



(12) 发明专利

(10) 授权公告号 CN 1783998 B

(45) 授权公告日 2010.07.21

(21) 申请号 200510118590.9

(22) 申请日 2005.10.25

(30) 优先权数据

10/978,172 2004.10.30 US

(73) 专利权人 微软公司

地址 美国华盛顿州

(72) 发明人 R·G·卡特勒

(74) 专利代理机构 上海专利商标事务所有限公司 31100

代理人 钱静芳

(51) Int. Cl.

H04N 7/15(2006.01)

G06F 17/30(2006.01)

(56) 对比文件

CN 1479525 A, 2004.03.03, 全文.

US 2003/0210329 A1, 2003.11.13, 说明书
41、141-136 段、图 2、15.

审查员 吴琼

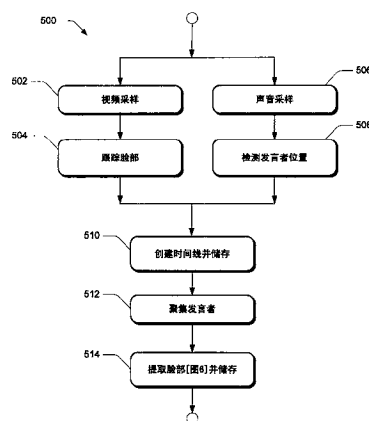
权利要求书 3 页 说明书 8 页 附图 6 页

(54) 发明名称

用于记录的会议时间线的自动脸部提取

(57) 摘要

一次聚会或会议中的发言者的脸部被自动检测,对应于每个发言者的脸部图像储存在脸部数据库中。创建一条时间线在会议记录的回放中图形化地标识发言者的发言时间。示出脸部图像以标识与时间线相关联的每个发言者,取代了一般地在时间线中识别每个用户。



1. 一种自动脸部提取方法,包括:

在一个视频样本中检测两幅或多幅脸部图像;

在一个与所述视频样本相对应的音频样本中检测两个或多个发言者;

检测所述两个或多个发言者中的主要发言者;

在时间上和空间上聚集所述两个或多个发言者;

为每一所检测到的发言者储存一发言者时间线,所述发言者时间线通过发言者标识符以及沿所述发言者时间线的各个时间处的发言者位置来标识发言者;

为每一所检测到的发言者在脸部数据库中储存至少一幅脸部图像;以及

把发言者时间线和脸部图像与每一所检测到的发言者相关联。

2. 如权利要求 1 所述的方法,其特征在于,所述的检测两幅或多幅脸部图像进一步包括使用脸部跟踪来检测两幅或多幅脸部图像。

3. 如权利要求 1 所述的方法,其特征在于,所述的检测两个或多个发言者进一步包括使用声音源定位来检测两个或多个发言者。

4. 如权利要求 1 所述的方法,其特征在于,进一步包括:

为每个发言者标识多于一幅脸部图像;以及

选择一幅最佳脸部图像来储存在所述脸部数据库中。

5. 如权利要求 4 所述的方法,其特征在于,所述的选择进一步包括选择一幅包括最正面脸部视图的脸部图像作为所述最佳脸部图像。

6. 如权利要求 4 所述的方法,其特征在于,所述的选择进一步包括选择一幅呈现最少运动的脸部图像作为所述最佳脸部图像。

7. 如权利要求 4 所述的方法,其特征在于,所述的选择进一步包括选择一幅呈现最大对称性的脸部图像作为所述最佳脸部图像。

8. 如权利要求 1 所述的方法,其特征在于,所述的发言者位置由视频样本坐标所标识的发言者边界框来标记。

9. 如权利要求 1 所述的方法,其特征在于,所述的发言者位置由视频样本中的方位角和仰角所标识的发言者脸部角度来标记。

10. 一种自动脸部提取方法,包括:

显示一个其中包含有两个或多个发言者的音频/视频(A/V)样本;

检测所述两个或多个发言者中的主要发言者;

在时间上和空间上聚集所述两个或多个发言者;

显示一个对应于所述两个或多个发言者中的每个发言者的发言者时间线,所述发言者时间线指示了对应于该发言者时间线的发言者在沿着时间连续区的哪些点上正在进行发言;

将发言者脸部图像与每个发言者时间线相关联,所述发言者脸部图像对应于与所述发言者时间线相关联的发言者;以及

与相应发言者时间线一起显示脸部图像。

11. 如权利要求 10 所述的方法,其特征在于,进一步包括从时间线数据库中检索所述发言者时间线,所述时间线数据库通过发言者标识符、发言者位置以及发言者正在进行发言的一个或多个时间来标识每个发言者。

12. 如权利要求 10 所述的方法,其特征在于,进一步包括从脸部数据库中获取发言者脸部图像,所述脸部数据库将每个发言者标识符与对应于该发言者标识符的发言者的至少一幅脸部图像相关联。

13. 一种自动脸部提取方法,包括:

在音频/视频样本中通过发言者标识符标识每个发言者;

在音频/视频样本中为每个发言者标识位置;

检测主要发言者;

在时间上和空间上聚集每个所标识的发言者;

为音频/视频样本中所标识的每个发言者提取至少一副脸部图像;

为音频/视频样本中所标识的每个发言者创建一条发言者时间线,每条发言者时间线指示了时间、发言者标识符和发言者位置;以及

将发言者的脸部图像与对应于同一发言者的发言者时间线相关联。

14. 如权利要求 13 所述的方法,其特征在于,进一步包括使用声音源定位来标识每个发言者。

15. 如权利要求 13 所述的方法,其特征在于,进一步包括使用脸部跟踪器来标识每个发言者。

16. 如权利要求 13 所述的方法,其特征在于,所述发言者位置由音频/视频样本中的发言者边界框来标识。

17. 如权利要求 13 所述的方法,其特征在于,进一步包括储存所述发言者时间线和所述脸部图像,并将每条发言者时间线与合适的脸部图像相链接。

18. 如权利要求 13 所述的方法,其特征在于,进一步包括为每个发言者提取多于一幅脸部图像。

19. 如权利要求 18 所述的方法,其特征在于,进一步包括选择一幅最佳脸部图像来与发言者时间线相关联。

20. 如权利要求 19 所述的方法,其特征在于,所述的选择一最佳脸部图像进一步包括选择一幅具有最多正面脸部图像的脸部图像。

21. 如权利要求 19 所述的方法,其特征在于,所述的选择一幅最佳脸部图像进一步包括选择一幅呈现最少运动的脸部图像。

22. 如权利要求 19 所述的方法,其特征在于,所述的选择一幅最佳脸部图像进一步包括选择一幅呈现最大脸部对称性的脸部图像。

23. 一种自动脸部提取系统,包括:

音频/视频样本;

用于标识音频/视频样本中所出现的每个发言者的装置;

用于标识音频/视频样本中所标识的每个发言者的脸部图像的装置;

用于检测主要发言者的装置;

用于在时间上和空间上聚集每个发言者的装置;

用于为音频/视频样本中所标识的每个发言者创建发言者时间线的装置;以及

用于将脸部图像与合适的发言者时间线相关联的装置。

24. 如权利要求 23 所述的系统,其特征在于,所述用于标识每个发言者的装置进一步

包括声音源定位器。

25. 如权利要求 23 所述的系统,其特征在于,所述用于标识脸部图像的装置进一步包括脸部跟踪器。

26. 如权利要求 23 所述的系统,其特征在于,所述发言者时间线通过发言者标识符和沿时间连续区域的多个时间中的每一个时间的发言者位置来标识与发言者时间线相关联的发言者。

27. 如权利要求 26 所述的系统,其特征在于,所述的将脸部图像与合适的发言者时间线相关联进一步包括将每个脸部图像与发言者标识符相关联。

28. 如权利要求 23 所述的系统,其特征在于,进一步包括储存所述发言者时间线和所述脸部图像。

29. 如权利要求 28 所述的系统,其特征在于,所述发言者时间线和所述脸部图像被分开存储。

30. 如权利要求 23 所述的系统,其特征在于,所述音频/视频样本进一步包括一记录的会议。

用于记录的会议时间线的自动脸部提取

[0001] 有关申请的交叉引用

[0002] 本申请是于 2002 年 6 月 21 日提交的序号为 10/177315, 发明名称为“分布式会议的系统与方法 (A System and Method for Distributed Meetings)”的美国专利申请的部分延续, 该申请已转让给微软公司, 本申请的受让人。申请人要求所述申请的申请日的优先权, 该专利申请整体引用在此作为参考。

(1) 技术领域

[0003] 下文的说明一般地涉及视频图像处理。更具体地, 下文的说明涉及提供一条索引的时间线用于视频回放。

(2) 背景技术

[0004] 包含多于一个发言者的场景的记录的視頻的回放——如记录的会议的回放——通常是通过一条索引的时间线同期显示的。使用时间线, 用户可以快速移至会议中的一个特定的时间通过操纵一个或多个时间线的控件。当视频包含多于一个发言者时, 可以使用多条时间线其中每条时间线与一个特定的发言者相关联。每条时间线指示了对应的发言者何时发言。那样, 用户可以导航至会议中某个特定发言者正在发言的部分。

[0005] 这种多条时间线可以用一般的方法标上标签以识别每一个发言者, 例如, “发言者 1”、“发言者 2”等。用于自动将时间线标记于具体的发言者姓名的现有技术是不准确的并且还可能需要一个用户的数据库以及用户的相关的声音样本和脸部样本, 这可能会蒙受安全性与隐私的问题。

(3) 附图说明

[0006] 通过参考下文的具体的说明书, 并结合附图, 前述的方面与本发明的许多伴随的优势将变得更容易体会以及变得更好理解, 其中:

[0007] 图 1 是一幅方框图, 描述了构成用于实现本发明的示例性系统的通用计算 / 摄像设备。

[0008] 图 2 是一幅方框图, 表述了一个示例性的全景摄像机和客户端设备。

[0009] 图 3 是一个示例性的回放屏幕的示意图, 带有一个全景图像和一条面部图像时间线。

[0010] 图 4 是一个示例性的回放屏幕, 带有一个全景图像和一条面部图像时间线。

[0011] 图 5 是一个示例性的用于创建带有面部图像的时间线的方法的实施例的流程图。

[0012] 图 6 是一个示例性的用于创建面部数据库的方法的实施例的流程图。

(4) 具体实施方式

[0013] 下文的描述涉及用于在多发言者环境中自动检测每个发言者的脸部并将一个或多个发言者的脸部图像与对应于该发言者的一部分时间线相关联的各种实施方式与实施

例。这类特殊的标记对于一般标记的优势在于观看者可以更容易地确定哪一部分时间线对应于多个发言者中的某一个。

[0014] 在下面的讨论中,将描述一个全景摄像机的例子,其中全景摄像机用于记录一个具有多于一个参与者和 / 或发言者的会议。虽然所描述全景摄像机包括多个摄像机,但下面的描述还是涉及到单个摄像机和具有两个或更多摄像头的多摄像头设备。

[0015] 全景图像被输入至一个脸部跟踪器 (FT),该跟踪器在会议中检测并跟踪脸部。一个麦克风阵列被输入至一个声音源定位器 (SSL),该定位器根据声音检测发言者的位置。脸部追踪系统和声音源定位器的输出被输入至一个虚拟电影放映机以检测发言者的位置。

[0016] 发言者通过一个发言者聚集模块被后置处理,该模块将发言者在时间上和在空间上聚集以更好地描绘具有两根或多条独立时间线的集合时间线。该 (集合) 时间线储存在一个时间线数据库中。创建一个脸部数据库为每一个发言者储存一张或多张照片,每个脸部的至少一张将用于与发言者相关联的时间线。

[0017] 这里表述和声明的概念在下文中会非常详细的描述,有关于一个或多个适合的工作环境。下文描述的元件中的一些也描述于 2002 年 6 月 21 日提交的第 10/177,315 号,题为“分布式会议的系统和方法”(A System and Method for Distributed Meetings) 美国专利申请中。参考并合并于此。

[0018] 示例性工作环境

[0019] 图 1 是一幅方框图,描述了一个通用计算 / 摄像设备。图 1 示出了适用的计算系统环境 100 的例子。计算系统环境 100 仅仅是适用的计算环境的一个例子,并不打算暗示对本发明的使用范围或功能的任何限制。也不应将计算环境 100 解释成相对于示例性操作环境 100 中示出的组件的任一个或其组合具有任何依赖或要求。

[0020] 本发明可以用各种其它通用或专用计算系统环境或配置来运行。适用于本发明使用的公知的计算系统、环境和 / 或配置的例子包括但不限于:个人计算机、服务器计算机、手持或膝上型设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费电子设备、网络 PC、小型机、大型机、包括任何以上系统或设备的任一个的分布式计算环境等等。

[0021] 本发明可以在诸如由计算机执行的程序模块等计算机可执行指令的通用上下文中描述。一般而言,程序模块包括完成特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。本发明还可以在分布式计算环境中实践,其中任务由通过通信网络链接的远程处理设备来完成。在分布式计算环境中,程序模块可以位于本地或远程计算机存储介质中,包括存储器存储设备。

[0022] 参照图 1,用于实现本发明的示例性系统包括计算机 110 形式的通用计算设备。计算机 110 的组件可包括但不限于:处理单元 120、系统存储器 130 和将包括系统存储器的各种系统组件耦合到处理单元 120 的系统总线 121。系统总线 121 可以是若干种总线结构的任何一种,包括存储器总线或存储器控件、外围总线、和使用多种总线体系结构的任何一种的局部总线。作为例子,而非限制,这种体系结构包括工业标准结构 (ISA) 总线、微通道结构 (MCA)、增强型 ISA (EISA) 总线、视频电子技术标准协会 (VESA) 局部总线、以及外围部件互连 (PCI) 总线 (也称为 Mezzanine 总线)。

[0023] 计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是可由计算机 110 访问的任何可用介质,包括易失性和非易失性介质、可移动和不可移动介质。作为例

子,但非限制,计算机可读介质可包含计算机储存介质或通信介质。计算机储存介质包括以任何方法和技术实现来存储诸如计算机可读指令、数据结构、程序模块或其它数据等信息的易失性和非易失性、可移动和不可移动介质。计算机储存介质包括,但不限于:RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘(DVD)或其它光盘存储、磁带盒、磁带、磁盘储存或其它磁储存设备,或者能够用来存储所要求的信息并能够由计算机 110 访问的任何其它介质。通信介质通常在载波或其它传输机制等已调制数据信号中具体化计算机可读指令、数据结构、程序模块或其它数据,并且包括任何信息递送介质。术语“已调制数据信号”指其一个或多个特征以在信号中编码信息的方式而设置或改变的信号。作为例子,但非限制,通信介质包括有限介质,如有限网络或直接线路连接,以及无线介质,如声学、RF、红外和其它无线介质。以上各种组合也应该被包括在计算机可读介质的范围之内。

[0024] 系统存储器 130 包括易失性和 / 或非易失性存储器形式的计算机储存介质,例如只读存储器 (ROM) 131 和随机存取存储器 (RAM) 132。基本输入 / 输出系统 133 (BIOS) 包含如在启动时帮助在计算机 110 中的元件之间传输信息的基本例程,通常储存在 ROM 131 中。RAM 132 通常包含处理单元 120 可直接访问和 / 或当前正在操作的数据和 / 或程序模块。作为例子,但非限制,图 1 示出操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

[0025] 计算机 110 还可包括其它可移动 / 不可移动、易失性 / 非易失性计算机储存介质。仅仅作为例子,图 1 示出读取或写入不可移动、非易失性磁介质的硬盘驱动器 141、读取或写入可移动、非易失性磁盘 152 的磁盘驱动器 151、以及读取或写入可移动、非易失性光盘 156,例如 CD-ROM 或其它光介质的光盘驱动器 155。可用于示例性操作环境中的其它可移动 / 不可移动、易失性 / 非易失性计算机储存介质包括但不限于:磁带盒、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 通常通过不可移动存储器接口 (如接口 140) 连接到系统总线 121,而磁盘驱动器 151 和光盘驱动器 155 通常由可移动存储器接口 (例如接口 150) 连接到系统总线 121。

[0026] 以上讨论并且在图 1 中示出的驱动器及其相关联的计算机储存介质为计算机 110 提供了计算机可读指令、数据结构、程序模块和其它数据的存储。例如,在图 1 中,例如,在图 1 中,硬盘驱动器 141 被示出为储存操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147。注意,这些组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同或不同。操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 在此被给以不同的标号以表示至少它们是不同的副本。用户可以通过输入设备,如键盘 162 和定点设备 161 (通常指鼠标、跟踪球或触摸垫) 输入命令和信息到计算机 110。其它输入设备 (未示出) 可包括话筒、操纵杆、游戏手柄、圆盘式卫星天线、扫描仪等等。这些和其它输入设备经常通过耦合到系统总线 121 的用户输入接口 160 连接到处理单元 120,但是也可以由其它接口和总线结构,如并行端口、游戏端口或通用串行总线 (USB) 连接。监视器 191 或其它类型显示设备也通过接口,如视频接口 190 连接到系统总线 121。除了监视器以外,计算机还可包括其它外围输入设备,如扬声器 197 和打印机 196,它们可以通过输出外围接口 195 连接。能够捕捉图像序列 193 的摄像机 192 (如数字 / 电子静态或视频摄像机,或者胶卷 / 照片扫描仪) 也能够作为个人计算机 110 的输入设备被包括在内。此外,尽管仅仅描述了一台摄像机,然而也可包括多台摄像机,作为个人计算机 110 的输入设备。来自一台

或多台摄像机的图像 193 通过适当的摄像机接口 194 输入到计算机 110。该接口 194 连接到系统总线 121, 因此允许图像被路由到并储存在 RAM 132 中, 或者与计算机 110 相关联的其它数据储存设备之一中。然而, 要注意, 图像数据也能够从上述任一计算机可读介质输入到计算机 110, 而不要求使用摄像机 192。

[0027] 计算机 110 可以使用到一个或多个远程计算机 (如远程计算机 180) 的逻辑连接在网络环境中操作。远程计算机 180 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其它公共网络结点, 并且通常包括许多或所有关于计算机 110 所描述的元件, 尽管在图 1 中仅仅示出了存储器储存设备 181。在图 1 中描述的逻辑连接包括局域网 (LAN) 171 和广域网 (WAN) 173, 但是还可以包括其它网络。这样的网络环境普遍存在于办公室、企业范围计算机网络、内联网和因特网中。

[0028] 当在 LAN 网络环境中使用时, 计算机 110 通过网络接口或适配器 170 连接到 LAN 171。当在 WAN 网络环境中使用时, 计算机 110 通常包括调制解调器 172 或通过 WAN 173 (例如因特网) 建立通信的其它装置。调制解调器 172 可以是内置或者外置的, 它可以通过用户输入接口 160 或者其它适当的机制连接到系统总线 121。在网络环境中, 相对于计算机 110 所描述的模块或其部分可以存储在远程存储器储存设备中。作为例子, 但非限制, 图 1 示出远程应用程序 185 驻留在存储器设备 181 上。将会明白, 示出的网络连接是示例性的, 并且可以使用在计算机之间建立通信链路的其它手段。

[0029] 示例性全景摄像机和客户机设备

[0030] 图 2 是一幅方框图, 表述了一个示例性的全景摄像机设备 200 和示例性的客户端设备 222。虽然示出了一个详细的配置, 但需注意的是全景摄像机 200 可以是任何包含全景摄像机或等同功能的设备。比全景摄像机设备 200 中所包括的比所示的更多或更少的组件可能包含在一个实际的申请中, 结合了此处所述的一项或多项技术特征。

[0031] 全景摄像机设备 200 包括一个处理器 202 和存储器 204。全景摄像机设备 200 通过将多个摄像机 206 (从 206₁ 至 206_n) 产生的几幅独立图像拼接在一起生成一个全景图像。全景图像可以是一幅完整的 360 度图像或者是其中的一部分。需注意的是虽然这里示出并描绘了全景摄像机设备 200, 但所述的技术也可用单个摄像机完成。

[0032] 全景摄像机设备 200 也包括一个麦克风阵列 208。如下文将要更详细的描述的, 配置一个麦克风阵列使声音方向可以被定位。换句话说, 输入至麦克风阵列的声音的分析产生一个被检测到的声音产生的方向。发言者 210 也可包含在全景摄像机设备 200 中以打开一个扩音器或者发送通知信号及其类似至用户。

[0033] 存储器 204 存储几个摄像机设置 212 如校准数据、曝光设置、拼接表等。控制摄像机功能的操作系统 214 与一个或多个摄像机软件应用程序 216 一起也存储在存储器 204 中。

[0034] 全景摄像机设备 200 还包括一个输入 / 输出 (I/O) 模块 218 用于发送来自全景摄像机设备 200 的数据并接受送至全景摄像机设备 200 以及各种其它摄像功能中所需的硬件 220 元件的数据。

[0035] 全景摄像机设备 200 与至少一个客户机设备 222 通信, 该客户机设备包括一个处理器 224、存储器 226、大容量存储设备 242 (如硬盘驱动器) 以及执行下文的客户机设备 222 的功能所需要的其它硬件 230。

[0036] 存储器 226 储存一个脸部跟踪器 (FT) 模块 230 和一个声音源定位 (SSL) 模块 232。脸部跟踪器模块 230 和声音源定位 (SSL) 模块 232 连同—个虚拟电影放映机 234 用于检测摄像机场景中的人物并确定该人物是否在发言以及何时在发言。几种传统的声音定位方法中的任一种都可使用。各种脸部跟踪方法 (或人物检测和跟踪系统), 包括在相关的源申请中的所描述的方法, 可如所描述的用在此处。

[0037] 存储器 226 还储存一个发言者聚集模块 236, 该模块配置为当两个或多个人在发言时确定一个主要的发言者并集中一个特殊的时间线部分给主要发言者。在大多数会议环境中, 存在同时又多于一个人发言的实例。一般地, 一个主要的发言者在发言而另一个人打断该发言者一小段时间或者与发言者进行讨论。发言者聚集模块 236 配置成将发言者在时间上和空间上聚集以理清时间线。

[0038] 虚拟电影放映机 234 生成时间线 238。时间线 238 储存在大容量储存设备 242 上的时间线数据库 244 中。时间线数据库 238 包括多个域, 包括但不限于, 时间、发言者编号、以及在一幅摄像机图像 (x、y、宽、高) 内的发言者边界框。时间线数据库 238 还可包括一个或多个发言者脸部角度 (方位角和仰角)。

[0039] 脸部提取模块 240 也储存在存储器 226 中并配置成从摄像机图像的脸部边界框中 (由脸部跟踪器 230 识别) 提取发言者脸部的图像。脸部提取模块 240 将经提取的脸部图像存储到大容量储存设备 242 上的脸部数据库 246 中。

[0040] 在至少一个实施例中, 可以为一个或多个发言者储存多个脸部图像。可以限定参数以确定在什么特殊时间使用哪幅脸部图像。或者, 用户可以手动从多个脸部图像中选择一个特殊的脸部图像。

[0041] 在至少一个可选择的实施例中, 只为每个发言者储存了一幅单独的脸部图像。储存的脸部图像可以使一幅由脸部提取模块提取的单独图像, 但脸部提取图像 240 也可配置成选择一幅发言者最佳的图像。

[0042] 选择一幅发言者最佳的图像可以通过识别正面脸部角度 (假设一幅具有正面脸部图像的图像相比可选择图像有更好的表现性), 通过识别存在最小量的动作的脸部图像、或者通过识别最大化脸部对称性的脸部图像来实现。

[0043] 经记录的会议 248 也储存在大容量储存设备 242 上这使它可以在之后重播和回放。

[0044] 关于图 2 所示出并描述的元件和功能, 结合下面的附图, 将会在下文更全面的描述。

[0045] 示例性的回放屏幕

[0046] 图 3 是一个示例性的回放屏幕 300 的示意图, 带有一个全景图像 302 和一条面部图像时间线 304。全景图像 302 显示了具有第一会议参与者 303 以及第二会议参与者 305。回放屏幕 300 还显示了具有一个标题栏 306 和一个独立图像 308。独立图像 308 是一个可任选的特征, 其中聚焦了一个特殊的个人, 典型地是主要发言者。在图 3 中, 独立图像 308 显示了第一会议参与者 303 的脸部图像。

[0047] 示例性的回放屏幕 300 还包括一个控制部分 310, 该控制部分包含了控件, 典型地可以在一个媒体播放器中找到, 比如一个播放按钮、一个快进按钮、一个后退按钮等。回放屏幕 300 中包括一个信息区域 312, 在其中可以显示有关于回放屏幕 300 的主体。例如, 会

议题目、会议室编号、会议出席者名单、诸如此类可以显示在信息区域 312 中。

[0048] 脸部图像时间线 304 包括一个第一副时间线 314 对应于第一会议参与者 303 以及一个第二副时间线 316 对应于第二会议参与者。每个副时间线 314、316 显示了时间连接区域的相应会议参与者正在发言的部分。用户可以直接访问副时间线 314、316 上的任何时间点以立即访问特殊的会议参与者正在发言的那部分会议。

[0049] 第一会议参与者 303 的第一脸部图像 318 出现在第一副时间线 314 相邻处以表示第一副时间线 314 与第一会议参与者 303 相关联。第二会议参与者 305 的脸部图像 320 出现在第二副时间线 316 相邻处以表示第二副时间线 316 与第二会议参与者 305 相关联。

[0050] 图 4 显示了一个示例性的回放屏幕 400, 包括与图 3 所示并描述的示例性的回放屏幕 300 相似的元件。关于图 3 中显示和描述的元件和参考数字也会使用在关于图 4 的示例性回放屏幕 400 中。

[0051] 示例性回放屏幕 400 包括一个全景图像 302 和一个脸部图像时间线 304。全景图像 302 显示了第一会议参与者 303 和第二会议参与者 305。标题栏 306 横跨回放屏幕 400 的顶部, 一幅独立图像 408 显示了第二会议参与者 303。

[0052] 示例性的回放屏幕 400 还包括一个白色书写板发言者图像 402 显示了一个位于白色书写板前面的会议参与者 (在这个例子中, 是第二会议参与者 305)。白色书写板发言者图像 402 并不包括在图 3 的回放屏幕 300 中, 并在此处用于显示其它图像是如何被包括在任何特殊的回放屏幕 300 和 400 中的。

[0053] 控制部分 310 包括多媒体控件和一个显示回放屏幕 400 上所显示的会议的有关信息的信息区域 312。

[0054] 脸部图像时间线 304 包括第一副时间线 314、第二副时间线 316 以及第三副时间线 404。需要注意的是虽然图 3 中只有两根副时间线, 但一条时间线可包含任何可管理的数量的副时间线。例如, 在图 4 中, 有三根时间线。

[0055] 需要注意的是虽然在本例中只有两个会议参与者, 但却有三根时间线。这是因为单个发言者可与多于一条的副时间线相关联。在本例中, 当第二会议参与者 305 处于白色书写板时, 第二副时间线 316 与第二会议参与者 305 相关联, 而当第二会议参与者 305 处于白色书写板之外的其它位置时, 第三副时间线 404 与第二会议参与者 305 相关联。

[0056] 这种情况可发生于在会议中一个会议参与者具有多于一个位置时。在本例中虚拟电影放映机 234 在三个位置检测到发言者。不需要知道在那些位置上只有两个发言者在场。这个特性在用户主要对一个发言者在某一位置上的发言感兴趣的情况下对用户起到帮助。例如, 用户可能只想播放发言者在白色书写板前时的那部分经记录的会议。

[0057] 示例性的回放屏幕 400 还包括第一会议参与者 303 的第一脸部图像 318 以及第二会议参与者 305 的第二脸部图像 320。另外, 还包括第三脸部图像 406, 其与第三副时间线 404 相关联。第三脸部图像 406 对应于第二会议参与者 305 的第二个位置。

[0058] 用于表述示例性的回放屏幕 300 和 400 的技术, 结合其它附图, 将会在下文更详细的描述。

[0059] 示例性的方法的实施例: 脸部图像时间线的创立

[0060] 图 5 是一个示例性的用于创建带有面部图像的时间线的方法的实施例的流程图 500。在下面的示例性的流程图 500 的讨论中, 以前的附图中所示的元件和参考数字将继续

沿用。

[0061] 在方框 502,全景摄像机设备 200 对一幅或多幅视频图像进行采样以生成一幅全景图像。全景图像被输入至脸部跟踪器 230(方框 504),脸部追踪器在图像中检测并跟踪脸部。在方框 506 大致同步地,麦克风阵列 208 对与全景图像相对应的声音进行采样,并输入该声音至声音源定位器 232,在方框 508,音源定位器根据经采样的声音检测发言者的位置。

[0062] 在方框 510,虚拟电影放映机 234 处理来自脸部跟踪器 230 和声音源定位器 232 的数据,并创建时间线 238。在方框 512,发言者聚集模块 236 将发言者在时间上和在空间上聚集以巩固和阐明时间线 238 的部分,如前文所述。

[0063] 时间线储存在时间线数据库 244,数据库有下述域:时间、发言者编号、图像内的发言者边界框(x、y、宽、高)、发言者脸部角度(方位角和仰角)等。

[0064] 使用全景图像和得自脸部跟踪器 230 的脸部识别坐标(即脸部边界框),在方框 514,脸部提取器 240 提取一幅发言者的面部图像。经提取的脸部图像储存在脸部数据库 246 中并与发言者编号相关联。

[0065] 如以前所述的,脸部跟踪器 240 可配置成为每位发言者提取多于一幅的图像并使用脸部提取器 240 确定的时间线 238 中最佳的那幅图像。

[0066] 选择一幅“最佳”面部图像和创建脸部数据库 246 的示例性的方法的实施例将结合图 6 在下文中示出并描述。

[0067] 示例性的方法的实施例:创建一个脸部数据库

[0068] 图 6 是一个示例性的流程图 600 表述了用于创建面部数据库的方法的实施例。在下文的关于图 6 的讨论中,以前的一幅或多幅附图中所示的元件和参考数字将继续沿用。

[0069] 在方框 602,脸部提取器 240 从以前所述的全景图像中提取一幅脸部图像。如果发言者的脸部图像还没有储存在脸部数据库 246 中(“否”分支,方框 604),那么在方框 610,脸部图像会存入脸部数据库 246。需要注意的是确定脸部图像是否已存储不必取决于在脸部图像中出现的个人是否具有已存储的他们的相似的图像,而取决于经识别的发言者是否具有已存储的与该发言者相对的图像。这样,如果位于第一个位置的一个发言者具有已存储的面部图像而随后发言者被检测到处于第二位置,处于第二位置的该发言者的面部图像不会与已存储的处于第一位置的该发言者的面部图像做比较以确定是否发言者已有面部图像被存储。

[0070] 如果该发言者的面部图像已存储在面部数据库 246——在下文中,“已存储的面部图像”——(“是”分支,方框 604),则在 606,面部图像与已存储的面部图像做比较。如果脸部提取器 240 确定脸部图像比已存储的面部图像更好或更容易接受(“是”分支,方框 608),则面部图像被存入面部数据库 246,这样就覆盖了原先已存储的面部图像。

[0071] 如果确定脸部图像不比已存储的面部图像更好(“否”分支,方框 608),则该面部图像被丢弃而保留已存储的面部图像。

[0072] 确定哪幅面部图像是较好的面部图像的标准可以是多种多样的。例如,脸部提取器 240 可配置成确定一幅“最佳的”面部图像是一个发言者处于某一位置时所拍摄的,在该位置发言者的脸部大多数是处于正面的位置。或者,如果一个第一面部图像显示了运动的迹象而第二面部图像没有,则提取器 240 可确定第二面部图像是最佳面部图像。或者,脸部

提取器 240 可配置成确定发言者的多幅图像中哪一幅具有最大的对称性,就将那幅面部图像用于时间线中。其它此处未列举的标准也可用作确定最适用于时间线的面部图像。

[0073] 如果有另一个发言者(“是”分支,方框 612),则流程回到方框 602 并为每个唯一的发言者重复该流程。重申的是,在上下文中使用的“唯一的发言者”指的并不必定指的是一个唯一的发言者,因为一个人出现在不同的发言位置也解释成是不同的发言者。当没有更多的唯一的发言者需识别时(“否”分支,方框 612) 流程终止。

[0074] 总结

[0075] 虽然已图示并描述了一个或多个示例性实施例,可以理解的是可以做出各种改变而不背离附加的权利要求的精神和范围。

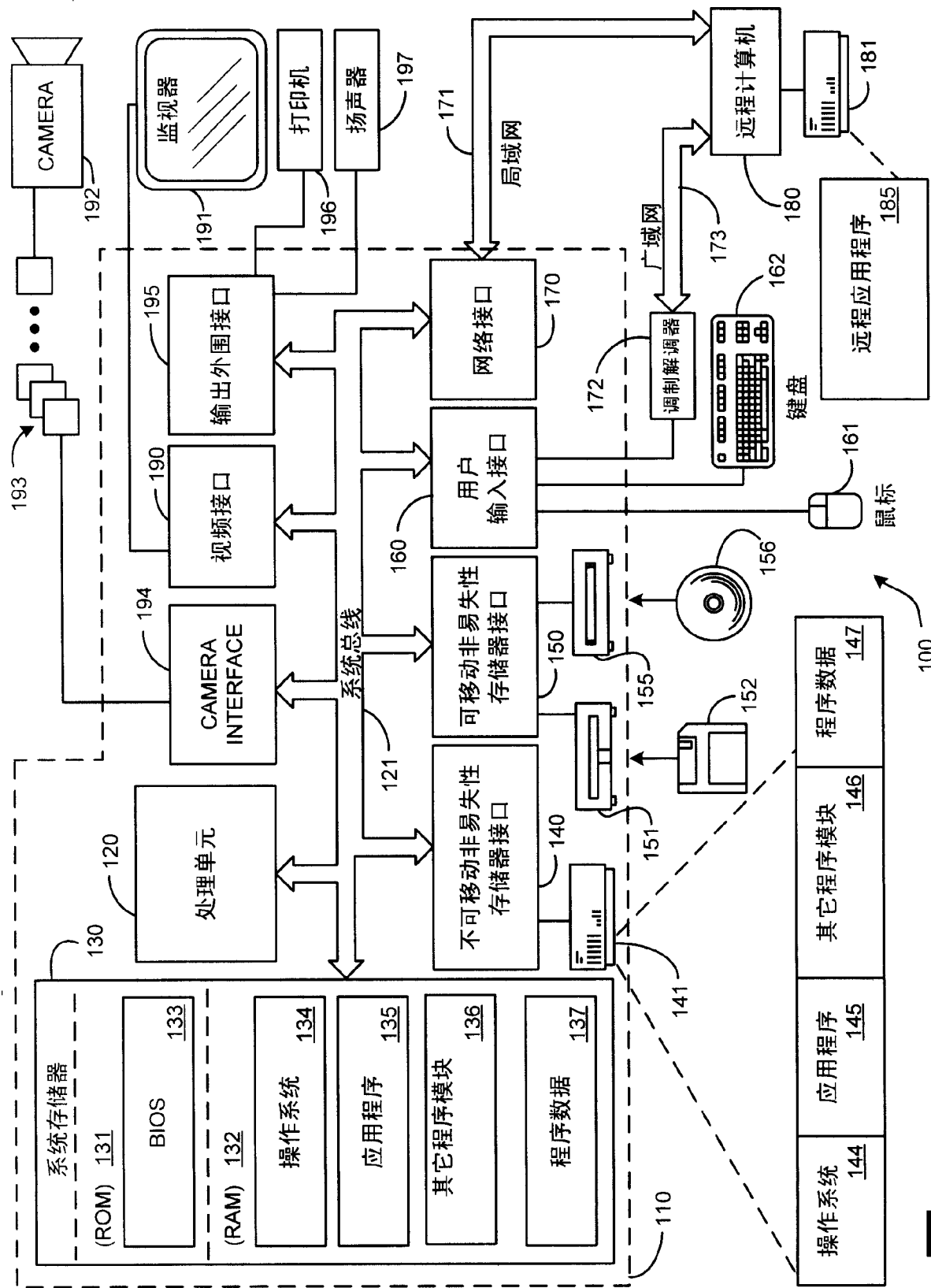


图 1

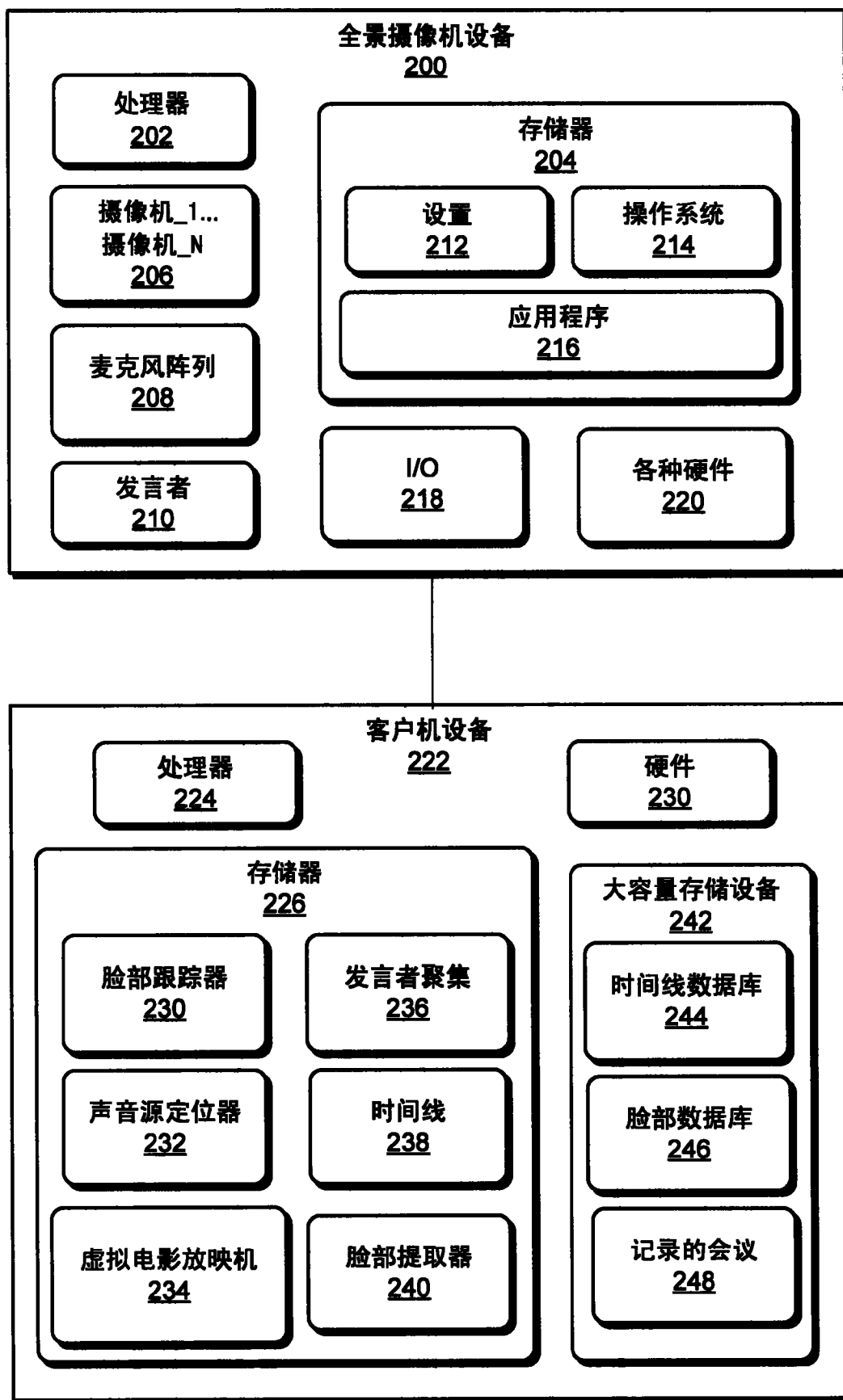


图 2

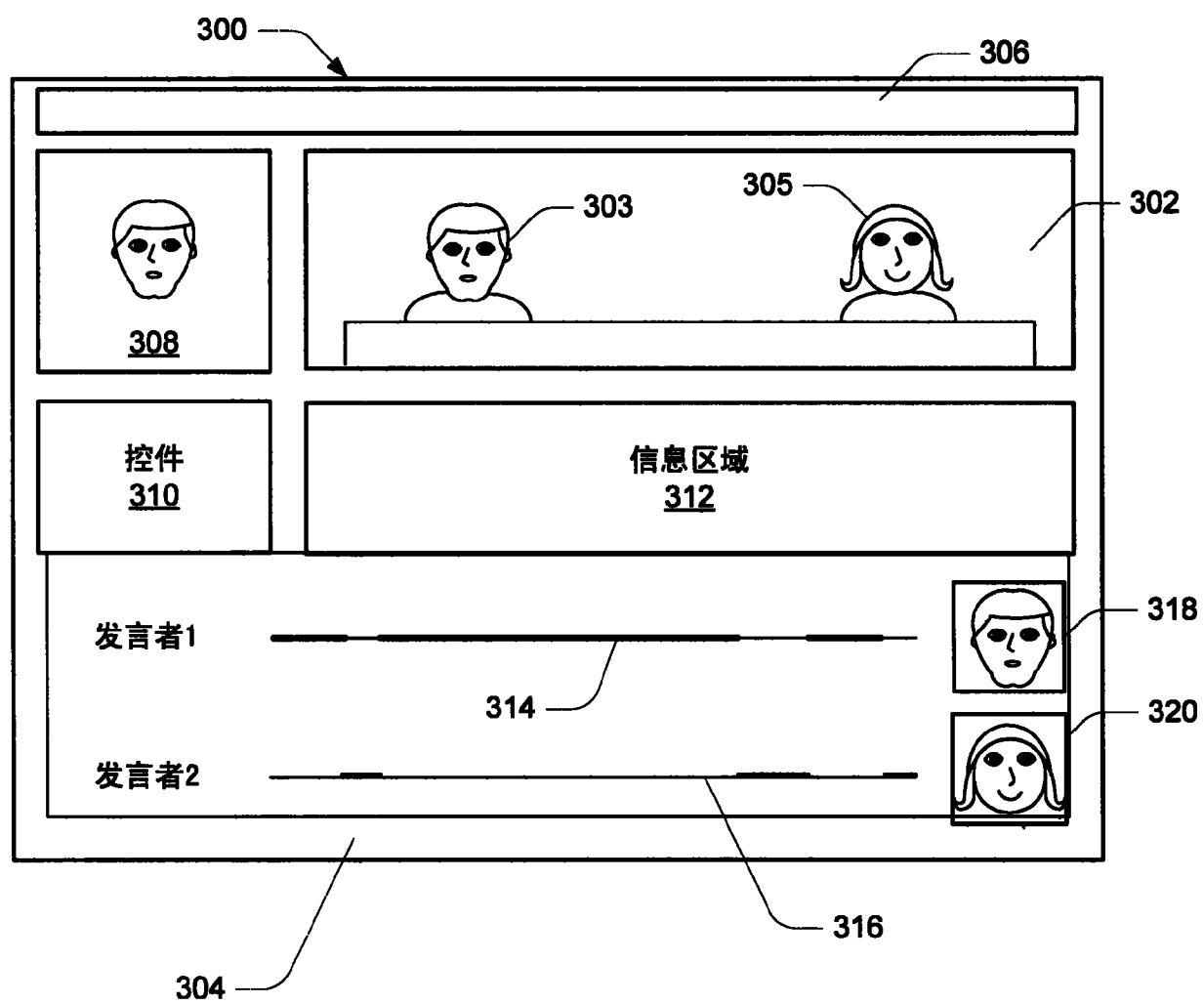


图 3

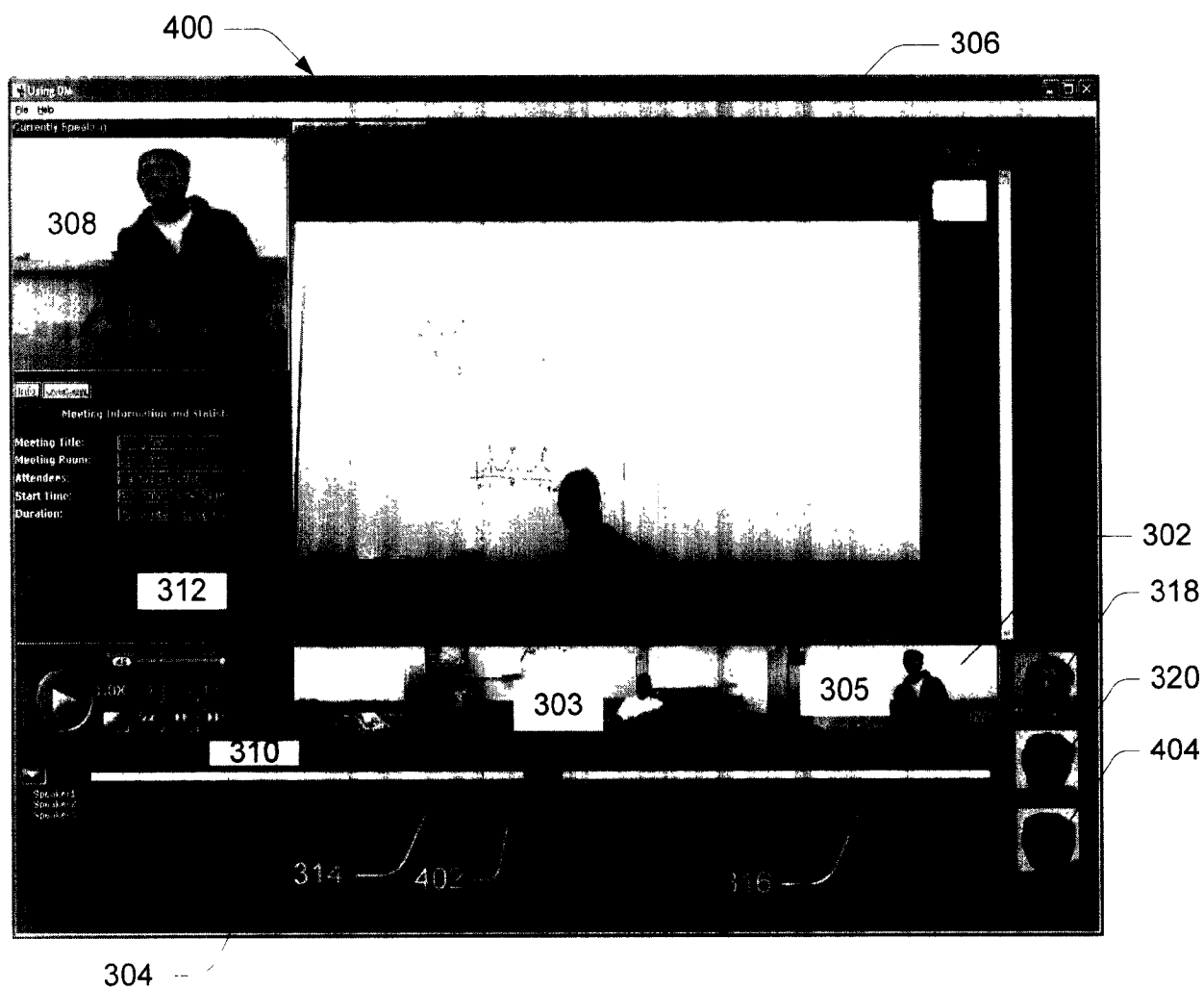


图 4

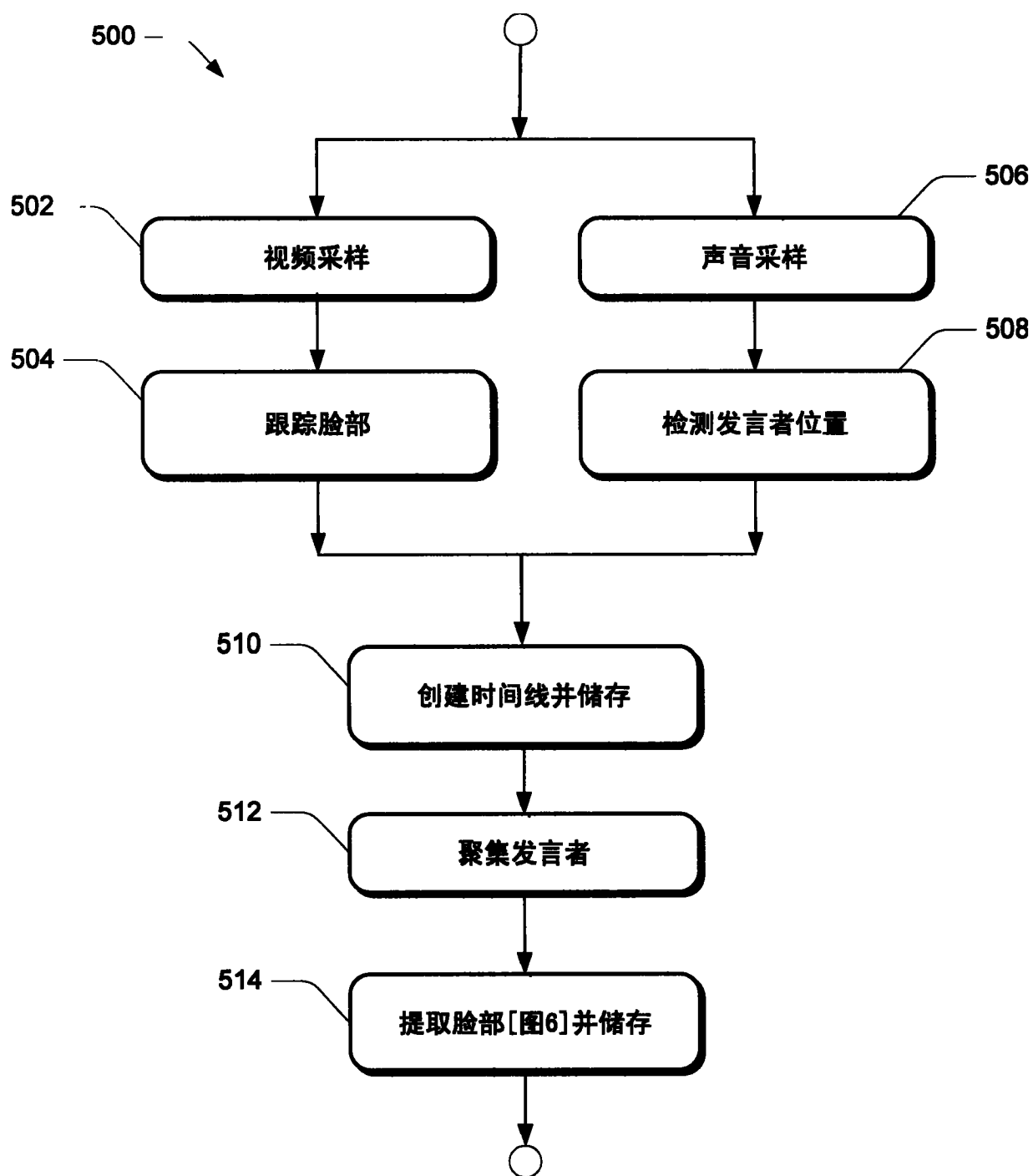


图 5

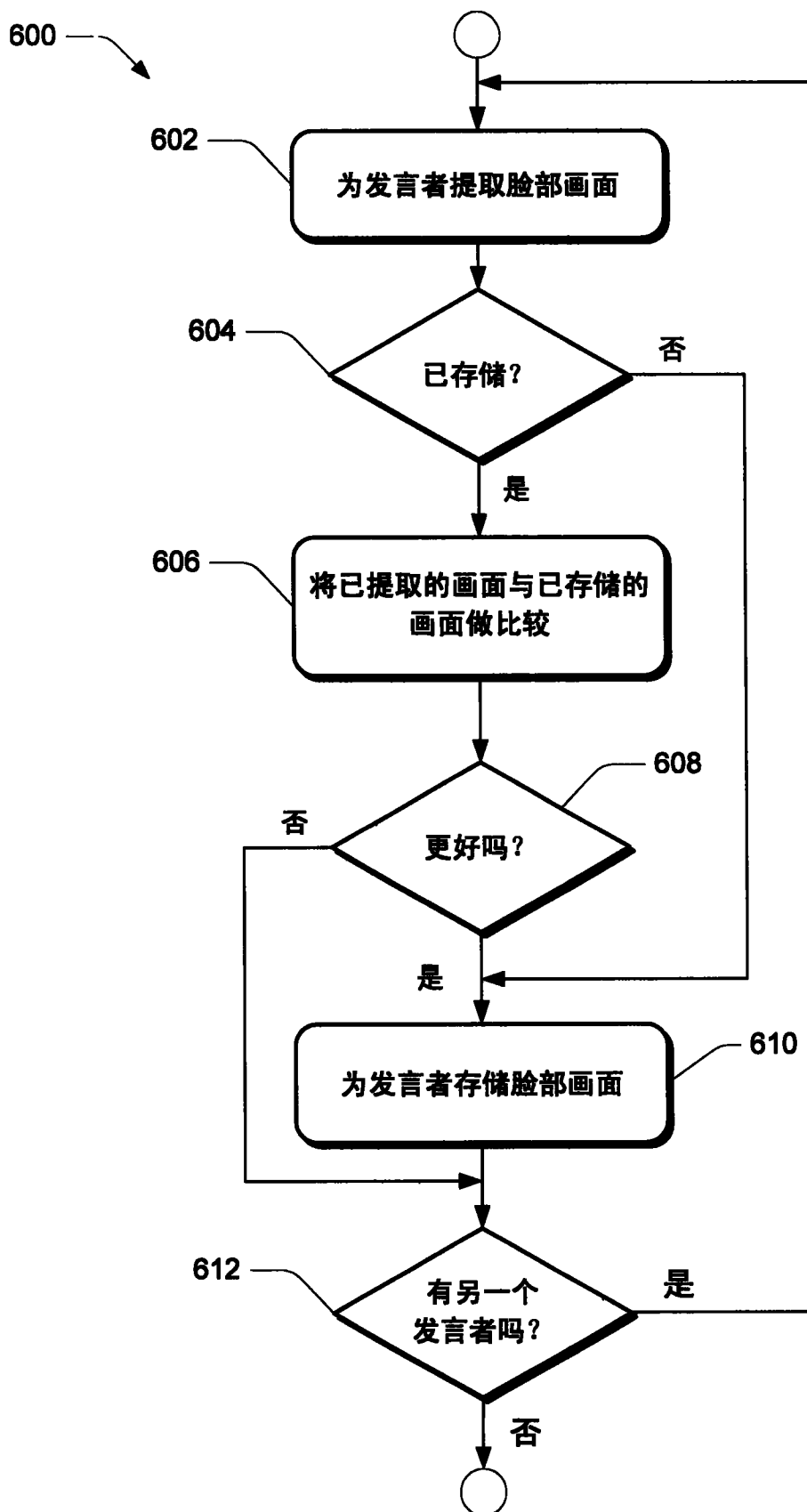


图 6