(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 106024003 B
(45)授权公告日 2020.01.31

(21)申请号 201610304047.6

(22)申请日 2016.05.10

(65)同一申请的已公布的文献号
申请公布号 CN 106024003 A

(43)申请公布日 2016.10.12

(73)专利权人 北京地平线信息技术有限公司
地址 100080 北京市海淀区海淀大街3号B座10层10-031

(72)发明人 徐荣强

(74)专利代理机构 北京市正见永申律师事务所 11497
代理人 黄小临 王怀章

(51)Int.Cl.
G10L 21/0216(2013.01)
G06K 9/00(2006.01)
G01S 5/18(2006.01)

(56)对比文件
US 2015/0022636 A1,2015.01.22,
CN 105204628 A,2015.12.30,
CN 102160398 A,2011.08.17,
CN 103716540 A,2014.04.09,
CN 103841357 A,2014.06.04,

审查员 陈宬

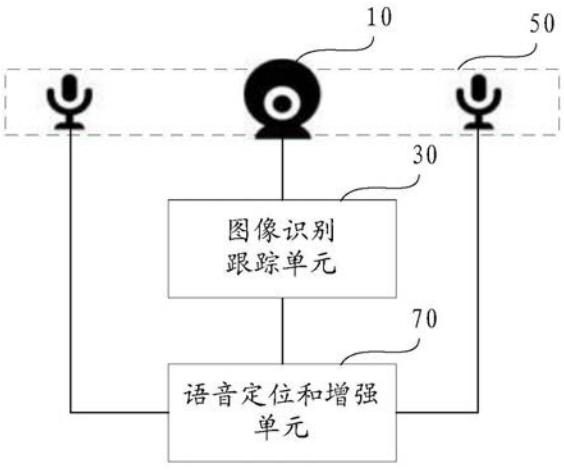
权利要求书2页 说明书6页 附图4页

(54)发明名称

结合图像的语音定位和增强系统及方法

(57)摘要

本发明提供一种结合图像的语音定位和增强系统及方法，所述定位系统包括图像识别跟踪子系统和语音定位和增强子系统。图像识别跟踪子系统包括：摄像头，用于采集图像序列；图像识别跟踪单元，用于识别人员并缓存脸部三维坐标；通过识别人员执行的第一预定义操作唤醒语音定位和增强子系统，并发送脸部三维坐标；跟踪识别所述人员，并发送更新的脸部三维坐标。语音定位和增强子系统包括：麦克风阵列，用于采集语音信息；语音定位和增强单元，用于根据空间滤波算法和接收的脸部三维坐标控制麦克风阵列定向聚焦采集所述人员的语音信息，并根据所采集的语音信息对所述人员进行定位。本发明实现了结合图像的语音跟踪定位，且具备适用于复杂环境的优点。



CN 106024003 B

1.一种结合图像的语音定位和增强系统,其特征在于,所述系统包括图像识别跟踪子系统和语音定位和增强子系统;

所述图像识别跟踪子系统包括:

摄像头,用于采集当前场景的图像序列;

图像识别跟踪单元,用于识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标;通过识别所述人员执行的第一预定义肢体动作唤醒语音定位和增强子系统,并将所缓存的执行所述第一预定义肢体动作的人员的脸部三维坐标发送至所述语音定位和增强子系统;跟踪识别执行所述第一预定义肢体动作的人员,并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统;

所述语音定位和增强子系统包括:

麦克风阵列,用于采集语音信息;

语音定位和增强单元,用于根据接收的执行所述第一预定义肢体动作的人员的脸部三维坐标计算脸部角度信息,根据空间滤波算法和所述脸部角度信息控制所述麦克风阵列定向聚焦采集所述人员的语音信息,并根据所采集的语音信息对执行所述第一预定义肢体动作的人员进行定位和语音增强;

所述图像识别跟踪单元还用于识别执行所述第一预定义肢体动作的人员执行的第二预定义肢体操作;若识别出,则停止跟踪识别执行所述第一预定义肢体动作的人员,进入并维持等待唤醒状态。

2.根据权利要求1所述的系统,其特征在于,所述图像识别跟踪单元和所述语音定位和增强单元还用于根据所述摄像头的位置和所述麦克风阵列的位置统一三维坐标系。

3.根据权利要求1所述的系统,其特征在于,所述语音定位和增强单元还用于利用所述空间滤波算法,根据所述接收的脸部三维坐标进行实时的空域滤波调整。

4.根据权利要求1所述的系统,其特征在于,所述语音增强通过对根据所述脸部角度信息所定位方向的声音信号进行加强、同时对其它方向的声音信号进行抑制实现。

5.根据权利要求1-4任一项所述的系统,其特征在于,所述麦克风阵列包括一组双麦克风阵列。

6.一种结合图像的语音定位和增强方法,其特征在于,所述方法包括:

采集当前场景的图像序列;

识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标;

通过识别所述人员执行的第一预定义肢体动作唤醒语音定位和增强子系统,并将所缓存的执行所述第一预定义肢体动作的人员的脸部三维坐标发送至所述语音定位和增强子系统;

根据接收的执行所述第一预定义肢体动作的人员的脸部三维坐标计算脸部角度信息,根据空间滤波算法和所述脸部角度信息控制麦克风阵列定向聚焦采集所述人员的语音信息,并根据所采集的语音信息对执行所述第一预定义肢体动作的人员进行定位和语音增强;

跟踪识别执行所述第一预定义肢体动作的人员,并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统;返回上一步进行循环,直至无法跟踪识别执行所述第一预定义肢体动作的人员;

其中，跟踪识别执行所述第一预定义肢体动作的人员，并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统；返回上一步进行循环，直至无法跟踪识别执行所述第一预定义肢体动作的人员，包括：

识别执行所述第一预定义肢体动作的人员执行的第二预定义操作：

若识别出执行所述第一预定义肢体动作的人员执行第二预定义操作，则停止跟踪识别执行所述第一预定义肢体动作的人员，进入并维持等待唤醒状态；

若未识别出执行所述第一预定义肢体动作的人员执行第二预定义操作，则将更新的执行所述第一预定义肢体动作的人员的脸部三维坐标发送至所述语音定位和增强子系统；返回上一步进行循环，直至无法跟踪识别执行所述第一预定义肢体动作的人员。

7.根据权利要求6所述的方法，其特征在于，所述跟踪识别执行所述第一预定义肢体动作的人员，并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统；返回上一步进行循环，直至无法跟踪识别执行所述第一预定义肢体动作的人员包括：

跟踪识别所述人员，若无法跟踪识别，则停止跟踪识别所述人员，进入并维持等待唤醒状态。

8.根据权利要求6所述的方法，其特征在于，所述采集当前场景的图像序列之前还包括：根据摄像头的位置和麦克风阵列的位置统一三维坐标系。

9.根据权利要求6所述的方法，其特征在于，所述根据空间滤波算法和接收的脸部三维坐标控制麦克风阵列定向聚焦采集所述人员的语音信息还包括利用所述空间滤波算法，根据所述接收的脸部三维坐标进行实时的空域滤波调整。

10.根据权利要求6所述的方法，其特征在于，所述语音增强通过对根据所述脸部角度信息所定位方向的声音信号进行加强、同时对其它方向的声音信号进行抑制实现。

11.根据权利要求6-10任一项所述的方法，其特征在于，所述麦克风阵列包括一组双麦克风阵列。

结合图像的语音定位和增强系统及方法

技术领域

[0001] 本申请涉及语音定位技术领域，具体涉及一种结合图像的语音定位和增强系统及方法。

背景技术

[0002] 现有的语音定位系统和方法都是基于麦克风阵列来完成定位，无法实现实时跟踪，只能通过语音唤醒定位系统重新进行麦克风阵列的定位，无法实时跟踪监控，用户体验效果较差。

[0003] 同时，现有的语音定位系统和方法因自身的限制对适用环境的要求较高：一方面，抗干扰能力较差，例如抗回声干扰的能力较差，又例如集成在电视、音响等设备中的语音定位系统，因设备本身发音，自身发声内容同样会对定位干扰；另一方面，复杂环境的适应能力较差，噪声环境会降低定位精度，非稳态噪声的干扰，例如同时有多人说话，房间混响也会对定位精度造成影响，例如周围硬反射介质的高混响环境，如玻璃等。

[0004] 此外，现有的语音定位系统和方法还受到麦克风阵列的限制，例如双麦克风阵列只能满足180°的平面定位，四阵列麦克风只能满足360°的平面定位，通常需要通过复杂阵型的麦克风阵列实现空间定位，而难以通过较简单设备实现立体的空间定位。

发明内容

[0005] 鉴于现有技术中的上述缺陷或不足，期望提供一种能实现跟踪的语音定位且适用于复杂环境的结合图像的语音定位和增强系统及方法。

[0006] 第一方面，本发明提供一种结合图像的语音定位和增强系统，所述系统包括图像识别跟踪子系统和语音定位和增强子系统。

[0007] 所述图像识别跟踪子系统包括：

[0008] 摄像头，用于采集当前场景的图像序列；

[0009] 图像识别跟踪单元，用于识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标；通过识别所述人员执行的第一预定义操作唤醒语音定位和增强子系统，并将所缓存的所述人员的脸部三维坐标发送至所述语音定位和增强子系统；跟踪识别所述人员，并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统。

[0010] 所述语音定位和增强子系统包括：

[0011] 麦克风阵列，用于采集语音信息；

[0012] 语音定位和增强单元，用于根据接收的脸部三维坐标计算脸部角度信息，根据空间滤波算法和所述脸部角度信息控制所述麦克风阵列定向聚焦采集所述人员的语音信息，并根据所采集的语音信息对所述人员进行定位和语音增强。

[0013] 第二方面，本发明提供一种结合图像的语音定位和增强方法，所述方法包括：

[0014] 采集当前场景的图像序列；

[0015] 识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标；

[0016] 唤醒语音定位和增强子系统，并将所述脸部三维坐标发送至所述语音定位和增强子系统；

[0017] 根据接收的脸部三维坐标计算脸部角度信息，根据空间滤波算法和所述脸部角度信息控制麦克风阵列定向聚焦采集所述人员的语音信息，并根据所采集的语音信息对所述人员进行定位和语音增强；

[0018] 跟踪识别所述人员，并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统；返回上一步进行循环，直至无法跟踪识别所述人员。

[0019] 本发明诸多实施例提供的结合图像的语音定位和增强系统及方法通过摄像头识别人员并缓存脸部三维坐标，通过自定义的动作开启跟踪识别，为语音定位和增强子系统提供了实时的脸部三维坐标，语音定位和增强单元根据空间滤波算法和实时的脸部三维坐标控制所述麦克风阵列定向聚焦采集所跟踪人员的语音信息，最终实现了结合图像的语音跟踪定位和语音增强，同时实现了避免其它音源的干扰，具备了适用于复杂环境的优点；

[0020] 本发明一些实施例提供的结合图像的语音定位和增强系统及方法进一步通过识别自定义的动作关闭跟踪定位和语音增强，实现了智能控制跟踪定位和语音增强的开关；

[0021] 本发明一些实施例提供的结合图像的语音定位和增强系统及方法进一步通过根据摄像头的位置和麦克风阵列的位置统一三维坐标系，使语音定位和增强单元无需对接收的脸部三维坐标进行换算，减少了计算的工作量，降低了设备的硬件要求；

[0022] 本发明一些实施例提供的结合图像的语音定位和增强系统及方法进一步通过利用所述空间滤波算法根据实时脸部三维坐标进行实时的空域滤波调整，优化了语音信息的采集效果，从而优化了最终跟踪定位的效果；

[0023] 本发明一些实施例提供的结合图像的语音定位和增强系统及方法进一步通过采用一组双麦克风阵列，即实现了通过双麦克风阵列和摄像头实现立体的空间定位。

附图说明

[0024] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本申请的其它特征、目的和优点将会变得更明显：

[0025] 图1为本发明一实施例中结合图像的语音定位和增强系统的结构示意图。

[0026] 图2为本发明一实施例中结合图像的语音定位和增强方法的流程图。

[0027] 图3为本发明一优选实施例中步骤S60的流程图。

[0028] 图4为本发明一优选实施例中结合图像的语音定位和增强方法的流程图。

具体实施方式

[0029] 下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与发明相关的部分。

[0030] 需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

[0031] 图1为本发明一实施例中结合图像的语音定位和增强系统的结构示意图。

[0032] 如图1所示，在本实施例中，本发明提供的定位系统包括图像识别跟踪子系统和语

音定位和增强子系统。

[0033] 所述图像识别跟踪子系统包括摄像头10和图像识别跟踪单元30。摄像头10用于采集当前场景的图像序列。图像识别跟踪单元30用于识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标;通过识别所述人员执行的第一预定义操作唤醒语音定位和增强子系统,并将所缓存的所述人员的脸部三维坐标发送至所述语音定位和增强子系统;跟踪识别所述人员,并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统。

[0034] 所述语音定位和增强子系统包括麦克风阵列50和语音定位和增强单元70。麦克风阵列50用于采集语音信息。语音定位和增强单元70用于根据接收的脸部三维坐标计算脸部角度信息,根据空间滤波算法和所述脸部角度信息控制麦克风阵列50定向聚焦采集所述人员的语音信息,并根据所采集的语音信息对所述人员进行定位和语音增强。

[0035] 图2为本发明一实施例中结合图像的语音定位和增强方法的流程图。图2所示的定位方法可应用在图1所示的定位系统中。

[0036] 如图2所示,在本实施例中,本发明提供的结合图像的语音定位和增强方法具体包括:

[0037] S20:采集当前场景的图像序列。

[0038] S30:识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标。

[0039] S40:通过识别所述人员执行的第一预定义操作唤醒语音定位和增强子系统,并将所缓存的所述人员的脸部三维坐标发送至所述语音定位和增强子系统。

[0040] S50:根据接收的脸部三维坐标计算脸部角度信息,根据空间滤波算法和所述脸部角度信息控制麦克风阵列定向聚焦采集所述人员的语音信息,并根据所采集的语音信息对所述人员进行定位和语音增强。

[0041] S60:跟踪识别所述人员,并将更新的所述脸部三维坐标发送至所述语音定位和增强子系统;返回步骤S50进行循环,直至无法跟踪识别所述人员。

[0042] 例如在设置了上述定位系统的一间房间内,同时有甲、乙、丙、丁四个人,图像识别跟踪单元30识别四人的脸部并分别缓存各人的脸部三维坐标。图像识别跟踪单元30中预设了可以通过招手3秒开启跟踪识别。在本实施例中,所述第一预定义操作设置为招手3秒,在更多实施例中,还可以根据实际需求将所述第一预定义操作设置为各种不同的手势、各类肢体动作等不同的操作,只要可以通过摄像头10采集并通过图像识别跟踪单元30识别,即可实现相同的技术效果。

[0043] 当甲招手3秒之后,摄像头10采集到这一序列图像,图像识别跟踪单元30识别出甲执行了招手3秒的操作,随即唤醒了语音定位和增强子系统,并将所缓存的甲的脸部三维坐标发送至所述语音定位和增强子系统,同时对甲进行快速注册,开始进行跟踪识别,同时对于未进行注册的乙丙丁三人不进行跟踪识别。

[0044] 当甲未移动时,语音定位和增强单元70根据空间滤波算法和之前接收的脸部三维坐标持续控制麦克风阵列50定向聚焦采集甲的语音信息,同时进行语音增强;

[0045] 当甲移动时,摄像头10采集到相关的图像序列,图像识别跟踪单元30识别出甲进行了移动,缓存新的脸部三维坐标并发送至语音定位和增强单元70,语音定位和增强单元70根据实时接收的脸部三维坐标计算实时的脸部角度信息,根据空间滤波算法和实时的脸部角度信息控制麦克风阵列50跟踪甲,定向聚焦采集语音信息,同时对甲进行语音增强;

[0046] 当甲移动出了摄像头10的采集范围之后，图像识别跟踪单元30无法跟踪识别甲，停止跟踪识别，进入并维持等待唤醒状态，同时发送提示信息至语音定位和增强子系统，提示停止进行定位。

[0047] 在一些实施例中，所述图像识别跟踪子系统和语音定位和增强子系统设置为一体集成的装置；在另一些实施例中，所述图像识别跟踪子系统和语音定位和增强子系统可根据实际需求设置为通过通用接口连接的装置，例如采用USB接口的摄像头、标准通用接口的麦克风等。

[0048] 上述实施例提供的系统和方法通过摄像头识别人员并缓存脸部三维坐标，通过自定义的动作开启跟踪识别，为语音定位和增强子系统提供了实时的脸部三维坐标，语音定位和增强单元根据空间滤波算法和实时的脸部三维坐标控制所述麦克风阵列定向聚焦采集所跟踪人员的语音信息，最终实现了结合图像的语音跟踪定位和语音增强，同时实现了避免其它音源的干扰，具备了适用于复杂环境的优点。

[0049] 在一优选实施例中，图像识别跟踪单元50还用于识别所述人员执行的第二预定义操作：若识别出，则停止跟踪识别所述人员，进入并维持等待唤醒状态。

[0050] 图3为本发明一优选实施例中步骤S60的流程图。该定位方法可应用在上述实施例提供的定位系统中。

[0051] 如图3所示，在一优选实施例中，步骤S60具体包括：

[0052] S61：跟踪识别人员，若无法跟踪识别，则停止跟踪识别所述人员，进入并维持等待唤醒状态；

[0053] S63：识别所述人员执行的第二预定义操作：

[0054] 若识别出所述人员执行第二预定义操作，则停止跟踪识别所述人员，进入并维持等待唤醒状态；

[0055] S65：若未识别出所述人员执行第二预定义操作，则将更新的人员的脸部三维坐标发送至所述语音定位和增强子系统；返回步骤S50。

[0056] 具体地，同样以上述甲、乙、丙、丁四个人共处一室的场景为例，图像识别跟踪单元30中还预设了可以通过握拳3秒关闭跟踪识别。在本实施例中，所述第二预定义操作设置为握拳3秒，在更多实施例中，还可以根据实际需求将所述第二预定义操作设置为各种不同的手势、各类肢体动作等不同的操作，只要可以通过摄像头10采集并通过图像识别跟踪单元30识别，即可实现相同的技术效果。

[0057] 当甲握拳3秒之后，摄像头10采集到这一序列图像，图像识别跟踪单元30识别出甲执行了握拳3秒的操作，停止对甲的跟踪识别，同时向所述语音定位和增强子系统发送提示信息，所述语音定位和增强子系统接收到提示信息后，停止对甲进行定位，同时取消对甲的语音增强。此时所述系统可响应乙、丙或丁通过手势开启跟踪识别。

[0058] 上述实施例提供的结合图像的语音定位和增强系统及方法进一步通过识别自定义的动作关闭跟踪定位和语音增强，实现了智能控制跟踪定位和语音增强的开关。

[0059] 在一优选实施例中，图像识别跟踪单元30和语音定位和增强单元70还用于根据摄像头10的位置和麦克风阵列50的位置统一三维坐标系。

[0060] 图4为本发明一优选实施例中结合图像的语音定位和增强方法的流程图。该定位方法可应用在上述实施例提供的定位系统中。

[0061] 如图4所示，在一优选实施例中，步骤S20之前还包括：

[0062] S10：根据摄像头的位置和麦克风阵列的位置统一三维坐标系。

[0063] 上述实施例提供的系统和方法进一步通过根据摄像头的位置和麦克风阵列的位置统一三维坐标系，使语音定位和增强单元无需对接收的脸部三维坐标进行坐标换算，减少了计算的工作量，降低了设备的硬件要求。

[0064] 在一优选实施例中，语音定位和增强单元70还用于利用所述空间滤波算法，根据所述接收的脸部三维坐标进行实时的空域滤波调整。

[0065] 在对应的方法实施例中，步骤S50中所述的根据空间滤波算法和接收的脸部三维坐标控制麦克风阵列定向聚焦采集所述人员的语音信息还包括利用所述空间滤波算法，根据所述接收的脸部三维坐标进行实时的空域滤波调整。

[0066] 上述实施例提供的系统和方法进一步通过利用所述空间滤波算法根据实时脸部三维坐标进行实时的空域滤波调整，优化了语音信息的采集效果，从而优化了最终跟踪定位的效果。

[0067] 在一优选实施例中，所述语音增强通过对根据所述脸部角度信息所定位方向的声音信号进行加强、同时对其它方向的声音信号进行抑制实现。

[0068] 在一优选实施例中，麦克风阵列50包括一组双麦克风阵列。具体地，在更多实施例中，麦克风阵列50可以包括多对麦克风阵列以实现多线程跟踪定位和语音加强，也可以采用其它不同组成结构的麦克风阵列，只要能实现语音采集和定向语音加强，即可实现同样的技术效果。

[0069] 上述实施例提供的系统和方法进一步通过采用一组双麦克风阵列，即实现了通过双麦克风阵列和摄像头实现立体的空间定位。

[0070] 附图中的流程图和框图，图示了按照本发明各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，所述模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这根据所涉及的功能而定。也要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以通过执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以通过专用硬件与计算机指令的组合来实现。

[0071] 描述于本申请实施例中所涉及到的单元或模块可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元或模块也可以设置在处理器中，例如，图像识别跟踪单元30和语音定位和增强单元70可以是设置在计算机或移动智能设备中的软件程序，通过有线或无线的方式与摄像头10和麦克风阵列50连接；也可以是单独进行图像跟踪识别或语音定位的硬件芯片。其中，这些单元或模块的名称在某种情况下并不构成对该单元或模块本身的限定，例如，图像识别跟踪单元30还可以被描述为“用于定位摄像头所跟踪人员的定位单元”。

[0072] 作为另一方面，本申请还提供了一种计算机可读存储介质，该计算机可读存储介质可以是上述实施例中所述装置中所包含的计算机可读存储介质；也可以是单独存在，未

装配入设备中的计算机可读存储介质。计算机可读存储介质存储有一个或者一个以上程序，所述程序被一个或者一个以上的处理器用来执行描述于本申请的公式输入方法。

[0073] 以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离所述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

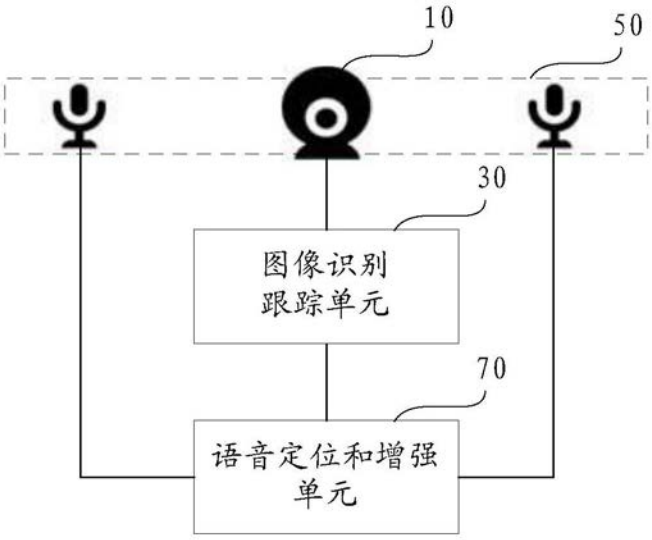


图1

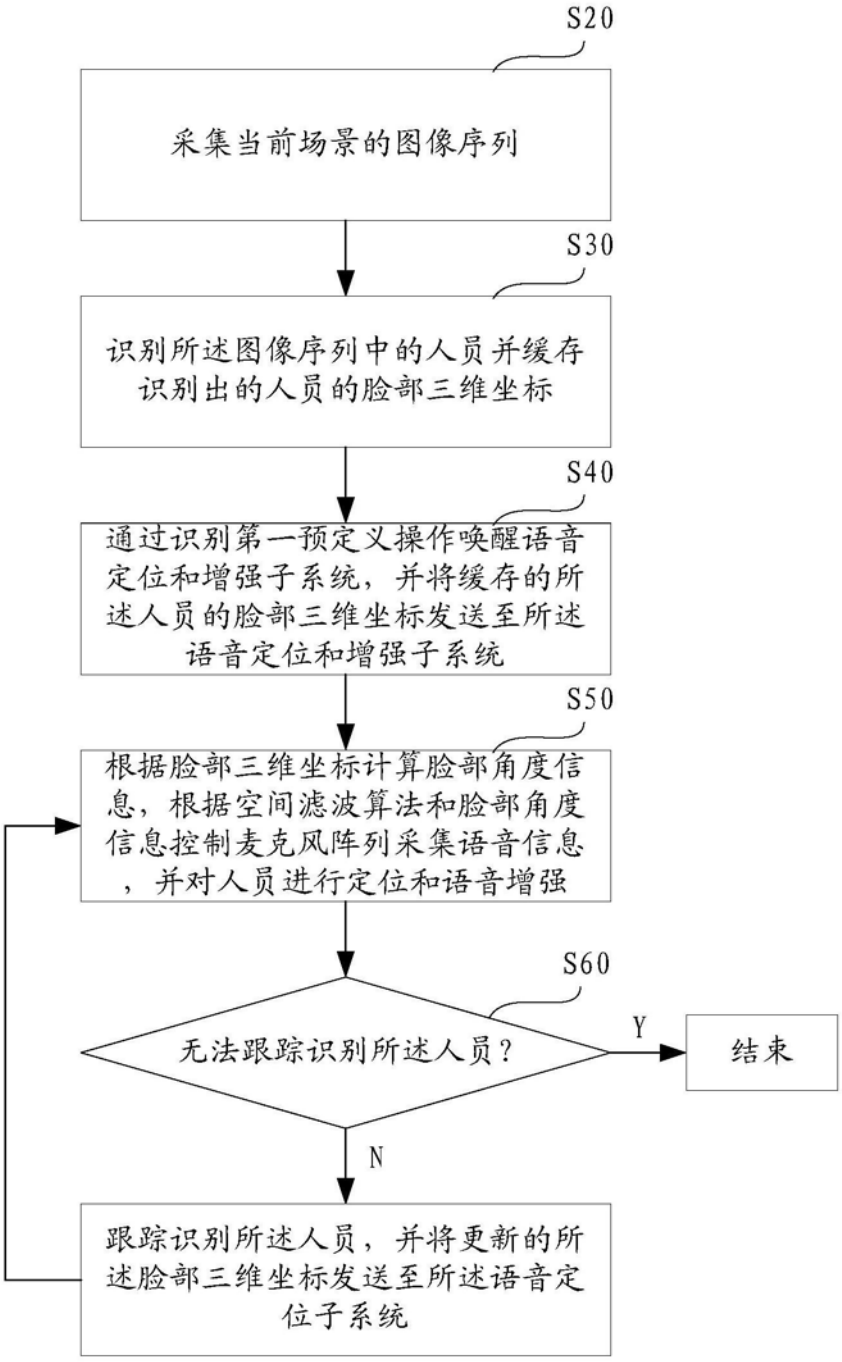

S20
采集当前场景的图像序列
S30
识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标
S40
通过识别第一预定义操作唤醒语音定位和增强子系统，并将缓存的所述人员的脸部三维坐标发送至所述语音定位和增强子系统
S50
根据脸部三维坐标计算脸部角度信息，根据空间滤波算法和脸部角度信息控制麦克风阵列采集语音信息，并对人员进行定位和语音增强
S60
无法跟踪识别所述人员？
Y
结束
N
跟踪识别所述人员，并将更新的所述脸部三维坐标发送至所述语音定位子系统

图2

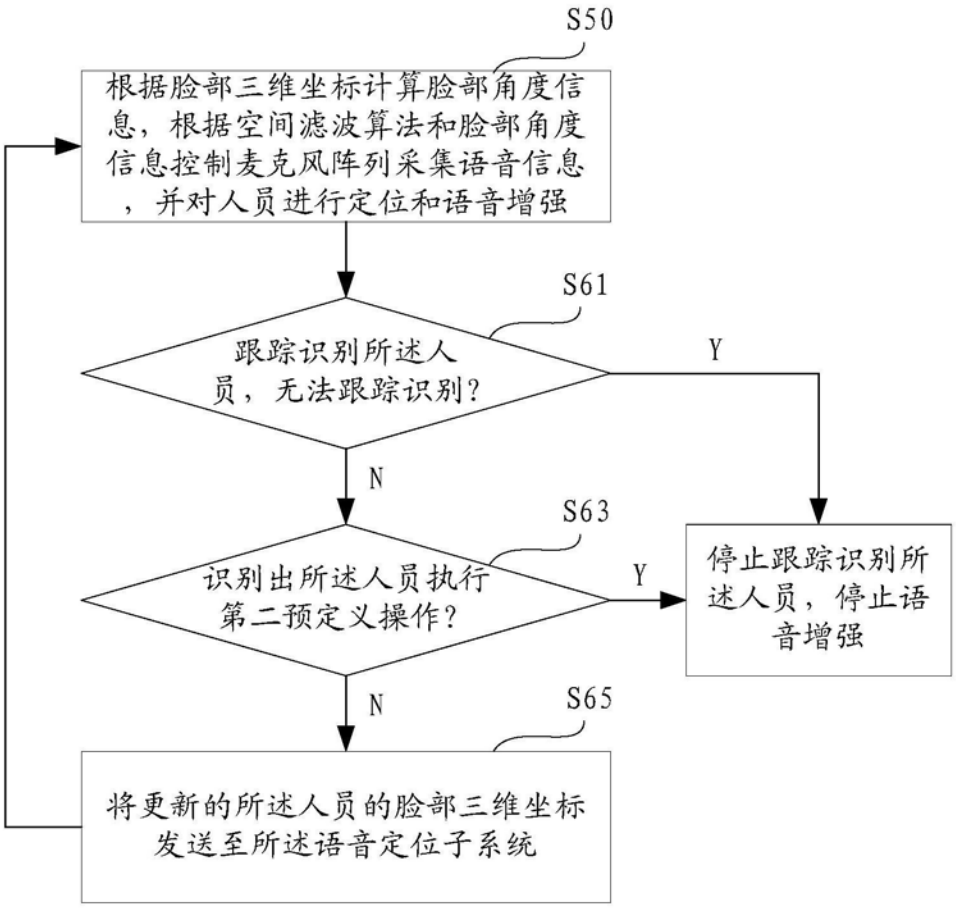

S50
根据脸部三维坐标计算脸部角度信息，根据空间滤波算法和脸部角度信息控制麦克风阵列采集语音信息，并对人员进行定位和语音增强
S61
跟踪识别所述人员，无法跟踪识别？
Y
N
S63
识别出所述人员执行第二预定义操作？
Y
停止跟踪识别所述人员，停止语音增强
N
S65
将更新的所述人员的脸部三维坐标发送至所述语音定位子系统

图3

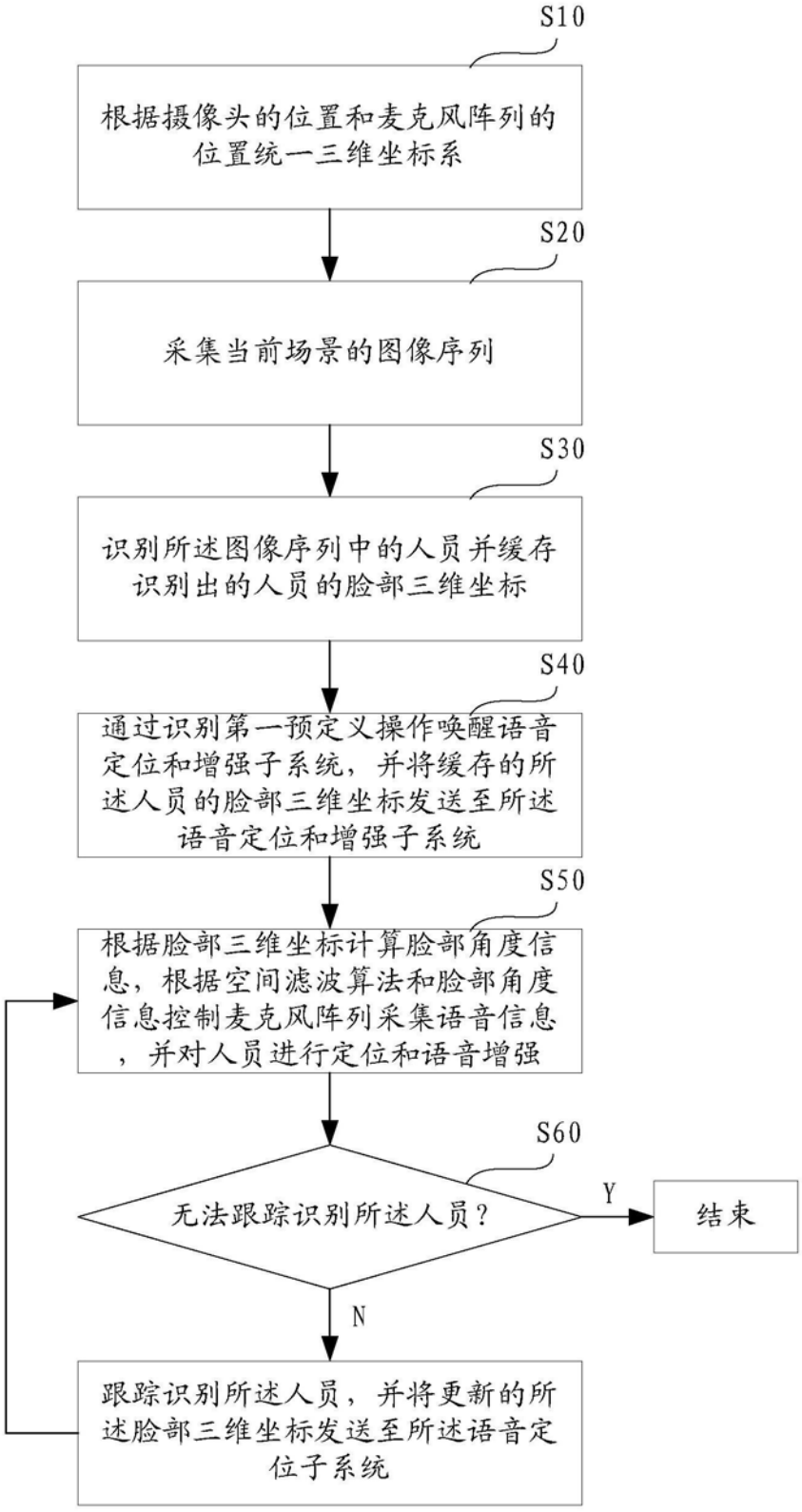

S10
根据摄像头的位置和麦克风阵列的位置统一三维坐标系
S20
采集当前场景的图像序列
S30
识别所述图像序列中的人员并缓存识别出的人员的脸部三维坐标
S40
通过识别第一预定义操作唤醒语音定位和增强子系统，并将缓存的所述人员的脸部三维坐标发送至所述语音定位和增强子系统
S50
根据脸部三维坐标计算脸部角度信息，根据空间滤波算法和脸部角度信息控制麦克风阵列采集语音信息，并对人员进行定位和语音增强
S60
无法跟踪识别所述人员？
Y
结束
N
跟踪识别所述人员，并将更新的所述脸部三维坐标发送至所述语音定位子系统

图4