



(12) 发明专利

(10) 授权公告号 CN 113705280 B

(45) 授权公告日 2024.05.10

(21) 申请号 202010436211.5

G06V 40/20 (2022.01)

(22) 申请日 2020.05.21

G06F 3/01 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113705280 A

(56) 对比文件

CN 102054291 A, 2011.05.11

CN 103324283 A, 2013.09.25

(43) 申请公布日 2021.11.26

CN 103605466 A, 2014.02.26

(73) 专利权人 北京聚匠艺传媒有限公司

CN 106991367 A, 2017.07.28

地址 100162 北京市大兴区北兴路(东段)2号1幢三层305房间

CN 108197534 A, 2018.06.22

CN 108537143 A, 2018.09.14

(72) 发明人 李华栋

CN 109151540 A, 2019.01.04

CN 110378994 A, 2019.10.25

(74) 专利代理机构 北京睿智保诚专利代理事务所(普通合伙) 11732

CN 111091031 A, 2020.05.01

US 2015038222 A1, 2015.02.05

专利代理师 周新楣

审查员 罗彩珠

(51) Int. Cl.

G06V 40/16 (2022.01)

G06V 20/40 (2022.01)

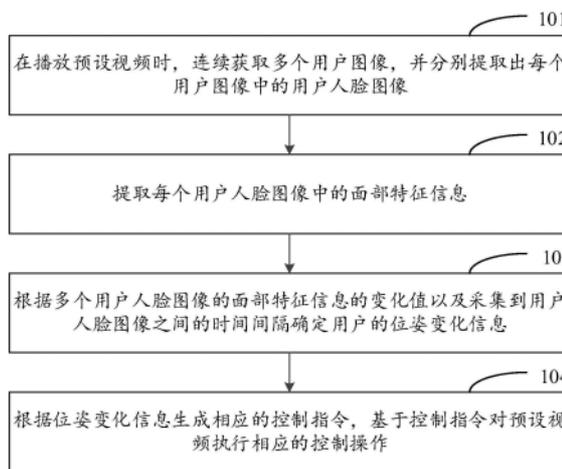
权利要求书3页 说明书10页 附图5页

(54) 发明名称

一种基于面部特征的人机交互方法及装置

(57) 摘要

本发明提供了一种基于面部特征的人机交互方法及装置,其中,该方法包括:在播放预设视频时,连续获取多个用户图像,并分别提取出每个用户图像中的用户人脸图像;提取每个用户人脸图像中的面部特征信息;根据多个用户人脸图像的面部特征信息的变化值以及采集到用户人脸图像之间的时间间隔确定用户的位姿变化信息,位姿变化信息包括角度变化值和角速度变化值;根据位姿变化信息生成相应的控制指令,基于控制指令对预设视频执行相应的控制操作。通过本发明实施例提供的基于面部特征的人机交互方法及装置,通过检测人体面部生物特征和其运动姿态,判断用户人脸当前的运动状态,进而可以发出暂停或播放等控制信号,实现通过人脸运动对视频播放控制。



1. 一种基于面部特征的人机交互方法,其特征在于,包括:

在播放预设视频时,连续获取多个用户图像,并分别提取出每个所述用户图像中的用户人脸图像;

所述提取出每个所述用户图像中的用户人脸图像包括:对所述用户图像进行预处理,所述预处理包括灰度增强、滤波、二值化中的一项或多项;提取出每个预处理后的所述用户图像中的用户人脸图像;

提取每个所述用户人脸图像中的面部特征信息;所述面部特征信息为特征点;人脸的特征点指的是人脸的外轮廓以及五官附近的关键点;

根据多个所述用户人脸图像的面部特征信息的变化值以及采集到所述用户人脸图像之间的时间间隔确定用户的位姿变化信息,所述位姿变化信息包括角度变化值和角速度变化值;面部特征信息的变化值,为人脸的特征点指的是人脸的外轮廓以及五官附近的关键点的变化值;

所述面部特征信息包括特征点的坐标值,所述确定用户的位姿变化信息包括:

预先从用户的特征点中选取至少四个标准特征点 S_a, S_b, S_c, S_d ;其中,在标准人脸图像中,所述标准特征点 S_a 与 S_b 之间的线段与所述标准特征点 S_c 与 S_d 之间的线段之间的夹角与90度之间的差值小于预设值;

确定所述面部特征信息中与四个所述标准特征点分别对应的四个特征点 f_a, f_b, f_c, f_d ,依次确定每个所述用户人脸图像中特征点 f_a 与 f_b 之间的距离、特征点 f_c 与 f_d 之间的距离,并确定特征点 f_a 与 f_b 之间的最大距离值 L_{ab}^{\max} 、特征点 f_c 与 f_d 之间的最大距离值 L_{cd}^{\max} ;

确定所述用户人脸图像对应的第一角度和第二角度:

$$Y_i = \arccos \frac{L_{ab}^i}{L_{ab}^{\max}}, P_i = \arccos \frac{L_{cd}^i}{L_{cd}^{\max}}; \text{其中, } i \in [1, n], n \text{ 为所述用户人脸图像的数量, } Y_i$$

表示第*i*个用户人脸图像的第一角度, P_i 表示第*i*个用户人脸图像的第二角度, L_{ab}^i 表示在第*i*个用户人脸图像中特征点 f_a 与 f_b 之间的距离, L_{cd}^i 表示在第*i*个用户人脸图像中特征点 f_c 与 f_d 之间的距离;

确定第*i*个与第*j*个用户人脸图像之间的第一角度变化值 ΔY 、第二角度变化值 ΔP 、第一角速度变化值 $\Delta \omega_Y$ 、第二角速度变化值 $\Delta \omega_P$:

$$\Delta Y = Y_i - Y_j, \Delta P = P_i - P_j, \Delta \omega_Y = \frac{\Delta Y}{\Delta t_{ij}}, \Delta \omega_P = \frac{\Delta P}{\Delta t_{ij}}; \text{其中, } \Delta t_{ij} \text{ 表示采集到第 } i \text{ 个用户人}$$

脸图像与第*j*个用户人脸图像之间的时间间隔;

确定第*i*个与第*j*个用户人脸图像之间的第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$:

$$\overrightarrow{D_{\Delta Y}} = (x_i^a - x_j^a + x_i^b - x_j^b, y_i^a - y_j^a + y_i^b - y_j^b),$$

$$\overrightarrow{D_{\Delta P}} = (x_i^c - x_j^c + x_i^d - x_j^d, y_i^c - y_j^c + y_i^d - y_j^d);$$

其中,四个特征点 f_a, f_b, f_c, f_d 在第 i 个用户人脸图像中的坐标分别为 $(x_{ia}, y_{ia})、(x_{ib}, y_{ib})、(x_{ic}, y_{ic})、(x_{id}, y_{id})$;特征点 f_a, f_b, f_c, f_d 在第 j 个用户人脸图像中的坐标分别为 $(x_{ja}, y_{ja})、(x_{jb}, y_{jb})、(x_{jc}, y_{jc})、(x_{jd}, y_{jd})$;

根据所述位姿变化信息生成相应的控制指令,基于所述控制指令对所述预设视频执行相应的控制操作;

所述根据所述位姿变化信息生成相应的控制指令包括:

在所述第二角度变化值 ΔP 小于第一预设阈值且所述第一角度变化值 ΔY 大于第二预设阈值时,根据所述第一角度变化值 ΔY 、第一角速度变化值 $\Delta \omega_Y$ 和第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 生成相应的控制指令;

在所述第一角度变化值 ΔY 小于第一预设阈值且所述第二角度变化值 ΔP 大于第二预设阈值时,根据所述第二角度变化值 ΔP 、第二角速度变化值 $\Delta \omega_P$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$ 生成相应的控制指令。

2. 根据权利要求1所述的方法,其特征在于,所述提取每个所述用户人脸图像中的面部特征信息,包括:

提取所述用户人脸图像中的特征点,依次将所述特征点邻域内预设大小的图片与训练好的人脸五官滤波器进行相似性检测;

基于相机内参设置统一的坐标系,确定通过相似性检测的特征点在所述坐标系下的坐标值,所述坐标值为所述面部特征信息中的一项信息。

3. 一种基于面部特征的人机交互装置,其特征在于,包括:

图像获取模块,用于在播放预设视频时,连续获取多个用户图像,并分别提取出每个所述用户图像中的用户人脸图像;

特征提取模块,用于提取每个所述用户人脸图像中的面部特征信息,所述面部特征信息为特征点;人脸的特征点指的是人脸的外轮廓以及五官附近的关键点;

位姿确定模块,用于根据多个所述用户人脸图像的面部特征信息的变化值以及采集到所述用户人脸图像之间的时间间隔确定用户的位姿变化信息,所述位姿变化信息包括角度变化值和角速度变化值;面部特征信息的变化值,为人脸的特征点指的是人脸的外轮廓以及五官附近的关键点的变化值;

所述面部特征信息包括特征点的坐标值,所述位姿确定模块确定用户的位姿变化信息包括:

预先从用户的特征点中选取至少四个标准特征点 S_a, S_b, S_c, S_d ;其中,在标准人脸图像中,所述标准特征点 S_a 与 S_b 之间的线段与所述标准特征点 S_c 与 S_d 之间的线段之间的夹角与90度之间的差值小于预设值;

确定所述面部特征信息中与四个所述标准特征点分别对应的四个特征点 f_a, f_b, f_c, f_d ,依次确定每个所述用户人脸图像中特征点 f_a 与 f_b 之间的距离、特征点 f_c 与 f_d 之间的距离,并确定特征点 f_a 与 f_b 之间的最大距离值 L_{ab}^{\max} 、特征点 f_c 与 f_d 之间的最大距离值 L_{cd}^{\max} ;

确定所述用户人脸图像对应的第一角度和第二角度:

$$Y_i = \arccos \frac{L_{ab}^i}{L_{ab}^{\max}}, P_i = \arccos \frac{L_{cd}^i}{L_{cd}^{\max}}; \text{其中, } i \in [1, n], n \text{ 为所述用户人脸图像的数量, } Y_i$$

表示第*i*个用户人脸图像的第一角度, P_i 表示第*i*个用户人脸图像的第二角度, L_{ab}^i 表示在第*i*个用户人脸图像中特征点 f_a 与 f_b 之间的距离, L_{cd}^i 表示在第*i*个用户人脸图像中特征点 f_c 与 f_d 之间的距离;

确定第*i*个与第*j*个用户人脸图像之间的第一角度变化值 ΔY 、第二角度变化值 ΔP 、第一角速度变化值 $\Delta \omega_Y$ 、第二角速度变化值 $\Delta \omega_P$:

$$\Delta Y = Y_i - Y_j, \Delta P = P_i - P_j, \Delta \omega_Y = \frac{\Delta Y}{\Delta t_{ij}}, \Delta \omega_P = \frac{\Delta P}{\Delta t_{ij}}; \text{其中, } \Delta t_{ij} \text{ 表示采集到第 } i \text{ 个用户人脸}$$

图像与第*j*个用户人脸图像之间的时间间隔;

确定第*i*个与第*j*个用户人脸图像之间的第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$:

$$\overrightarrow{D_{\Delta Y}} = (xi_a - xj_a + xi_b - xj_b, yi_a - yj_a + yi_b - yj_b),$$

$$\overrightarrow{D_{\Delta P}} = (xi_c - xj_c + xi_d - xj_d, yi_c - yj_c + yi_d - yj_d);$$

其中, 四个特征点 f_a, f_b, f_c, f_d 在第*i*个用户人脸图像中的坐标分别为 (xi_a, yi_a) 、 (xi_b, yi_b) 、 (xi_c, yi_c) 、 (xi_d, yi_d) ; 特征点 f_a, f_b, f_c, f_d 在第*j*个用户人脸图像中的坐标分别为 (xj_a, yj_a) 、 (xj_b, yj_b) 、 (xj_c, yj_c) 、 (xj_d, yj_d) ;

控制模块, 用于根据所述位姿变化信息生成相应的控制指令, 基于所述控制指令对所述预设视频执行相应的控制操作;

所述控制模块用于: 在所述第二角度变化值 ΔP 小于第一预设阈值且所述第一角度变化值 ΔY 大于第二预设阈值时, 根据所述第一角度变化值 ΔY 、第一角速度变化值 $\Delta \omega_Y$ 和第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 生成相应的控制指令;

在所述第一角度变化值 ΔY 小于第一预设阈值且所述第二角度变化值 ΔP 大于第二预设阈值时, 根据所述第二角度变化值 ΔP 、第二角速度变化值 $\Delta \omega_P$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$ 生成相应的控制指令。

4. 根据权利要求3所述的装置, 其特征在于, 所述特征提取模块包括:

提取单元, 用于提取所述用户人脸图像中的特征点, 依次将所述特征点邻域内预设大小的图片与训练好的人脸五官滤波器进行相似性检测;

坐标确定单元, 用于基于相机内参设置统一的坐标系, 确定通过相似性检测的特征点在所述坐标系下的坐标值, 所述坐标值为所述面部特征信息中的一项信息。

5. 根据权利要求3-4任一所述的装置, 其特征在于, 所述图像获取模块用于:

对所述用户图像进行预处理, 所述预处理包括灰度增强、滤波、二值化中的一项或多项;

提取出每个预处理后的所述用户图像中的用户人脸图像。

一种基于面部特征的人机交互方法及装置

技术领域

[0001] 本发明涉及人机交互技术领域,具体而言,涉及一种基于面部特征的人机交互方法及装置。

背景技术

[0002] 在书法教育过程中,身体状态包括持笔姿势,手部动作,身体姿态等。心理状态变化有注意力焦点转移,学习兴趣波动,学习节奏打断等。传统书法信息化教学过程中,学员可以一边看教学视频一边练习书法,而在观看教学视频时,可能需要学员操作播放教学视频的设备,从而可能改变学员的身体状态或心理状态,身体和心理变化构成了学生学习过程中的重大障碍。

[0003] 现有控制播放设备的方法有手动点击和语音控制两种方法。手动点击是指用手指点击移动端界面上的按钮来控制播放、暂停、快进和后退等。语音控制方法是用语音识别系统,将语音信号变成控制信号,控制视频播放、暂停、快进和后退等。

[0004] 手动点击控制方法,需要学习者放下笔,改变写字姿态,并且导致写字心态波动。语音进行控制播放暂停,需要学习者从静态转到说话的动态,会对心理状态有影响,且语音识别容易产生歧义,且不适用于多人同时学习场景。

发明内容

[0005] 为解决上述问题,本发明实施例的目的在于提供一种基于面部特征的人机交互方法及装置。

[0006] 第一方面,本发明实施例提供了一种基于面部特征的人机交互方法,包括:

[0007] 在播放预设视频时,连续获取多个用户图像,并分别提取出每个所述用户图像中的用户人脸图像;

[0008] 提取每个所述用户人脸图像中的面部特征信息;

[0009] 根据多个所述用户人脸图像的面部特征信息的变化值以及采集到所述用户人脸图像之间的时间间隔确定用户的位姿变化信息,所述位姿变化信息包括角度变化值和角速度变化值;

[0010] 根据所述位姿变化信息生成相应的控制指令,基于所述控制指令对所述预设视频执行相应的控制操作。

[0011] 第二方面,本发明实施例还提供了一种基于面部特征的人机交互装置,包括:

[0012] 图像获取模块,用于在播放预设视频时,连续获取多个用户图像,并分别提取出每个所述用户图像中的用户人脸图像;

[0013] 特征提取模块,用于提取每个所述用户人脸图像中的面部特征信息;

[0014] 位姿确定模块,用于根据多个所述用户人脸图像的面部特征信息的变化值以及采集到所述用户人脸图像之间的时间间隔确定用户的位姿变化信息,所述位姿变化信息包括角度变化值和角速度变化值;

[0015] 控制模块,用于根据所述位姿变化信息生成相应的控制指令,基于所述控制指令对所述预设视频执行相应的控制操作。

[0016] 本发明实施例上述第一方面提供的方案中,基于面部特征的人机交互方法,通过检测人体面部生物特征和其运动姿态,判断用户人脸当前的运动状态,进而可以发出暂停或播放等控制信号,实现通过人脸运动对视频播放控制。该交互方式简答,不需要用户大幅度移动身体,可以避免传统方法对书写状态的破坏,并且容易使学生保持学习心态的连续性。本申请提供的交互方式对于书法教育的信息化过程提供基础性的促进作用,具有重要的现实价值。

[0017] 为使本发明的上述目的、特征和优点能更明显易懂,下文特举较佳实施例,并配合所附附图,作详细说明如下。

附图说明

[0018] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0019] 图1示出了本发明实施例所提供的一种基于面部特征的人机交互方法的流程图;

[0020] 图2示出了本发明实施例所提供的基于面部特征的人机交互方法中,提取出的人脸特征点的示意图;

[0021] 图3示出了本发明实施例所提供的一种基于面部特征的人机交互方法中,提取每个用户人脸图像中的面部特征信息的方法流程图;

[0022] 图4示出了本发明实施例所提供的基于面部特征的人机交互方法中,人脸特征点的主视示意图;

[0023] 图5示出了本发明实施例所提供的基于面部特征的人机交互方法中,人脸特征点的俯视示意图;

[0024] 图6示出了本发明实施例所提供的基于面部特征的人机交互方法中,头部运动姿态的示意图;

[0025] 图7示出了本发明实施例所提供的一种基于面部特征的人机交互装置的结构示意图;

[0026] 图8示出了本发明实施例所提供的用于执行基于面部特征的人机交互方法的电子设备的结构示意图。

具体实施方式

[0027] 在本发明的描述中,需要理解的是,术语“中心”、“纵向”、“横向”、“长度”、“宽度”、“厚度”、“上”、“下”、“前”、“后”、“左”、“右”、“竖直”、“水平”、“顶”、“底”“内”、“外”、“顺时针”、“逆时针”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此不能理解为对本发明的限制。

[0028] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性

或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。在本发明的描述中,“多个”的含义是两个或两个以上,除非另有明确具体的限定。

[0029] 在本发明中,除非另有明确的规定和限定,术语“安装”、“相连”、“连接”、“固定”等术语应做广义理解,例如,可以是固定连接,也可以是可拆卸连接,或一体地连接;可以是机械连接,也可以是电连接;可以是直接相连,也可以通过中间媒介间接相连,可以是两个元件内部的连通。对于本领域的普通技术人员而言,可以根据具体情况理解上述术语在本发明中的具体含义。

[0030] 本发明实施例提供的一种基于面部特征的人机交互方法,参见图1所示,包括:

[0031] 步骤101:在播放预设视频时,连续获取多个用户图像,并分别提取出每个用户图像中的用户人脸图像。

[0032] 本发明实施例中,预设视频指的是用户当前需要观看的视频,比如用户在练习书法时所需要观看的教学视频等。在通过某个设备播放该预设视频时,可以通过摄像头等器件采集设备前方用户的用户图像。该设备具体可以为智能手机、平板电脑、计算机等,本实施例对此不做限定。在获取到用户图像后,通过人脸检测技术即可提取出用户图像中的用户人脸图像。

[0033] 一般来说,由于光照、拍摄角度等原因,会使用户图像存在过度曝光、不合理的对比度及扭曲等现象,同时由摄像头或者相机获得的图像背景还有可能具有很严重的污染噪声,因此存在失真、污染、断裂、模糊、扭曲等各种不利因素,而这些因素会影响后续特征提取的正确率。可选的,本实施例还包括对用户图像进行预处理的过程,通过预处理去除背景干扰和噪声以获得良好的识别率,可以提高后续特征提取的准确性。用户图像的预处理主要采用以下几个环节:图像的灰度增强、滤波、形态学处理及自适应二值化等。在对用户图像进行预处理后即可提取出预处理后的用户图像中的用户人脸图像。

[0034] 此外,本实施例中可通过人脸检测网络提取用户图像中的用户人脸图像。具体的,该人脸检测网络是一个总计16层的卷积神经网络,由1个卷积层、12个Inception层、3个池化层和1个全连接层组成。网络输入为 256×256 像素大小的用户图像,输出是一个256维的特征向量,其中包含已提取的人脸特征信息。

[0035] 网络的前端部分由卷积层和池化层组成,这部分结构的功能用以提取用户图像最基本的点、线、交点等低级特征。网络的主体部分由12个Inception层和2个池化层组成。从网络设计的角度分析,这14层结构负责从简单到复杂对前端输入进行排列组合,在网络训练的过程中学习能描述人脸差异的结构化特征,最终压缩至1024维特征向量。网络的输出端由全连接层构成,该层结构将输入的1024维特征向量压缩到256维进行输出。这种设计能随机屏蔽1024维向量到256维向量之间的连接,减轻网络训练时产生的过拟合现象,并最终提取出用户人脸图像。

[0036] 步骤102:提取每个用户人脸图像中的面部特征信息。

[0037] 本发明实施例中,面部特征信息包括从用户人脸图像中提取的特征点,或者与特征点相关的信息,比如特征点的坐标值等。其中,人脸的特征点指的是人脸的外轮廓以及五官附近的关键点,包括眉毛的外轮廓,眼睛的上下轮廓,鼻子的中线,嘴唇的上下轮廓等。如图2所示。使用人脸特征识别算法能够从用户人脸图像中定位出多个特征点,并可确定每个

特征点的坐标,该坐标包含了整个人脸的姿态信息。图2中以定位出68个特征点(从0~67)为例说明。

[0038] 步骤103:根据多个用户人脸图像的面部特征信息的变化值以及采集到用户人脸图像之间的时间间隔确定用户的位姿变化信息,位姿变化信息包括角度变化值和角速度变化值。

[0039] 本发明实施例中会连续采集多个用户图像,即可以连续获取到多个用户人脸图像,通过不同用户人脸图像的不同的面部特征信息,可以确定面部特征信息之间的变化值;同时,获取到不同用户图像之间存在时间间隔,基于该变化值以及时间间隔即可确定速度变化值。同时,本实施例中将与角度相关的信息作为位姿变化信息,以该位姿变化信息来描述用户的头部转动情况。

[0040] 步骤104:根据位姿变化信息生成相应的控制指令,基于控制指令对预设视频执行相应的控制操作。

[0041] 本发明实施例中,通过位姿变化信息来表征用户头部的转动情况,进而可以生产与该位姿变化信息相对应的控制指令。例如,用户头部左转则生成快退指令,用户头部右转则生成快进指令,用户头部向上转动则生成暂停指令等。在声称该控制指令后即可控制该预设视频的播放情况,比如快进播放等。

[0042] 本发明实施例提供的一种基于面部特征的人机交互方法,通过检测人体面部生物特征和其运动姿态,判断用户人脸当前的运动状态,进而可以发出暂停或播放等控制信号,实现通过人脸运动对视频播放控制。该交互方式简答,不需要用户大幅度移动身体,可以避免传统方法对书写状态的破坏,并且容易使学生保持学习心态的连续性。本申请提供的交互方式对于书法教育的信息化过程提供基础性的促进作用,具有重要的现实价值。

[0043] 在上述实施例的基础上,参见图3所示,步骤102“提取每个用户人脸图像中的面部特征信息”包括:

[0044] 步骤1021:提取用户人脸图像中的特征点,依次将特征点邻域内预设大小的图片与训练好的人脸五官滤波器进行相似性检测。

[0045] 步骤1022:基于相机内参设置统一的坐标系,确定通过相似性检测的特征点在坐标系下的坐标值,坐标值为面部特征信息中的一项信息。

[0046] 本发明实施例中,预先使用带标注的人脸五官样本集进行训练,训练完成后每一组参数都代表一个特定的小片滤波器,即人脸五官滤波器,比如眼睛滤波器、嘴巴滤波器等。通过训练好人脸五官滤波器,能对某特征点坐标邻域内的一小块图片进行相似性检测。例如,眼睛特征点附近采样的图片使用眼睛的小片模型进行检测,嘴巴特征点附近采样的图片使用嘴巴的小片模型进行检测。其中,可采用asm算法得到用户人脸图像中的特征点。

[0047] 同时,用户采集用户图像的相机具有特定的相机内参,该相机内参是固定的,基于该相机内参可以设置坐标系,比如世界坐标系。在特征点通过相似性检测时,说明该特征点是有效的特征点,可以作为后续处理过程的参考,故此时可以确定该特征点的坐标值,以方便后续位姿参数的计算。

[0048] 在上述实施例的基础上,若相机可以采集三维的人脸图像,则根据人脸特征点的三维坐标的变化可以方便确定头部转动情况。但是由于一般设备的相机只能采集二维的平面图像,即相机采集的用户人脸图像不具有深度信息,此时利用二维的平面图像来确定头

部转动时需要大量处理过程,会降低处理效率。本实施例中通过选取部分特征点的方式来确定用户头部转动情况,具体的,面部特征信息包括特征点的坐标值,上述步骤103“确定用户的位姿变化信息”包括:

[0049] 步骤A1:预先从用户的特征点中选取至少四个标准特征点 S_a, S_b, S_c, S_d ;其中,在标准人脸图像中,标准特征点 S_a 与 S_b 之间的线段与标准特征点 S_c 与 S_d 之间的线段之间的夹角与90度之间的差值小于预设值。

[0050] 本发明实施例中,预先确定至少四个特征点,即四个标准特征点,该四个特征点中两个特征点 S_a 与 S_b 之间的线段与特征点 S_c 与 S_d 之间的线段之间的夹角与90度之间的差值小于预设值,即两个线段之间的夹角近似为90度。

[0051] 具体的,人脸中某些特征点的位置是基本固定的,例如眼睛的位置等;同时,若平面内存在相互垂直的两条线段,在三维空间中,该两条线段围绕某些特定的轴旋转才可以使得该两条线段投影到该平面上的线段仍然是垂直的。用户人脸图像是二维图像,通过选取垂直的四个特征点可以正确识别出用户头部围绕某些特定的轴旋转的位姿变化。例如,如图2所示,四个标准特征点依次是左眼最左侧的特征点36、右眼最右侧的特征点45、鼻尖的特征点33、下巴的特征点8。图4示出了四个标准特征点的位置示意图,四个标准特征点 S_a, S_b, S_c, S_d 分别为图4中的A、B、C、D,AB与CD之间的夹角 θ 大约为90度。

[0052] 需要说明的是,步骤A1中确定四个标准特征点的过程为预先执行的过程,即不需要在采集到用户人脸图像后才确定该四个标准特征点。

[0053] 步骤A2:确定面部特征信息中与四个标准特征点分别对应的四个特征点 f_a, f_b, f_c, f_d ,依次确定每个用户人脸图像中特征点 f_a 与 f_b 之间的距离、特征点 f_c 与 f_d 之间的距离,并确定特征点 f_a 与 f_b 之间的最大距离值 L_{ab}^{\max} 、特征点 f_c 与 f_d 之间的最大距离值 L_{cd}^{\max} 。

[0054] 本发明实施例中,为了统一确定角度的基准,将特征点之间最大距离值作为基准。具体的,在不考虑用户人脸与相机之间的距离时,当用户人脸所在平面与相机所采集的平面平行时,特征点之间的距离最大;当用户人脸发生偏移时,则特征点之间的距离会减小,通过特征点之间距离以及最大距离值即可确定可以表征头部姿态的角度。

[0055] 步骤A3:确定用户人脸图像对应的第一角度和第二角度:

[0056] $Y_i = \arccos \frac{L_{ab}^i}{L_{ab}^{\max}}, P_i = \arccos \frac{L_{cd}^i}{L_{cd}^{\max}}$;其中, $i \in [1, n]$, n 为用户人脸图像的数量, Y_i

表示第 i 个用户人脸图像的第一角度, P_i 表示第 i 个用户人脸图像的第二角度, L_{ab}^i 表示在第 i 个用户人脸图像中特征点 f_a 与 f_b 之间的距离, L_{cd}^i 表示在第 i 个用户人脸图像中特征点 f_c 与 f_d 之间的距离。

[0057] 本发明实施例中,用户人脸图像两个特征点之间的线段可以认为是最大距离线段所在平面的投影,故第一角度为 $Y_i = \arccos \frac{L_{ab}^i}{L_{ab}^{\max}}$ 。具体的,为了方便说明,图4表示人脸的

正视图,图4中的四个特征点A、B、C、D(对应用户人脸图像中的四个特征点 f_a, f_b, f_c, f_d)处于同一个平面,且图4中的人脸对应特征点之间距离最大的情况,即图4中AB之间的距离为

L_{ab}^{\max} , CD之间的距离为 L_{cd}^{\max} 。此外,参见图5所示,图5表示人脸的俯视图,图5中的线段AB表示图4中四个特征点A、B、C、D所在的平面,且在图5中AB的距离仍为 L_{ab}^{\max} 。在实际情况下,以相机作为参考物,若用户的头部在左右方向上发生了偏转(即用户向左或向右摇头),即在真实的世界坐标系下,用户头部的特征点A和B此时位于A1位置和B1位置,在不考虑用户头部上下方向旋转(即用户没有点头或抬头)的情况下,图5中线段A1B1的距离仍然是 L_{ab}^{\max} ;但是由于相机只能采集到二维的图像,即相机采集到的用户人脸图像中,特征点A和B映射到了Ai和Bi处,对于采集的第i个用户人脸图像来说,Ai和Bi是该用户人脸图像中特征点的位置,即线段AiBi之间的距离即为特征点 f_a 与 f_b 之间的距离 L_{ab}^i ,故此时用户在左右方向上所偏转的角度为 $Y_i = \arccos \frac{L_{ab}^i}{L_{ab}^{\max}}$ 。同理,可以确定用户在上下方向上偏转的角度为 $P_i = \arccos \frac{L_{cd}^i}{L_{cd}^{\max}}$ 。

[0058] 步骤A4:确定第i个与第j个用户人脸图像之间的第一角度变化值 ΔY 、第二角度变化值 ΔP 、第一角速度变化值 $\Delta \omega_Y$ 、第二角速度变化值 $\Delta \omega_P$:

[0059] $\Delta Y = Y_i - Y_j$, $\Delta P = P_i - P_j$, $\Delta \omega_Y = \frac{\Delta Y}{\Delta t_{ij}}$, $\Delta \omega_P = \frac{\Delta P}{\Delta t_{ij}}$;其中, Δt_{ij} 表示采集到第i个用户人脸图像与第j个用户人脸图像之间的时间间隔。

[0060] 本发明实施例中,每个用户人脸图像均可以确定其本身的角度,即第j个用户人脸图像的第一角度为 Y_j ,第二角度为 P_j ;通过两个用户人脸图像(比如相邻的两帧用户人脸图像)的角度值确定角度变化值,并根据时间间隔可以确定角速度变化值。

[0061] 步骤A5:确定第i个与第j个用户人脸图像之间的第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$:

[0062] $\overrightarrow{D_{\Delta Y}} = (xi_a - xj_a + xi_b - xj_b, yi_a - yj_a + yi_b - yj_b)$,

[0063] $\overrightarrow{D_{\Delta P}} = (xi_c - xj_c + xi_d - xj_d, yi_c - yj_c + yi_d - yj_d)$;

[0064] 其中,四个特征点 f_a, f_b, f_c, f_d 在第i个用户人脸图像中的坐标分别为 (xi_a, yi_a) 、 (xi_b, yi_b) 、 (xi_c, yi_c) 、 (xi_d, yi_d) ;特征点 f_a, f_b, f_c, f_d 在第j个用户人脸图像中的坐标分别为 (xj_a, yj_a) 、 (xj_b, yj_b) 、 (xj_c, yj_c) 、 (xj_d, yj_d) 。

[0065] 本发明实施例中,在上述步骤A4中可以确定角度变化值,即可以确定用户头部发生了左右旋转或上下旋转,但是不能区分是向左旋转还是向右旋转,本实施例中利用特征点的坐标值的变化来确定旋转方向。具体的,如图6所示,人体头部旋转时,头部运动姿态可以看做头部在三维空间中的六个方向的运动,包括头部上、下、左、右、前、后六个方向以及上下左右四个方向和前后两个方向间的各种组合姿态。欧拉角在一个固定的坐标系下描述刚体运动的三个角度取向,任意方向都可以通过偏航角(Yaw)、俯仰角(Pitch)、滚转角(Roll)三个角度的组合来表达,它能精确表达头部旋转角度。如图6所示,偏航角是指头部的左右旋转产生的角度;俯仰角是指头部的上下旋转产生的角度;滚转角是指在平面内旋转产生的角度。基于人体的特点,头部在发生偏转时是基于颈部运动的,即头部旋转时的旋转轴位于颈部处,故当头部旋转时,人脸的特征点整体也会发生位移。例如,在颈部作用下

头部向右转动时,用户人脸图像中的四个特征点 f_a, f_b, f_c, f_d 整体也会向右移动,本实施例中基于该特点确定头部的旋转方向。

[0066] 具体的,先采集到第j个用户人脸图像,后采集到第i个用户人脸图像,即采集到第j个用户人脸图像的时间先于采集到第i个用户人脸图像的时间。在确定第一角度变化方向时,第j个用户人脸图像中特征点 f_a 和 f_b 的坐标分别为 (xj_a, yj_a) 、 (xj_b, yj_b) ;之后,在第i个用户人脸图像中,特征点 f_a 和 f_b 的坐标分别为 (xi_a, yi_a) 、 (xi_b, yi_b) ,即特征点 f_a 从 (xj_a, yj_a) 移动到了 (xi_a, yi_a) ,其移动向量为 $\vec{a} = (xi_a - xj_a, yi_a - yj_a)$;同理,特征点 f_b 从 (xj_b, yj_b) 移动到了 (xi_b, yi_b) ,同样的,其移动向量为 $\vec{b} = (xi_b - xj_b, yi_b - yj_b)$,则特征点 f_a 和 f_b 整体的移动向量 $\overrightarrow{D_{\Delta Y}} = \vec{a} + \vec{b}$,即:

$$[0067] \quad \overrightarrow{D_{\Delta Y}} = (xi_a - xj_a + xi_b - xj_b, yi_a - yj_a + yi_b - yj_b)。$$

[0068] 同理,在发生上下旋转时,对于特征点 f_c, f_d ,第二角度变化方向 $\overrightarrow{D_{\Delta P}}$ 为:

$$\overrightarrow{D_{\Delta P}} = (xi_c - xj_c + xi_d - xj_d, yi_c - yj_c + yi_d - yj_d)。$$

[0069] 本发明实施例中,基于头部旋转的特点,利用二维的用户人脸图像即可确定两个用户人脸图像之间的角度变化值、角速度变化值以及角度变化方向等位姿变化信息,从而可以确定用户头部的旋转位移、旋转速度、旋转方向,进而方便后续生成与用户旋转姿态相一致的控制指令。

[0070] 在上述实施例的基础上,上述确定位姿变化信息的过程需要用户人脸与相机之间的距离基本保持一致,若用户人脸与相机之间的距离发生变化,则可能导致确定的位姿变化信息不准确。故本实施例中,上述步骤104“根据位姿变化信息生成相应的控制指令”包括:

[0071] 步骤B1:在第二角度变化值 ΔP 小于第一预设阈值且第一角度变化值 ΔY 大于第二预设阈值时,根据第一角度变化值 ΔY 、第一角速度变化值 $\Delta \omega_Y$ 和第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 生成相应的控制指令。

[0072] 步骤B2:在第一角度变化值 ΔY 小于第一预设阈值且第二角度变化值 ΔP 大于第二预设阈值时,根据第二角度变化值 ΔP 、第二角速度变化值 $\Delta \omega_P$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$ 生成相应的控制指令。

[0073] 本发明实施例中,当第二角度变化值 ΔP 小于第一预设阈值时,由于在当前时间段最大距离值 L_{cd}^{\max} 是固定的,故说明两个用户人脸图像之间特征点 f_c 与 f_d 之间的距离变化不大,此时基本可以认为用户人脸与相机之间的距离基本保持一致,且用户头部在特征点 f_c 与 f_d 方向(比如上下旋转方向)上几乎没有旋转。同时,若第一角度变化值 ΔY 大于第二预设阈值,则说明用户头部在特征点 f_a 与 f_b 方向(比如左右旋转方向)上有较大的旋转,此时即可确定用户头部在 f_a 与 f_b 方向上发生了旋转,之后即可根据第一角度变化值 ΔY 、第一角速度变化值 $\Delta \omega_Y$ 和第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 生成相应的控制指令。例如,根据第一角度变化方向

$\overrightarrow{D_{\Delta Y}}$ 可以确定用户头部为向右旋转,根据第一角度变化值 ΔY 以及第一角速度变化值 $\Delta \omega_Y$ 的大小可以确定用户头部旋转的幅度足够大,此时则可生成与用户头部向右旋转对应的控制指令,比如快进指令。

[0074] 同理,在步骤B2中,当第一角度变化值 ΔY 小于第一预设阈值且第二角度变化值 ΔP 大于第二预设阈值时,也可生成其他相应的控制指令,比如用户头部向下旋转(即点头时)生成用于播放视频的控制指令。

[0075] 本发明实施例提供的一种基于面部特征的人机交互方法,通过检测人体面部生物特征和其运动姿态,判断用户人脸当前的运动状态,进而可以发出暂停或播放等控制信号,实现通过人脸运动对视频播放控制。该交互方式简答,不需要用户大幅度移动身体,可以避免传统方法对书写状态的破坏,并且容易使学生保持学习心态的连续性。本申请提供的交互方式对于书法教育的信息化过程提供基础性的促进作用,具有重要的现实价值。同时,利用二维的用户人脸图像即可确定两个用户人脸图像之间的角度变化值、角速度变化值以及角度变化方向等位姿变化信息,从而可以确定用户头部的旋转位移、旋转速度、旋转方向,进而方便后续生成与用户旋转姿态相一致的控制指令。

[0076] 以上详细介绍了基于面部特征的人机交互方法流程,该方法也可以通过相应的装置实现,下面详细介绍该装置的结构和功能。

[0077] 本发明实施例提供了一种基于面部特征的人机交互装置,参见图7所示,该装置包括:

[0078] 图像获取模块71,用于在播放预设视频时,连续获取多个用户图像,并分别提取出每个所述用户图像中的用户人脸图像;

[0079] 特征提取模块72,用于提取每个所述用户人脸图像中的面部特征信息;

[0080] 位姿确定模块73,用于根据多个所述用户人脸图像的面部特征信息的变化值以及采集到所述用户人脸图像之间的时间间隔确定用户的位姿变化信息,所述位姿变化信息包括角度变化值和角速度变化值;

[0081] 控制模块74,用于根据所述位姿变化信息生成相应的控制指令,基于所述控制指令对所述预设视频执行相应的控制操作。

[0082] 在上述实施例的基础上,所述特征提取模块72包括:

[0083] 提取单元,用于提取所述用户人脸图像中的特征点,依次将所述特征点邻域内预设大小的图片与训练好的人脸五官滤波器进行相似性检测;

[0084] 坐标确定单元,用于基于相机内参设置统一的坐标系,确定通过相似性检测的特征点在所述坐标系下的坐标值,所述坐标值为所述面部特征信息中的一项信息。

[0085] 在上述实施例的基础上,所述面部特征信息包括特征点的坐标值,所述位姿确定模块73确定用户的位姿变化信息包括:

[0086] 预先从用户的特征点中选取至少四个标准特征点 S_a, S_b, S_c, S_d ; 其中,在标准人脸图像中,所述标准特征点 S_a 与 S_b 之间的线段与所述标准特征点 S_c 与 S_d 之间的线段之间的夹角与90度之间的差值小于预设值;

[0087] 确定所述面部特征信息中与四个所述标准特征点分别对应的四个特征点 f_a, f_b, f_c, f_d , 依次确定每个所述用户人脸图像中特征点 f_a 与 f_b 之间的距离、特征点 f_c 与 f_d 之间的距

离,并确定特征点 f_a 与 f_b 之间的最大距离值 L_{ab}^{\max} 、特征点 f_c 与 f_d 之间的最大距离值 L_{cd}^{\max} ;

[0088] 确定所述用户人脸图像对应的第一角度和第二角度:

[0089] $Y_i = \arccos \frac{L_{ab}^i}{L_{ab}^{\max}}, P_i = \arccos \frac{L_{cd}^i}{L_{cd}^{\max}}$;其中, $i \in [1, n]$, n 为所述用户人脸图像的数量, Y_i 表示第 i 个用户人脸图像的第一角度, P_i 表示第 i 个用户人脸图像的第二角度, L_{ab}^i 表示在第 i 个用户人脸图像中特征点 f_a 与 f_b 之间的距离, L_{cd}^i 表示在第 i 个用户人脸图像中特征点 f_c 与 f_d 之间的距离;

[0090] 确定第 i 个与第 j 个用户人脸图像之间的第一角度变化值 ΔY 、第二角度变化值 ΔP 、第一角速度变化值 $\Delta \omega_Y$ 、第二角速度变化值 $\Delta \omega_P$:

[0091] $\Delta Y = Y_i - Y_j, \Delta P = P_i - P_j, \Delta \omega_Y = \frac{\Delta Y}{\Delta t_{ij}}, \Delta \omega_P = \frac{\Delta P}{\Delta t_{ij}}$;其中, Δt_{ij} 表示采集到第 i 个用户人脸图像与第 j 个用户人脸图像之间的时间间隔;

[0092] 确定第 i 个与第 j 个用户人脸图像之间的第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$:

[0093] $\overrightarrow{D_{\Delta Y}} = (xi_a - xj_a + xi_b - xj_b, yi_a - yj_a + yi_b - yj_b),$

[0094] $\overrightarrow{D_{\Delta P}} = (xi_c - xj_c + xi_d - xj_d, yi_c - yj_c + yi_d - yj_d);$

[0095] 其中,四个特征点 f_a, f_b, f_c, f_d 在第 i 个用户人脸图像中的坐标分别为 $(xi_a, yi_a), (xi_b, yi_b), (xi_c, yi_c), (xi_d, yi_d)$;特征点 f_a, f_b, f_c, f_d 在第 j 个用户人脸图像中的坐标分别为 $(xj_a, yj_a), (xj_b, yj_b), (xj_c, yj_c), (xj_d, yj_d)$ 。

[0096] 在上述实施例的基础上,所述控制模块74用于:

[0097] 在所述第二角度变化值 ΔP 小于第一预设阈值且所述第一角度变化值 ΔY 大于第二预设阈值时,根据所述第一角度变化值 ΔY 、第一角速度变化值 $\Delta \omega_Y$ 和第一角度变化方向 $\overrightarrow{D_{\Delta Y}}$ 生成相应的控制指令;

[0098] 在所述第一角度变化值 ΔY 小于第一预设阈值且所述第二角度变化值 ΔP 大于第二预设阈值时,根据所述第二角度变化值 ΔP 、第二角速度变化值 $\Delta \omega_P$ 和第二角度变化方向 $\overrightarrow{D_{\Delta P}}$ 生成相应的控制指令。

[0099] 在上述实施例的基础上,所述图像获取模块71用于:

[0100] 对所述用户图像进行预处理,所述预处理包括灰度增强、滤波、二值化中的一项或多项;

[0101] 提取出每个预处理后的所述用户图像中的用户人脸图像。

[0102] 本发明实施例提供了一种基于面部特征的人机交互装置,通过检测人体面部生物特征和其运动姿态,判断用户人脸当前的运动状态,进而可以发出暂停或播放等控制信号,实现通过人脸运动对视频播放控制。该交互方式简答,不需要用户大幅度移动身体,可以避

免传统方法对书写状态的破坏,并且容易使学生保持学习心态的连续性。本申请提供的交互方式对于书法教育的信息化过程提供基础性的促进作用,具有重要的现实价值。同时,利用二维的用户人脸图像即可确定两个用户人脸图像之间的角度变化值、角速度变化值以及角度变化方向等位姿变化信息,从而可以确定用户头部的旋转位移、旋转速度、旋转方向,进而方便后续生成与用户旋转姿态相一致的控制指令。

[0103] 本发明实施例还提供了一种计算机存储介质,所述计算机存储介质存储有计算机可执行指令,其包含用于执行上述的基于面部特征的人机交互方法的程序,该计算机可执行指令可执行上述任意方法实施例中的方法。

[0104] 其中,所述计算机存储介质可以是计算机能够存取的任何可用介质或数据存储设备,包括但不限于磁性存储器(例如软盘、硬盘、磁带、磁光盘(MO)等)、光学存储器(例如CD、DVD、BD、HVD等)、以及半导体存储器(例如ROM、EPROM、EEPROM、非易失性存储器(NAND FLASH)、固态硬盘(SSD))等。

[0105] 图8示出了本发明的另一个实施例的一种电子设备的结构框图。所述电子设备1100可以是具备计算能力的主机服务器、个人计算机PC、或者可携带的便携式计算机或终端等。本发明具体实施例并不对电子设备的具体实现做限定。

[0106] 该电子设备1100包括至少一个处理器(processor)1110、通信接口(Communications Interface)1120、存储器(memory array)1130和总线1140。其中,处理器1110、通信接口1120、以及存储器1130通过总线1140完成相互间的通信。

[0107] 通信接口1120用于与网元通信,其中网元包括例如虚拟机管理中心、共享存储等。

[0108] 处理器1110用于执行程序。处理器1110可能是一个中央处理器CPU,或者是专用集成电路ASIC(Application Specific Integrated Circuit),或者是被配置成实施本发明实施例的一个或多个集成电路。

[0109] 存储器1130用于可执行的指令。存储器1130可能包含高速RAM存储器,也可能还包括非易失性存储器(non-volatile memory),例如至少一个磁盘存储器。存储器1130也可以是存储器阵列。存储器1130还可能被分块,并且所述块可按一定的规则组合成虚拟卷。存储器1130存储的指令可被处理器1110执行,以使处理器1110能够执行上述任意方法实施例中的基于面部特征的人机交互方法。

[0110] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应所述以权利要求的保护范围为准。

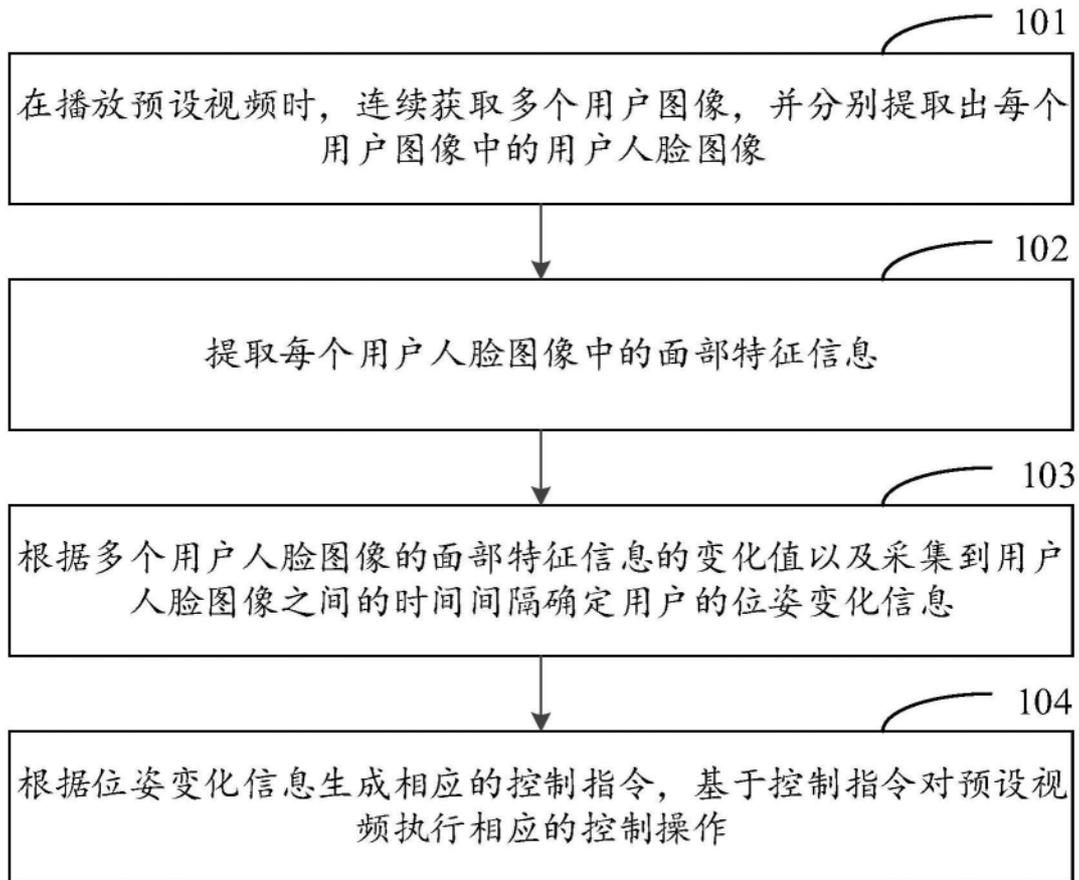


图1

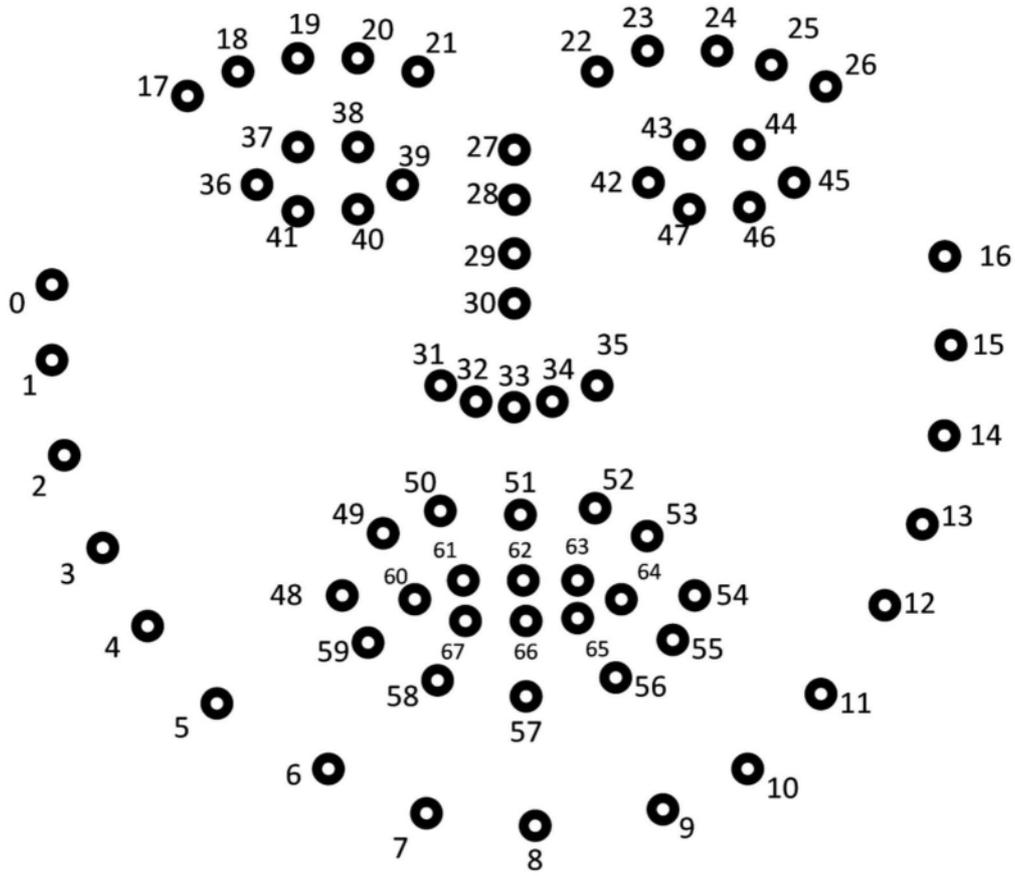


图2

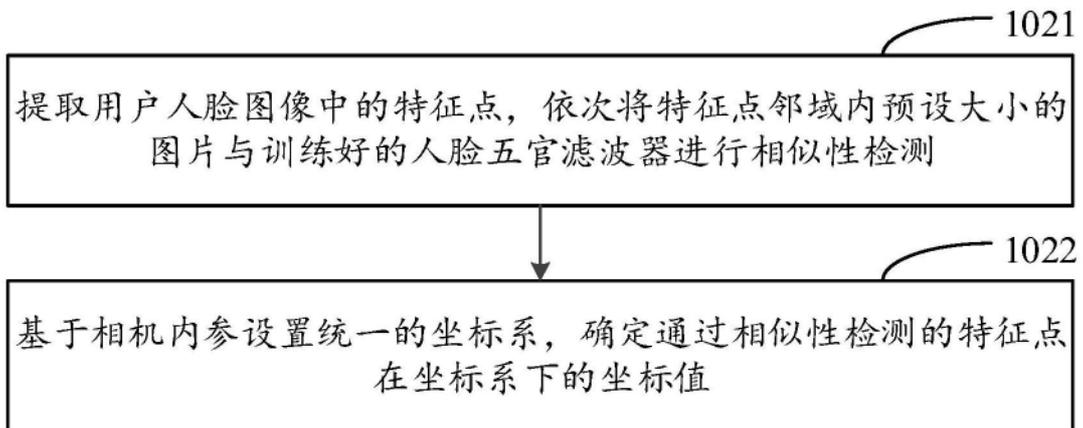


图3

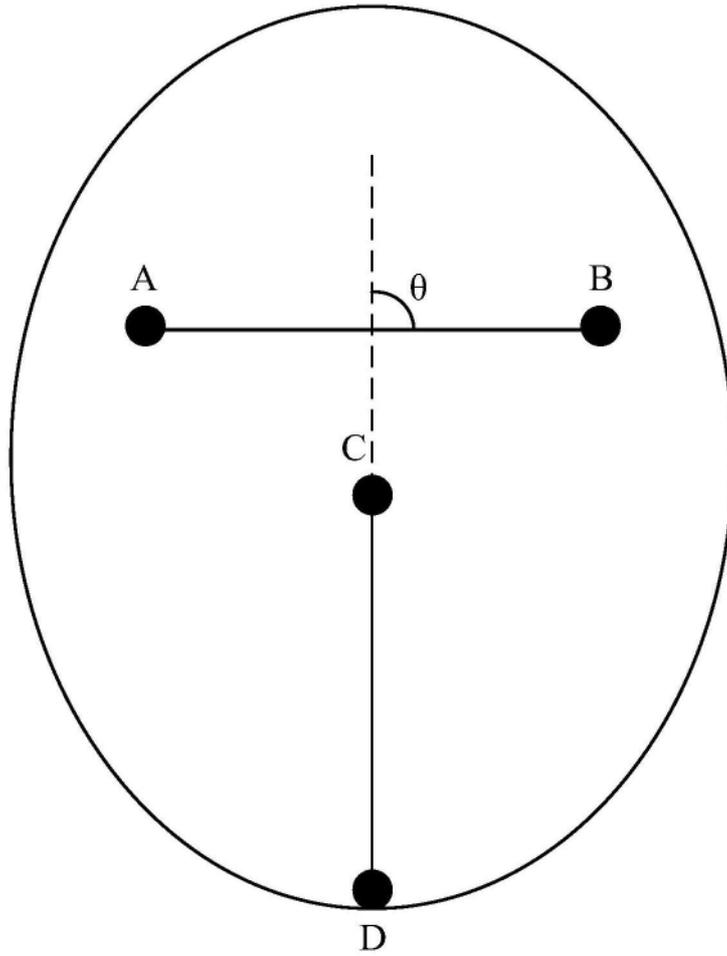


图4

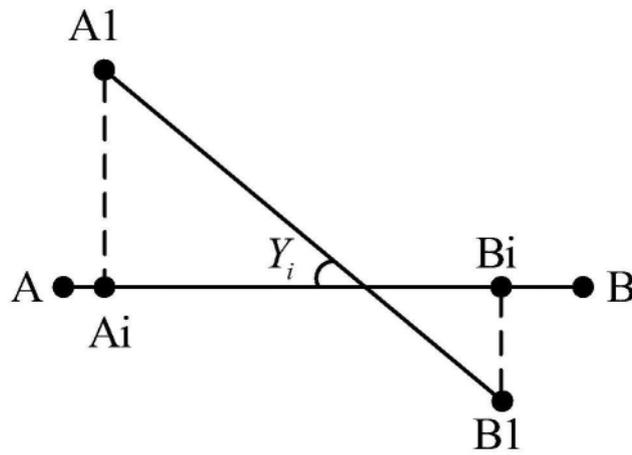


图5

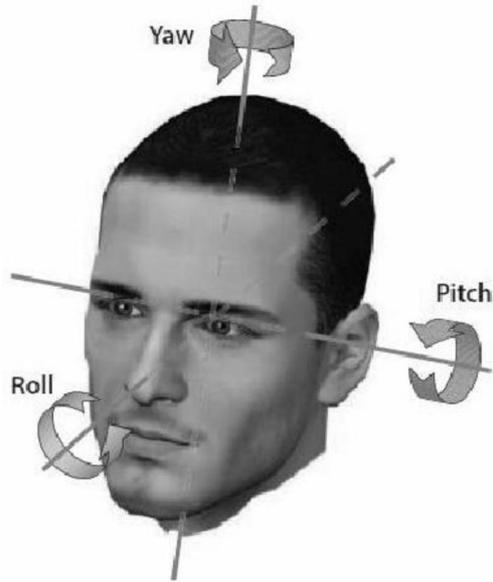


图6

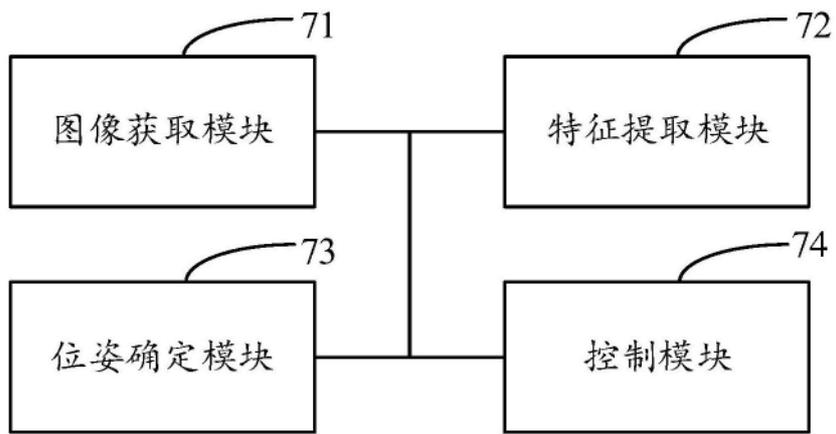


图7

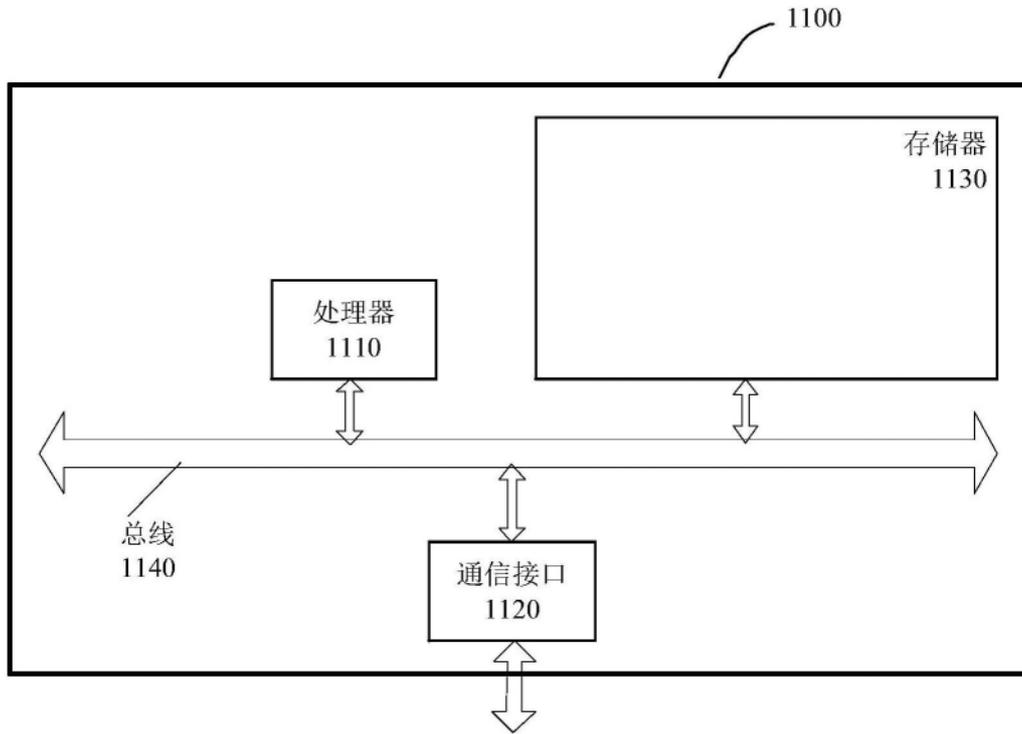


图8