprising a face of the plurality of faces having the largest face region area to be the first face region.

15. A computer-implemented system, comprising:

one or more computers; and

one or more computer memory devices interoperably coupled with the one or more computers and having tangible, non-transitory, machine-readable media storing one or more instructions that, when executed by the one or more computers, perform one or more operations comprising:

determining a first face region within a first image, the first face region including a location of a face within the first image;

based on the determined first face region within the first image, determining a predicted face region within a second image;

determining a first region of similarity within the predicted face region, the first region of similarity having at least a predetermined degree of similarity to the first face region within the first image;

determining whether a second face region is present within the second image; and

determining the location of the face within the second image based on the first region of similarity, the determination of whether the second face region is present within the second image, and a face region selection rule comprising:

when the second face region is not present within the second image, the location of the face is the first region of similarity;

2016352215 28 Aug 2020

when the second face region is present within the second image, and a spatial overlap coefficient between the second face region and the first region of similarity is 0, the location of the face is the first region of similarity;

when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is less than a preset threshold, the location of the face is the first region of similarity; and

when the second face region is present within the second image, and the spatial overlap coefficient between the second face region and the first region of similarity is greater than or equal to the preset threshold, the location of the face is the second face region.

16. The computer-implemented system of claim 15, wherein determining the first region of similarity within the predicted face region comprises:

setting a first comparison region within the predicted face region;

determining a first degree of similarity between the first comparison region and the first face region;

determining that the first degree of similarity satisfies the predetermined degree of similarity; and

based on the determination that the first degree of similarity satisfies the predetermined degree of similarity, determining the first comparison region to be the first region of similarity.

17. The computer-implemented system of claim 16, wherein determining the first degree of similarity between the first comparison region and the first face region comprises performing calculations according to the following equations:

$$\min X = \max\left(-left_{ori}, -left_{des}\right)$$

$$\max X = \max\left(width - left_{ori}, width - left_{des}\right)$$

$$\min Y = \max\left(-top_{ori}, -top_{des}\right)$$

$$\max Y = \max\left(height - top_{ori}, height - top_{des}\right)$$

$$sumDis = \left\{ \sum_{i=\max(1,\min X)}^{\min(width,\max X)} \sum_{j=\max(1,\min Y)}^{\min(height,\max X)} \min\left\{ \left| f(i,j) - g(i,j) \right|, x \right\} \right\}$$

$$effctiveNum = \left[\min\left(width, \max X\right) - \max\left(1, \min X\right)\right] * \left[\min\left(height, \max X\right) - \max\left(1, \min Y\right)\right]$$

$$dis = sumDis * \left(width * height\right) / effctiveNum$$

wherein $left_{ori}$, $left_{des}$, top_{ori}, and top_{des} respectively represent a left boundary location of the first face region, a left boundary location of the first comparison region, an upper boundary location of the first face region, and an upper boundary location of the first comparison region,

wherein *width* represents a width of the first face region and *height* represents a height of the first face region,

wherein $f(i,j)$ represents a grayscale value of a pixel of the first image having coordinates (i,j) in the first face region of the first image and $g(i,j)$ represents a grayscale value of a pixel of the second image having coordinates (i,j) in the first comparison region of the second image, and

wherein x represents a preset threshold and *dis* is the first degree of similarity between the first face region and the first comparison region.

18. The computer-implemented system of claim 16, wherein determining the first region of similarity within the predicted face region further comprises:

setting a second comparison region within the predicted face region, the second comparison region being different from the first comparison region;

determining a second degree of similarity between the second comparison region and the first face region; and

determining that the second degree of similarity satisfies the predetermined degree of similarity, and

wherein determining the first comparison region to be the first region of similarity comprises:

determining that the first degree of similarity is greater than the second degree of similarity; and

based on the determination that the first degree of similarity is greater than the second degree of similarity, determining the first comparison region to be the first region of similarity.

19. The computer-implemented system of claim 16, wherein determining the first region of

2016352215 28 Aug 2020

similarity within the predicted face region further comprises:

setting a second comparison region within the predicted face region by translating the first comparison region by a first step, the first step being two or more pixels of the second image.

20. The computer-implemented system of claim 19, further comprising:

based on the determination of the first region of similarity within the predicted face region, determining a second region of similarity within the predicted face region, comprising:

setting a third comparison region within the predicted face region by translating the first region of similarity by a second step smaller than the first step;

determining a third degree of similarity between the third comparison region and the first face region;

determining that the third degree of similarity is greater than the first degree of similarity; and

based on the determination that the third degree of similarity is greater than the first degree of similarity, determining the third comparison region to be the second region of similarity,

wherein the determination of the location of the face within the second image is further based on the second region of similarity.

Alibaba Group Holding Limited

Patent Attorneys for the Applicant/Nominated Person

SPRUSON & FERGUSON

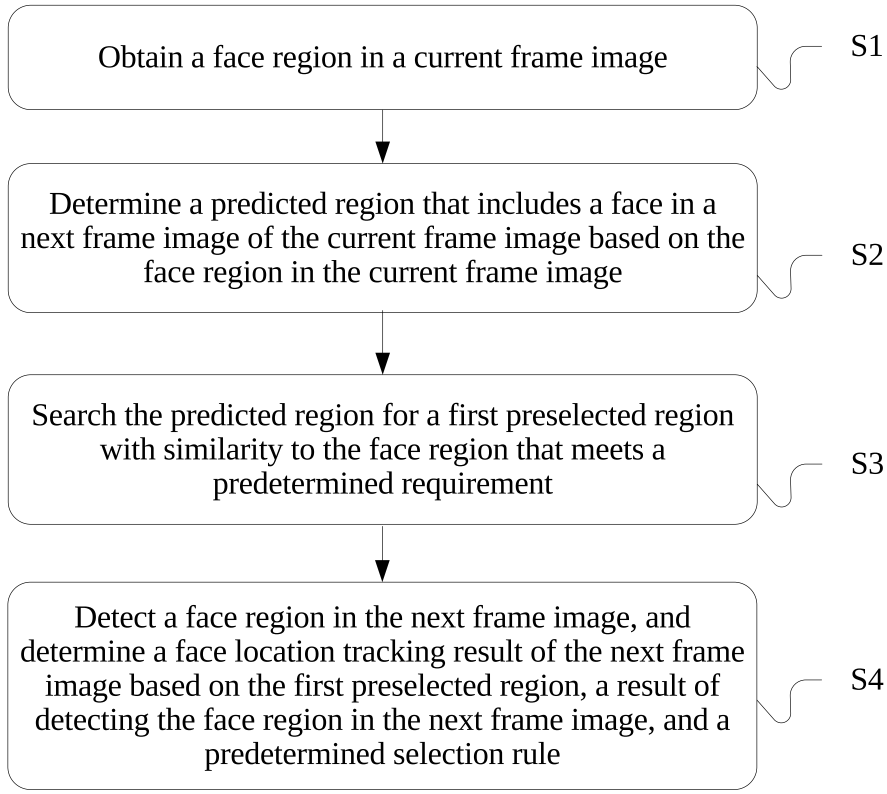
Obtain a face region in a current frame image
S1
Determine a predicted region that includes a face in a next frame image of the current frame image based on the face region in the current frame image
S2
Search the predicted region for a first preselected region with similarity to the face region that meets a predetermined requirement
S3
Detect a face region in the next frame image, and determine a face location tracking result of the next frame image based on the first preselected region, a result of detecting the face region in the next frame image, and a predetermined selection rule
S4

FIG. 1

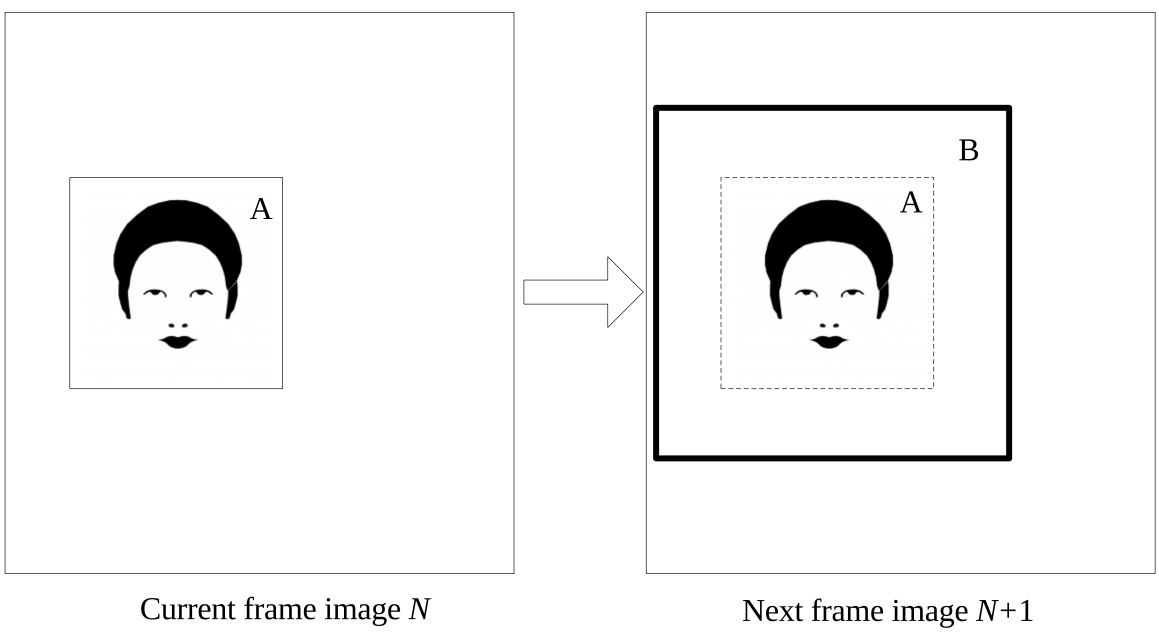
A
B
A
Current frame image N
Next frame image N+1

FIG. 2

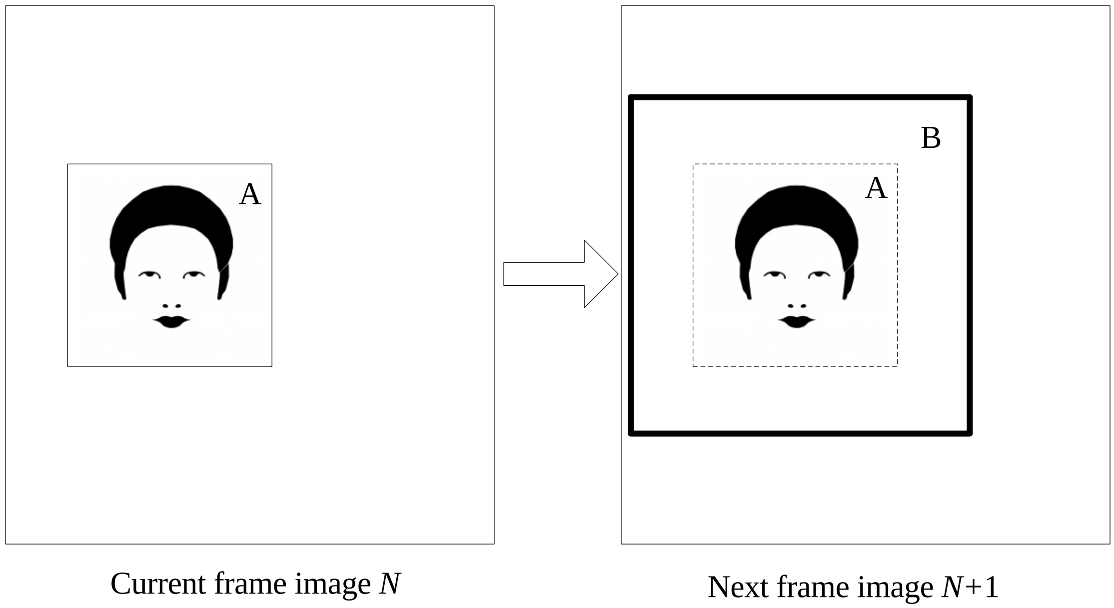

FIG. 3

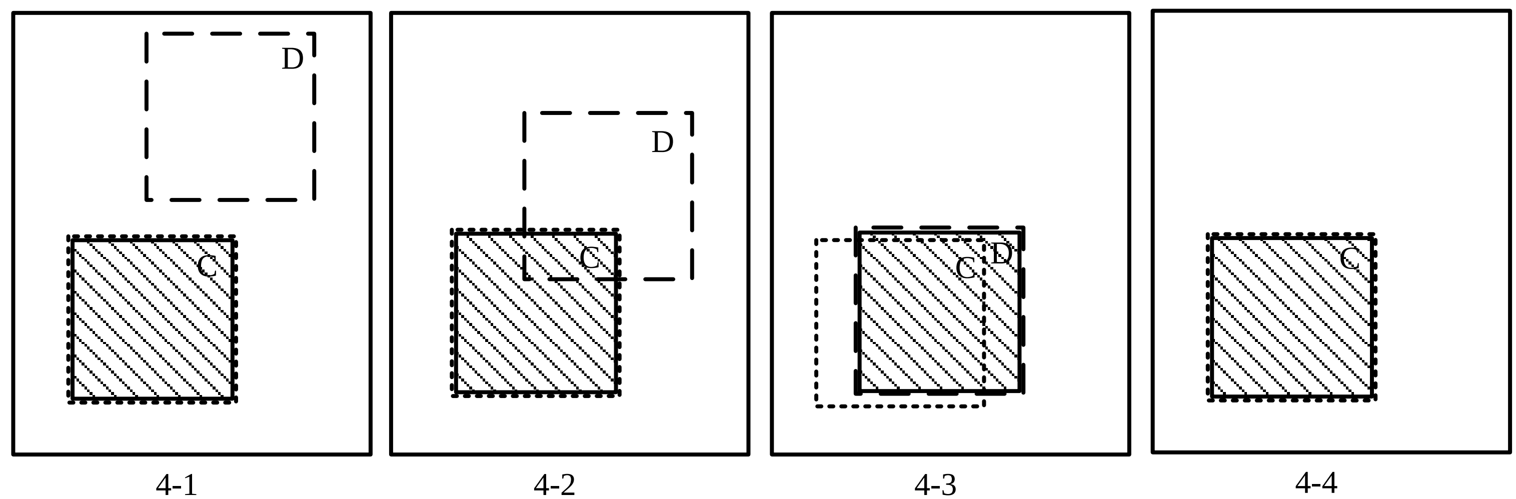

FIG. 4

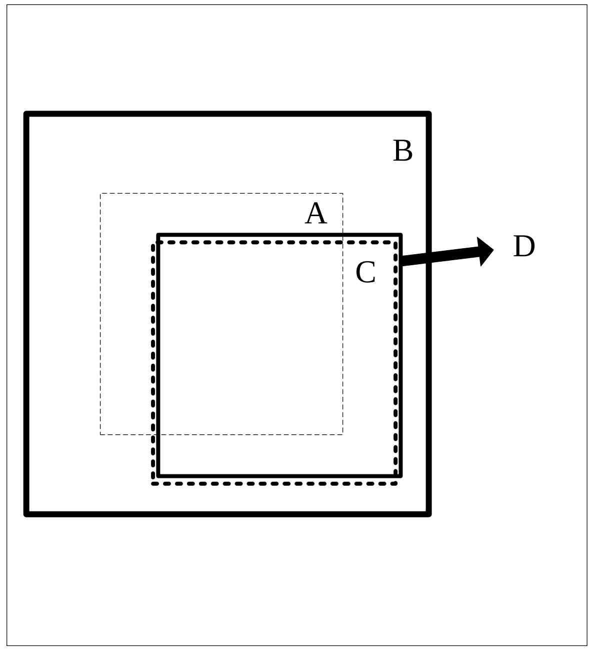
B
A
C
D

FIG. 5

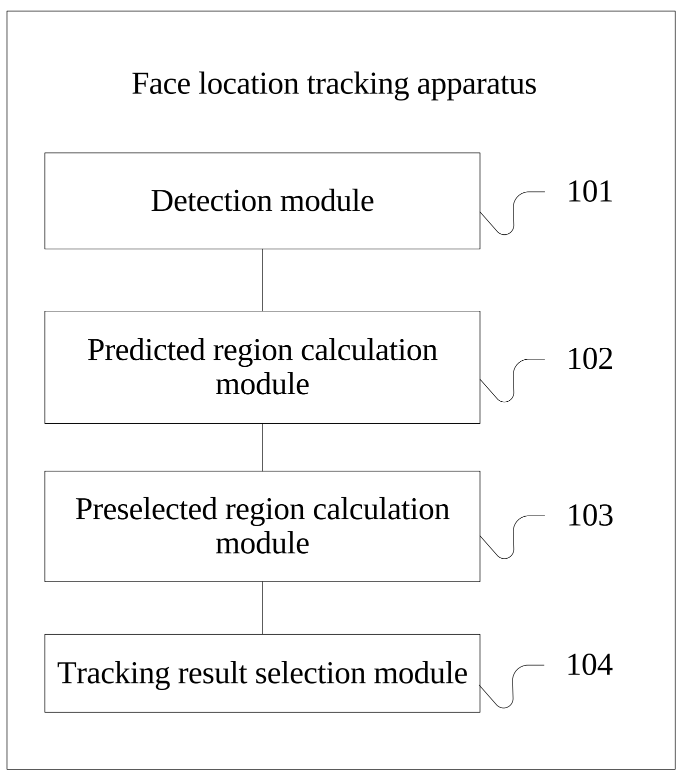
Face location tracking apparatus
Detection module
101
Predicted region calculation module
102
Preselected region calculation module
103
Tracking result selection module
104

FIG. 6

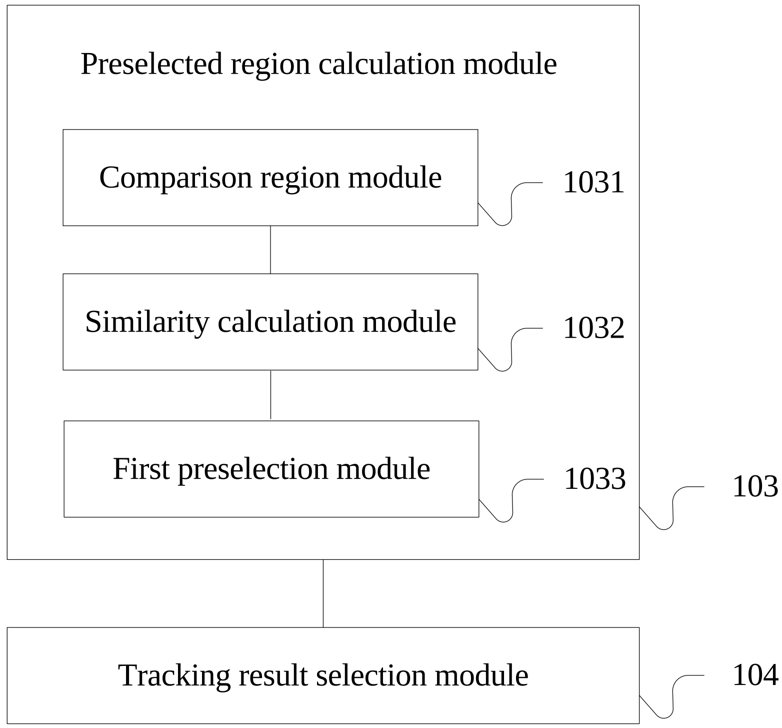
Preselected region calculation module
Comparison region module
1031
Similarity calculation module
1032
First preselection module
1033
103
Tracking result selection module
104

FIG. 7

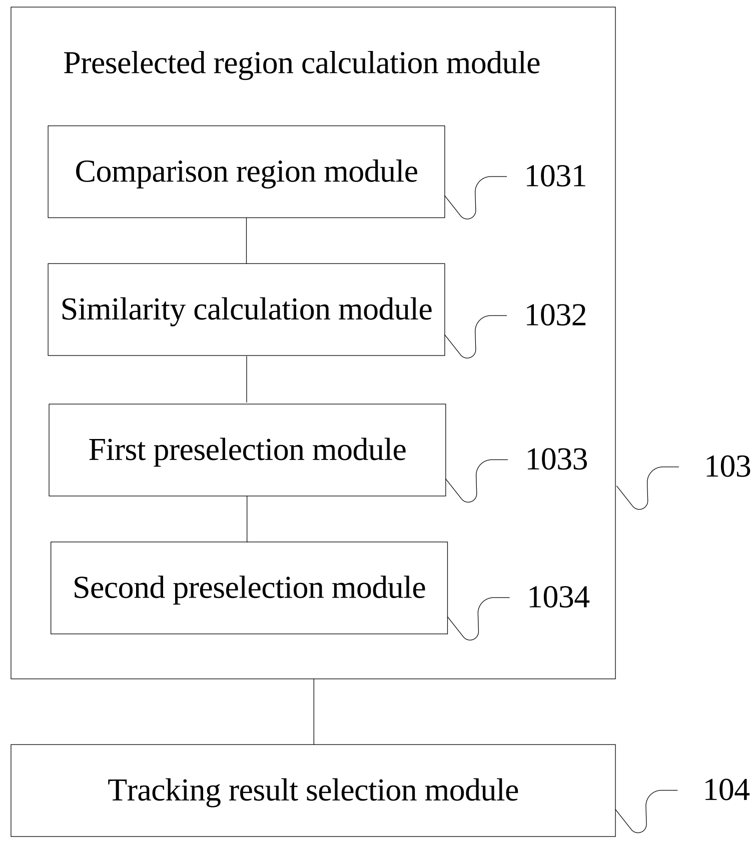
Preselected region calculation module
Comparison region module
1031
Similarity calculation module
1032
First preselection module
1033
Second preselection module
1034
103
Tracking result selection module
104

FIG. 8

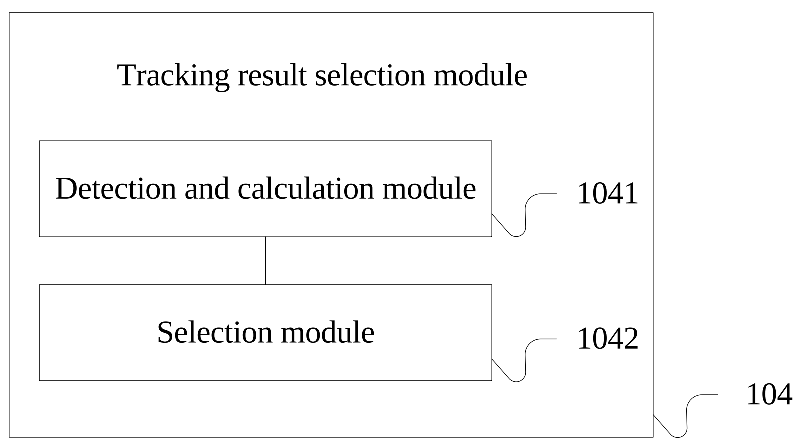
Tracking result selection module
Detection and calculation module
1041
Selection module
1042
104

FIG. 9

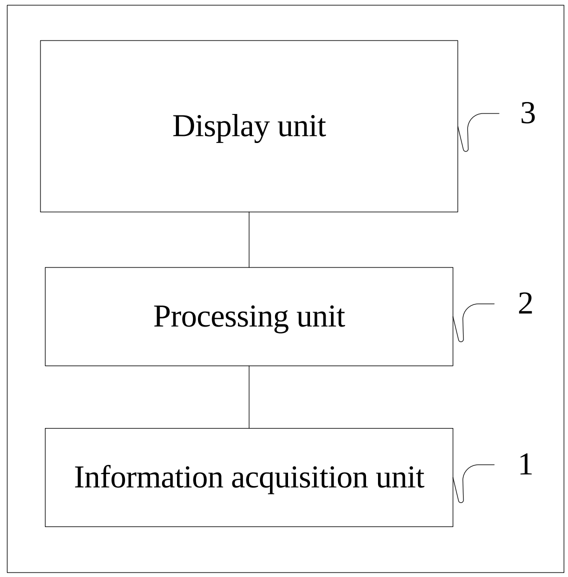
Display unit
3
Processing unit
2
Information acquisition unit
1

FIG. 10