

(12) 发明专利申请

(10) 申请公布号 CN 102945672 A

(43) 申请公布日 2013. 02. 27

(21) 申请号 201210374809. 1

(22) 申请日 2012. 09. 29

(71) 申请人 深圳市国华识别科技开发有限公司
地址 518026 广东省深圳市福田区彩田路西
红荔路南中银花园办公(综合)楼A栋
4B、Ca、cb、D4cb-20

(72) 发明人 王洪智 刘乐元 桑农 刘国华

(51) Int. Cl.

G10L 15/26(2006. 01)

G06F 3/01(2006. 01)

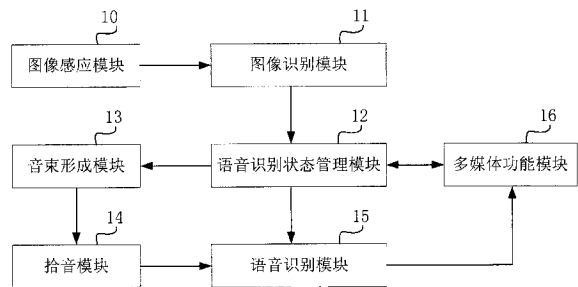
权利要求书 2 页 说明书 7 页 附图 5 页

(54) 发明名称

一种多媒体设备语音控制系统及方法

(57) 摘要

本发明提出一种多媒体设备语音控制系统,包括:图像感应模块,采集用户动作图像;图像识别模块,根据用户动作图像确定控制指令类型或状态;语音识别状态管理模块,根据当前的控制指令类型激活或暂停语音识别;拾音模块,采集语音数据;语音识别模块,对采集到的语音数据进行识别,形成控制指令;多媒体功能模块,执行控制指令,向用户提供相应的多媒体功能。本发明还提出一种多媒体设备语音控制方法。本发明结合图像识别和语音识别技术,实现不依赖手持遥控器,不使用近距离拾音模块实现自由便捷的语音控制,有效避免了多媒体设备输出的声音、环境背景声音以及用户的非控制指令语音信号对控制指令语音识别的干扰,实现准确识别用户发出的控制指令。



1. 一种多媒体设备语音控制系统,其特征在于,包括:
图像感应模块,采集用户动作图像;
图像识别模块,根据用户动作图像确定控制指令类型或状态;
语音识别状态管理模块,根据当前的控制指令类型激活或暂停语音识别;
拾音模块,采集语音数据;
语音识别模块,对采集到的语音数据进行识别,形成控制指令;
多媒体功能模块,执行控制指令,向用户提供相应的多媒体功能。
2. 如权利要求 1 所述的一种多媒体设备语音控制系统,其特征在于:
所述图像识别模块将用户动作图像与预设的图像模版比对,选取与用户动作图像匹配的控制指令类型;
若比对结果为找到与用户动作图像匹配的控制指令类型,则认为所述用户所在位置为目标音源位置,向语音识别状态管理模块发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型 ;或
若未找到与用户动作图像匹配的控制指令类型,则向语音识别状态管理模块发出比对失败信息。
3. 如权利要求 2 所述的一种多媒体设备语音控制系统,其特征在于还包括:
音束形成模块,根据目标音源所在位置信息确定拾音方向和拾音接收角。
4. 如权利要求 3 所述的一种多媒体设备语音控制系统,其特征在于:
所述拾音模块为阵列拾音模块,包括规则排列的至少一拾音传感器,根据拾音方向和拾音接收角的限定采集目标音源发出的语音信号,进行数字化处理形成语音数据后发送给语音识别模块。
5. 如权利要求 4 所述的一种多媒体设备语音控制系统,其特征在于:
所述语音识别状态管理模块根据接收到的启动语音识别信息,向语音识别模块发送启动指令及控制指令类型以激活语音识别,向音束形成模块发送目标音源所在位置信息,并控制多媒体功能模块减小多媒体输出声音的音量 ;或根据接收到的比对失败信息,向语音识别模块发送指令暂停语音识别。
6. 如权利要求 5 所述的一种多媒体设备语音控制系统,其特征在于:
语音识别模块根据来自语音识别状态管理模块的启动指令和控制指令类型,对来自拾音模块的语音数据进行识别,形成属于所述控制指令类型的控制指令,发送给多媒体功能模块。
7. 如权利要求 1 至 5 任意一项所述的一种多媒体设备语音控制系统,其特征在于:
所述语音识别模块包括本地语音识别模块和云端语音识别模块;
本地语音识别模块识别语音数据,形成属于所述控制指令类型的控制指令,发送给多媒体功能模块;
云端语音识别模块对本地语音识别模块无法识别的语音数据进行语义识别处理,形成属于所述控制指令类型的控制指令,发送给多媒体功能模块。
8. 一种多媒体设备语音控制方法,包括:
图像感应模块采集用户动作图像的步骤;
图像识别模块根据用户动作图像确定控制指令类型或状态的步骤;

语音识别状态管理模块根据当前的控制指令类型激活或暂停语音识别的步骤；

音束形成模块确定拾音方向和拾音接收角的步骤；

阵列拾音模块根据拾音方向和拾音接收角的限定采集用户发出的语音信号,进行数字化处理形成语音数据的步骤；

语音识别模块对采集到的语音数据进行识别,形成控制指令的步骤；

多媒体功能模块执行控制指令,向用户提供相应的多媒体功能的步骤。

9. 如权利要求 8 所述的一种多媒体设备语音控制方法,其特征在于：

图像感应模块采集用户动作图像；

所述图像识别模块将用户动作图像与预设的图像模版比对,选取与用户动作图像匹配的控制指令类型；若比对结果为找到与用户动作图像匹配的控制指令类型,则认为所述用户所在位置为目标音源位置,向语音识别状态管理模块发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型；若未找到与用户动作图像匹配的控制指令类型,则向语音识别状态管理模块发出比对失败信息；

所述语音识别状态管理模块根据接收到的启动语音识别信息,向语音识别模块发送启动指令及控制指令类型以激活语音识别,向音束形成模块发送目标音源所在位置信息,并控制多媒体功能模块减小多媒体输出声音的音量；或根据接收到的比对失败信息,向语音识别模块发送指令暂停语音识别；

音束形成模块根据目标音源所在位置信息确定拾音方向和拾音接收角；

阵列拾音模块根据拾音方向和拾音接收角的限定采集目标音源发出的语音信号,进行数字化处理形成语音数据后发送给语音识别模块；

语音识别模块根据来自语音识别状态管理模块的启动指令和控制指令类型,对来自阵列拾音模块的语音数据进行识别,形成属于所述控制指令类型的控制指令,发送给多媒体功能模块；

多媒体功能模块执行控制指令,向用户提供相应的多媒体功能。

10. 如权利要求 8 或 9 所述的一种多媒体设备语音控制方法,其中语音识别模块包括本地语音识别模块和云端语音识别模块,语音识别模块预设语音指令词典,其特征在于,包括：

本地语音识别模块识别语音数据,将语音数据与所述语音指令词典中的单词模型对比,若语音数据与至少一单词模型相似度大于预设阈值,则将所述语音数据解释为与所述单词模型对应的控制指令,发送给多媒体功能模块；

若语音数据与至少一单词模型相似度不大于预设阈值,则将语音数据通过网络发送给所述云端语音识别模块；

所述云端语音识别模块对语音数据进行语义识别处理,形成控制指令,通过网络发送给多媒体功能模块。

一种多媒体设备语音控制系统及方法

技术领域

[0001] 本发明涉及语音遥控技术,更具体地说,涉及一种多媒体设备语音控制系统及方法。

背景技术

[0002] 继手机智能化后,电视、投影仪、游戏机等多种多媒体设备的智能化成为必然趋势。现有多种多媒体设备电视一般配备高性能控制芯片,具有开放式平台和操作系统;可由用户自行安装和卸载应用程序,此类应用程序扩展多媒体设备的功能;支持用户通过网络进行信息浏览和网络社交。以智能电视机为例,电视机不再局限于传统的电视节目播放功能,还通过运行应用程序实现音视频分享,交互娱乐游戏等多种功能。而传统的按键式遥控器已经无法满足多种多媒体功能选择和操作需求。

[0003] 现有技术提出包括触摸控制、声音控制、手势识别、体感控制等多种人机交互方案以实现智能控制,但由于电视机的使用场景限制和使用习惯问题,仍然没有一种智能控制方式能够完全取代手持遥控器,用户必须借助手持遥控器上的特定功能键和数字键组合来进行操控。例如触摸控制方案需要在借助遥控器上安装的触摸感应模块;手势识别方案其无法快捷地进行常用的节目频道跳转控制;如用户想从当前的 1 频道切换到 55 频道,仅采用手势识别方式操作显然不如使用传统遥控器更加快捷;而体感控制方案的问题类似手势识别方案,通常还需要安装价格昂贵的深度图像感应模块才能实现准确的体感控制功能。而现有技术声音识别控制方案的问题在于,一般为清晰地采集用户的语音而在遥控器上安装麦克风模块,仍然需要借助手持遥控器。

[0004] 随着声音识别技术的发展,声音识别、语义识别等都已经基本达到实用的阶段,而且随着云计算技术的普及,许多基于云服务的语音识别服务商与智能电视结合实现语音控制电视。但目前现有技术方案大多在遥控器上加装麦克风拾音模块,获取用户语音经处理后再传至云端进行识别处理;即使是采用可以远距离拾音的麦克风阵列技术,也存在因电视输出的声音、环境声音干扰和用户的非控制指令语音被错误解读为控制指令等问题,影响语音控制电视机等多媒体设备的效果。

发明内容

[0005] 本发明要解决的技术问题在于,提出一种多媒体设备语音控制系统。

[0006] 本发明解决其技术问题所采用的技术方案是提供一种多媒体设备语音控制系统,包括:图像感应模块,采集用户动作图像;图像识别模块,根据用户动作图像确定控制指令类型或状态;语音识别状态管理模块,根据当前的控制指令类型激活或暂停语音识别;拾音模块,采集语音数据;语音识别模块,对采集到的语音数据进行识别,形成控制指令;多媒体功能模块,执行控制指令,向用户提供相应的多媒体功能。

[0007] 优选地,上述图像识别模块将用户动作图像与预设的图像模版比对,选取与用户动作图像匹配的控制指令类型;若比对结果为找到与用户动作图像匹配的控制指令类型,

则认为用户所在位置为目标音源位置,向语音识别状态管理模块发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型 ;或若未找到与用户动作图像匹配的控制指令类型,则向语音识别状态管理模块发出比对失败信息。

[0008] 优选地,上述多媒体设备语音控制系统还包括音束形成模块,根据目标音源所在位置信息确定拾音方向和拾音接收角。

[0009] 优选地,上述拾音模块为阵列拾音模块,包括规则排列的至少一拾音传感器,根据拾音方向和拾音接收角的限定采集目标音源发出的语音信号,进行数字化处理形成语音数据后发送给语音识别模块。

[0010] 优选地,上述语音识别状态管理模块根据接收到的启动语音识别信息,向语音识别模块发送启动指令及控制指令类型以激活语音识别,向音束形成模块发送目标音源所在位置信息,并控制多媒体功能模块减小多媒体输出声音的音量 ;或根据接收到的比对失败信息,向语音识别模块发送指令暂停语音识别。

[0011] 优选地,上述语音识别模块根据来自语音识别状态管理模块的启动指令和控制指令类型,对来自拾音模块的语音数据进行识别,形成属于控制指令类型的控制指令,发送给多媒体功能模块。

[0012] 优选地,上述语音识别模块包括本地语音识别模块和云端语音识别模块 ;本地语音识别模块识别语音数据,形成属于控制指令类型的控制指令,发送给多媒体功能模块 ;云端语音识别模块对本地语音识别模块无法识别的语音数据进行语义识别处理,形成属于控制指令类型的控制指令,发送给多媒体功能模块。

[0013] 本发明还提出一种多媒体设备语音控制方法,包括 :图像感应模块采集用户动作图像的步骤 ;图像识别模块根据用户动作图像确定控制指令类型或状态的步骤 ;语音识别状态管理模块根据当前的控制指令类型激活或暂停语音识别的步骤 ;音束形成模块确定拾音方向和拾音接收角的步骤 ;阵列拾音模块根据拾音方向和拾音接收角的限定采集用户发出的语音信号,进行数字化处理形成语音数据的步骤 ;语音识别模块对采集到的语音数据进行识别,形成控制指令的步骤 ;多媒体功能模块执行控制指令,向用户提供相应的多媒体功能的步骤。

[0014] 优选地,上述图像感应模块采集用户动作图像 ;图像识别模块将用户动作图像与预设的图像模版比对,选取与用户动作图像匹配的控制指令类型 ;若比对结果为找到与用户动作图像匹配的控制指令类型,则认为用户所在位置为目标音源位置,向语音识别状态管理模块发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型 ;若未找到与用户动作图像匹配的控制指令类型,则向语音识别状态管理模块发出比对失败信息 ;语音识别状态管理模块根据接收到的启动语音识别信息,向语音识别模块发送启动指令及控制指令类型以激活语音识别,向音束形成模块发送目标音源所在位置信息,并控制多媒体功能模块减小多媒体输出声音的音量 ;或根据接收到的比对失败信息,向语音识别模块发送指令暂停语音识别 ;音束形成模块根据目标音源所在位置信息确定拾音方向和拾音接收角 ;阵列拾音模块根据拾音方向和拾音接收角的限定采集目标音源发出的语音信号,进行数字化处理形成语音数据后发送给语音识别模块 ;语音识别模块根据来自语音识别状态管理模块的启动指令和控制指令类型,对来自阵列拾音模块的语音数据进行识别,形成属于控制指令类型的控制指令,发送给多媒体功能模块 ;多媒体功能模块执行控制指令,向用户

提供相应的多媒体功能。

[0015] 优选地,上述多媒体设备语音控制方法,其中语音识别模块包括本地语音识别模块和云端语音识别模块,语音识别模块预设语音指令词典,包括:本地语音识别模块识别语音数据,将语音数据与语音指令词典中的单词模型对比,若语音数据与至少一单词模型相似度大于预设阈值,则将语音数据解释为与单词模型对应的控制指令,发送给多媒体功能模块;若语音数据与至少一单词模型相似度不大于预设阈值,则将语音数据通过网络发送给云端语音识别模块;云端语音识别模块对语音数据进行语义识别处理,形成控制指令,通过网络发送给多媒体功能模块。

[0016] 本发明结合图像识别和语音识别技术,实现不依赖手持遥控器,不使用近距离拾音模块实现自由便捷的语音控制,有效避免了多媒体设备输出的声音、环境背景声音以及用户的非控制指令语音信号对控制指令语音识别的干扰,实现准确识别用户发出的控制指令。

附图说明

[0017] 下面将结合附图及实施例对本发明作进一步说明,附图中:

[0018] 图 1 是本发明一实施例多媒体设备语音控制系统模块结构示意图;

[0019] 图 2 是本发明一实施例预设图像模板示意图;

[0020] 图 3 是本发明一实施例多媒体设备语音控制系统详细工作流程图;

[0021] 图 4 是本发明一实施例阵列拾音模块 14 排布示意图;

[0022] 图 5 是本发明一实施例多媒体设备语音控制系统基本工作流程图;

[0023] 图 6 是本发明一实施例语音识别模块 15 详细流程示意图。

具体实施方式

[0024] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0025] 参照图 1 示出的多媒体设备语音控制系统模块结构示意图,本发明一实施例多媒体设备 1 包括图像感应模块 10,采集用户动作图像;图像识别模块 11,根据用户动作图像确定控制指令类型或状态;语音识别状态管理模块 12,根据当前的控制指令类型激活或暂停语音识别;拾音模块 14,采集语音数据;语音识别模块 15,对采集到的语音数据进行识别,形成控制指令;多媒体功能模块 16,执行控制指令,向用户提供相应的多媒体功能。

[0026] 参照图 2 所示的预设图像模板示意图,本发明一实施例的图像识别模块 11 预设至少一图像模板,不同的控制指令类型分别对应不同的图像模板,将对用户动作图像与至少一图像模板比对,若找到与用户动作图像相符的图像模板,则认为用户是目标音源,用户接下来发出的语音是属于相应控制指令类型的控制指令,如果比对失败,没有找到与用户动作图像相符的图像模板,则认为用户的动作不是发出控制指令,暂停对其语音进行识别。

[0027] 参照图 3 示出的多媒体设备语音控制系统详细工作流程图,图像识别模块 11 对图像感应模块 10 发来的用户动作图像进行处理,将处理结果与预设的图像模版数据进行比对,选取与用户动作图像匹配的控制指令类型;

[0028] 若比对结果为找到与用户动作图像匹配的控制指令类型,则认为用户所在位置为目标音源位置,向语音识别状态管理模块 12 发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型;

[0029] 若未找到与用户动作图像匹配的控制指令类型,则向语音识别状态管理模块 12 发出比对失败信息。

[0030] 本发明的最佳实施例中,图像识别模块 11 需要对特定的用户动作进行训练。例如多媒体设备 1 通过向用户播放人机交互内容,引导用户将右手自然放置嘴边做喊话状的动作,直到动作符合预设的与“开始语音遥控”控制指令类型相应的第一图像模版。又如,多媒体设备 1 可引导用户将手掌平摊遮住口部,直到动作符合预设的与“静音”控制指令类型相应的第二图像模板。

[0031] 本发明提出一实施例还包括音束形成模块 13,根据目标音源所在位置信息确定拾音方向和拾音接收角,结合阵列拾音技术可有效消除噪声,提高语音识别的准确度。

[0032] 本实施例的拾音模块 14 为阵列拾音模块,包括规则排列的至少一拾音传感器,根据拾音方向和拾音接收角的限定采集目标音源发出的语音信号,进行数字化处理,消除背景杂音,形成语音数据后发送给语音识别模块 15。参照图 4 示出的阵列拾音模块 14 排布示意图,阵列拾音模块 14 可包括多个按照规则几何形状排列的拾音传感器,例如使用等间隔直线排列方式,将多个拾音传感器水平等间隔排列在图像感应模块 10 的两侧。

[0033] 参照图 3 示出的多媒体设备语音控制系统详细工作流程图,音束形成模块 13 确定阵列拾音模块 14 采集声音信号的音束主瓣方向和范围大小,即拾音方向和拾音接收角,据此限定阵列拾音模块 14 采集目标音源发出的语音信号。现有常见的音束形成方法包括延迟-累加方法(传统波束法)、自适应波束法及基于后置自适应滤波法,这三种方法各有优缺点,延迟-累加波束法和后置自适应滤波法适用于非相干噪声和弱相干噪声消除;而自适应波束法适用于消除相干噪声,对非相干噪声或散射噪声消除效果较差。而在本实施例使用环境中,通常是既有相干噪音,又有非相干噪音,本实施例通过图像识别目标音源的方式来巧妙地实现确定拾音方向和拾音接收角。即使有多个电视观众且都处于图像感应识别范围内,也只对目标用户发出的语音信号进行识别。

[0034] 参照图 3 示出的多媒体设备语音控制系统详细工作流程图,本发明还提出语音识别状态管理模块 12 主要负责管理控制多媒体设备语音控制系统的识别状态。当接收到的是启动语音识别信息,则向语音识别模块 15 发送启动指令及控制指令类型以激活语音识别,向音束形成模块 13 发送目标音源所在位置信息,此时用户发出的语音信号才被当作控制指令,由阵列拾音模块 14 发给语音识别模块 15 进行处理;当接收到的是比对失败信息,则向语音识别模块 15 发送指令以暂停语音识别。

[0035] 更进一步的,语音识别状态管理模块 12 激活语音识别,并控制多媒体功能 16 模块减小多媒体输出声音的音量,以智能电视为例即控制电视输出的声音强度减小至适当小于当前目标音源的语音信号强度。不失一般性的,可以将智能电视输出的声音临时置为静音状态,从而避免电视背景成为噪音干扰语音识别。而若完成语音识别或比对失败而暂停语音识别,则不启动语音识别模块 15,智能电视输出的声音调至正常音量,用户的语音信号将被忽略,从而达到避免用户无意识语音命令干扰的目的。

[0036] 本发明又提出,语音识别模块 15 根据来自语音识别状态管理模块 12 的启动指令

和控制指令类型,对来自拾音模块 14 的语音数据进行识别,形成属于控制指令类型的控制指令,发送给多媒体功能模块 16。

[0037] 本实施例提出,语音识别模块 15 预设内置语音指令词典,该语音指令词典中保存经过处理的控制指令语音信号单词模型,包括但不限于“上一频道”、“下一频道”、“增大音量”、“减小音量”、“中央一台”、“湖南卫视”等。语音识别模块 15 将语音数据与语音指令词典中的单词模型对比,若语音数据与至少一单词模型相似度大于预设阈值,则将语音数据解释为与单词模型对应的控制指令,发送给多媒体功能模块 16。

[0038] 为实现复杂的语义识别控制指令,本发明更进一步提出,语音识别模块 15 包括本地语音识别模块 151 和云端语音识别模块 152;前者负责简单控制指令的识别和处理,包括但不限于换台、调整音量、开关机等;后者负责包含语义识别内容的复杂控制指令的识别和处理,采用语音识别云服务的方式实现。

[0039] 参照图 3 示出的多媒体设备语音控制系统详细工作流程图,本地语音识别模块 151 识别语音数据,形成属于控制指令类型的控制指令,发送给多媒体功能模块 16;

[0040] 云端语音识别模块 152 可采用具有语义识别能力的语音识别服务商如科大讯飞提供的在线服务。如果用户的语音数据在本地语音识别模块 152 中无法识别,即语音数据与语音指令词典中的所有单词模型相似度都不大于预设阈值,则将语音数据通过网络发送给云端语音识别模块 152 进行语义识别处理,形成属于控制指令类型的控制指令,发送给多媒体功能模块 16。

[0041] 本发明还提出一种多媒体设备语音控制方法,参照图 5 示出的多媒体设备语音控制系统基本工作流程图,包括:

[0042] 步骤 S1、图像感应模块 10 采集用户动作图像;

[0043] 步骤 S2、图像识别模块 11 根据用户动作图像确定控制指令类型或状态;

[0044] 步骤 S3、语音识别状态管理模块 12 根据当前的控制指令类型激活或暂停语音识别;

[0045] 步骤 S4、音束形成模块 13 确定拾音方向和拾音接收角;

[0046] 步骤 S5、阵列拾音模块 14 根据拾音方向和拾音接收角的限定采集用户发出的语音信号,进行数字化处理形成语音数据;

[0047] 步骤 S6、语音识别模块 15 对采集到的语音数据进行识别,形成控制指令;

[0048] 步骤 S7、多媒体功能模块 16 执行控制指令,向用户提供相应的多媒体功能。

[0049] 参照图 3 示出的多媒体设备语音控制系统详细工作流程图,本发明提出一实施例,包括:

[0050] 步骤 S1、图像感应模块 10 采集用户动作图像;

[0051] 步骤 S21、图像识别模块 11 将用户动作图像与预设的图像模版比对,选取与用户动作图像匹配的控制指令类型;若比对结果为找到与用户动作图像匹配的控制指令类型,则进行步骤 S22;若未找到与用户动作图像匹配的控制指令类型,则进行步骤 S23;

[0052] 步骤 S22、图像识别模块 11 认为用户所在位置为目标音源位置,向语音识别状态管理模块 12 发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型;

[0053] 步骤 S23、图像识别模块 11 向语音识别状态管理模块 12 发出比对失败信息;

[0054] 步骤 S31、语音识别状态管理模块 12 分析接收到的信息,如果是启动语音识别信

息则进行步骤 S32 ;如果是比对失败信息则进行步骤 S35 ;

[0055] 步骤 S32、语音识别状态管理模块 12 向语音识别模块 15 发送启动指令及控制指令类型以激活语音识别 ;

[0056] 步骤 S33、语音识别状态管理模块 12 向音束形成模块 13 发送目标音源所在位置信息 ;

[0057] 步骤 S34、语音识别状态管理模块 12 控制多媒体功能模块 16 减小多媒体输出声音的音量 ;

[0058] 步骤 S35、语音识别状态管理模块 12 向语音识别模块 15 发送指令暂停语音识别 ;

[0059] 步骤 S4、音束形成模块 13 根据目标音源所在位置信息确定拾音方向和拾音接收角 ;

[0060] 步骤 S51、阵列拾音模块 14 根据拾音方向和拾音接收角的限定采集目标音源发出的语音信号 ;

[0061] 步骤 S52、阵列拾音模块 14 对采集到的语音信号进行数字化处理形成语音数据,发送给语音识别模块 15 ;

[0062] 步骤 S61、语音识别模块 15 根据来自语音识别状态管理模块 12 的启动指令和控制指令类型,对来自阵列拾音模块 14 的语音数据进行识别,形成属于控制指令类型的控制指令,发送给多媒体功能模块 16 ;

[0063] 步骤 S7、多媒体功能模块 16 执行控制指令,向用户提供相应的多媒体功能。

[0064] 本实施例的多个模块及其间的工作关系均与上述实施例相似,故不赘述。

[0065] 以下举一具体应用示例 :

[0066] 智能电视 1 的图像感应模块 10 采集到用户甲在感应范围内做出如图 2 所示动作。图像识别模块 11 将用户动作图像与预设的图像模版比对,发现与预设的“开始语音遥控”控制指令类型对应的图像模板相一致,则认为用户甲所在位置为目标音源位置,向语音识别状态管理模块 12 发送目标音源所在位置信息、启动语音识别信息和 / 或控制指令类型 ;语音识别状态管理模块 12 根据收到的启动语音识别信息向语音识别模块 15 发送启动指令及控制指令类型以激活语音识别,向音束形成模块 13 发送目标音源所在位置信息,保证即使有多个电视观众且都处于图像感应识别范围内,只有用户甲才是目标用户,只对他发出的语音信号进行识别。音束形成模块 13 根据目标音源所在位置信息确定拾音方向和拾音接收角 ;阵列拾音模块 14 根据拾音方向和拾音接收角的限定采集用户甲发出的语音信号“湖南卫视”,对其进行数字化处理形成语音数据,发送给语音识别模块 15。语音识别模块 15 对该语音数据进行识别,发现语音数据与一单词模型的相似度大于预设阈值,形成“频道调整至湖南卫视频道”控制指令,发送给多媒体功能模块 16。多媒体功能模块 16 执行控制指令,将频道调整至湖南卫视频道。

[0067] 基于上述实施例,本发明还提出一种多媒体设备语音控制方法。参照图 6 示出的语音识别模块 15 详细流程示意图,语音识别模块 15 包括本地语音识别模块 151 和云端语音识别模块 152,语音识别模块 15 预设语音指令词典,还包括 :

[0068] 步骤 S611、本地语音识别模块 151 识别语音数据,将语音数据与语音指令词典中的单词模型对比,若语音数据与至少一单词模型相似度大于预设阈值,则进行步骤 S612,否则进行步骤 S613 ;

[0069] 步骤 S612、本地语音识别模块 151 将语音数据解释为与单词模型对应的控制指令,发送给多媒体功能模块 16;

[0070] 步骤 S613、将语音数据通过网络发送给云端语音识别模块 152;

[0071] 步骤 S614、云端语音识别模块 152 对语音数据进行语义识别处理,形成控制指令,通过网络发送给多媒体功能模块 16。

[0072] 以下举一具体应用示例,本示例的步骤 S1 到 S51 均与上一具体应用示例相同,故不赘述。阵列拾音模块 14 根据拾音方向和拾音接收角的限定采集用户甲发出的语音信号“给我来一首刘德华的歌”,对其进行数字化处理形成语音数据,发送给语音识别模块 15。语音识别模块 15 的本地语音识别模块 151 识别语音数据,将语音数据与语音指令词典中的单词模型对比,结果没有找到与语音数据相似度大于预设阈值的单词模型,则将语音数据通过网络发送给云端语音识别模块 152。云端语音识别模块 152 对语音数据进行语义识别处理,根据用户语音数据形成“播放刘德华的歌曲”控制指令,通过网络发送给多媒体功能模块 16。多媒体功能模块 16 执行该控制指令,通过搜索引擎自动搜索一首刘德华的歌曲,将音频数据下载并发送给智能电视 1 内置的音乐播放模块,播放音频数据。

[0073] 本发明结合图像识别和语音识别技术,实现不依赖手持遥控器,不使用近距离拾音器件实现自由便捷的语音控制,有效避免了多媒体设备输出的声音、环境背景声音以及用户的非控制指令语音信号对控制指令语音识别的干扰,实现准确识别用户发出的控制指令,还可实现多用户分别或共同控制多媒体设备。

[0074] 以上仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

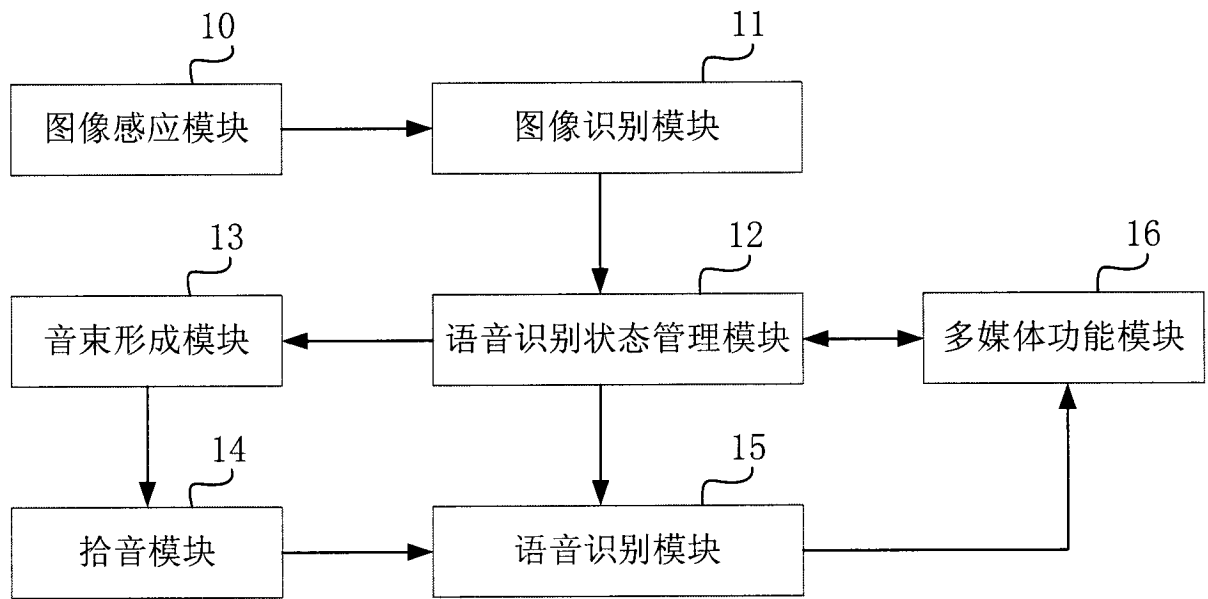


图 1



图 2

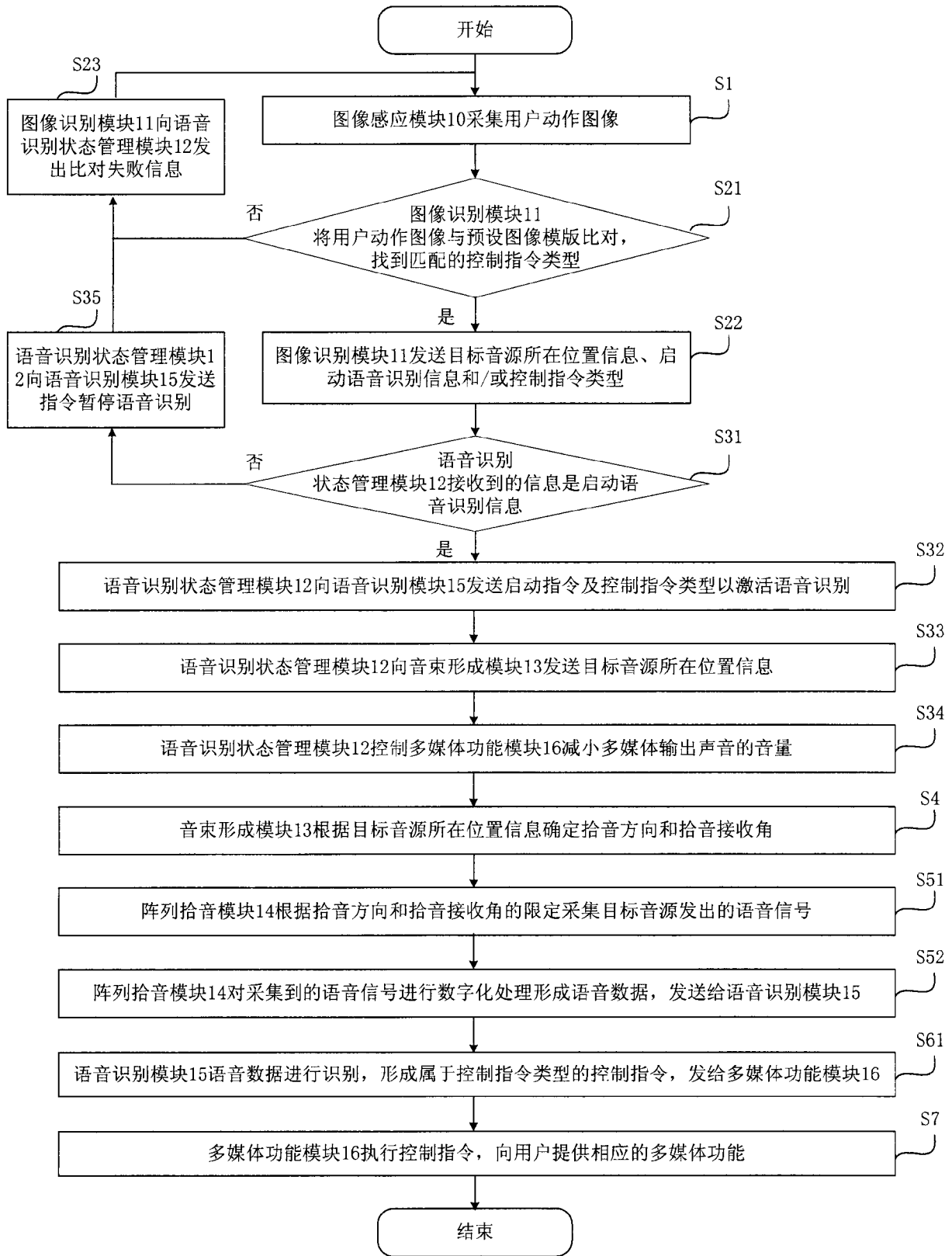


图 3

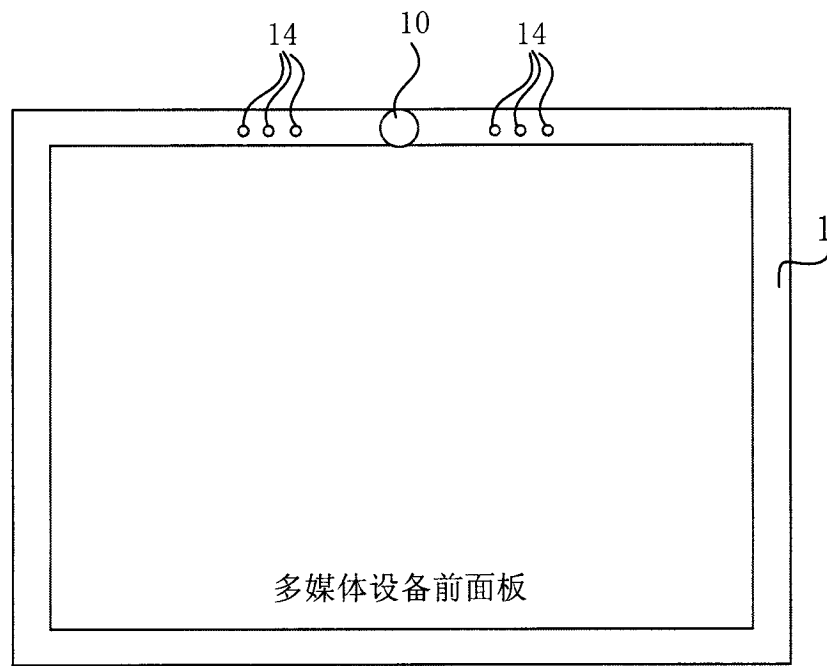


图 4

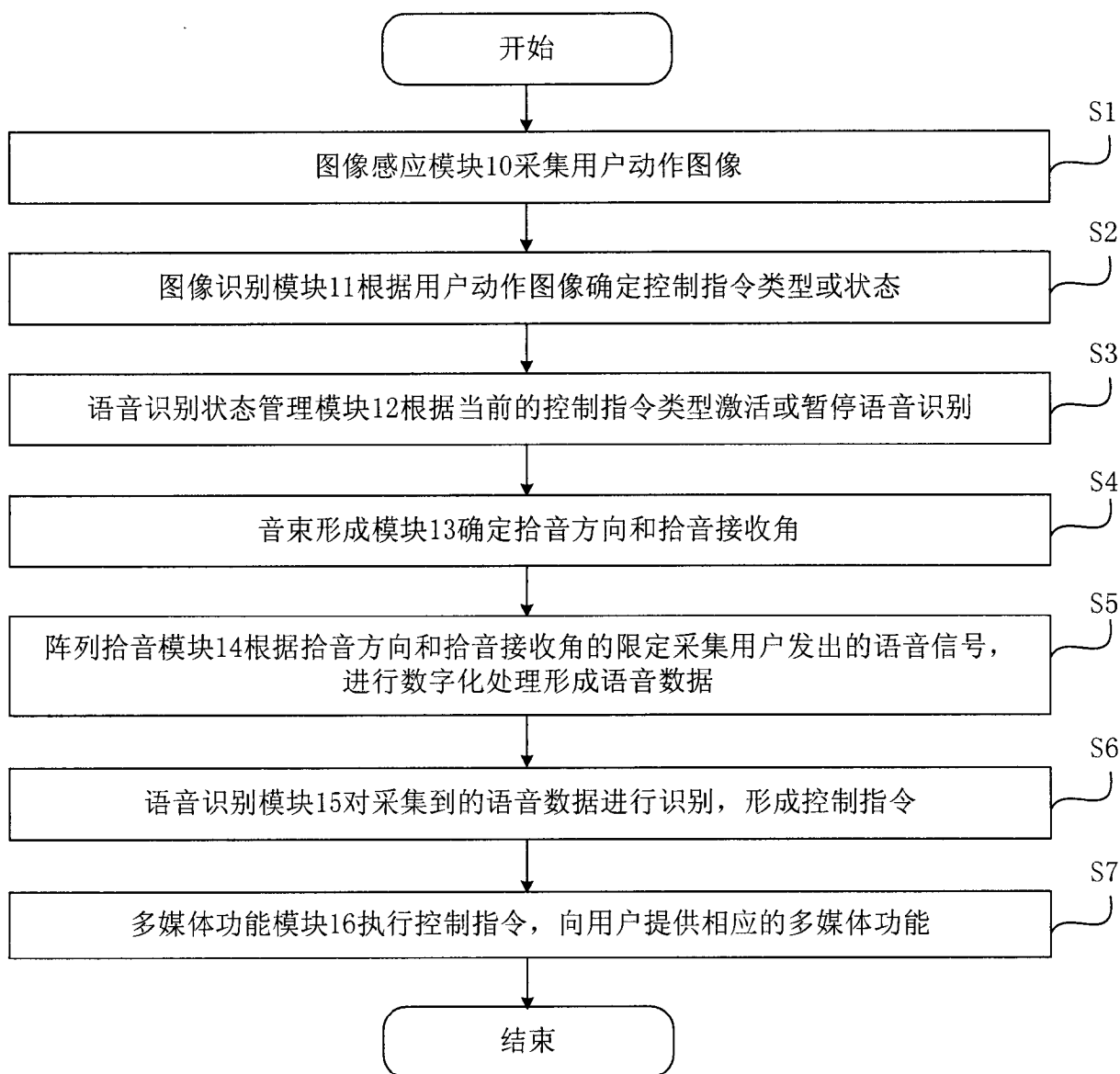


图 5

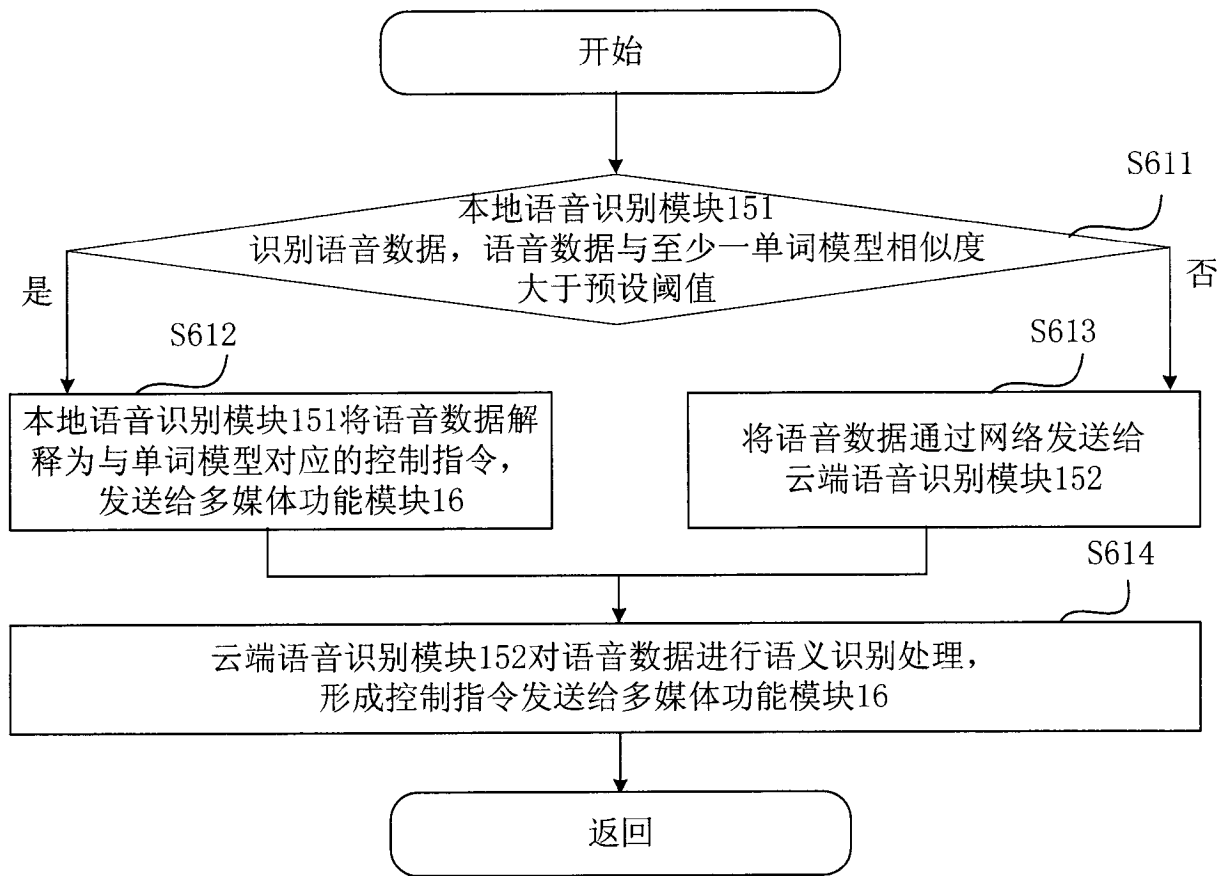


图 6