



(12) 发明专利

(10) 授权公告号 CN 102906810 B

(45) 授权公告日 2015. 03. 18

(21) 申请号 201180020715. 8

(51) Int. Cl.

(22) 申请日 2011. 02. 24

H04N 7/18(2006. 01)

(30) 优先权数据

61/307, 675 2010. 02. 24 US

61/339, 071 2010. 02. 26 US

(56) 对比文件

US 2008/0198222 A1, 2008. 08. 21, 说明书第 0007-0049, 0070-0072, 0100 段、附图 1-3.

US 2008/0024594 A1, 2008. 01. 31, 说明书第 0056, 0098-0100, 0208-0255, 0283-0284 段、附图 1-29.

(85) PCT国际申请进入国家阶段日
2012. 10. 24

US 2008/0198159 A1, 2008. 08. 21, 说明书第 2 页第 2 栏第 0029 段, 第 0093-0094 段、附图 1-11.

(86) PCT国际申请的申请数据
PCT/US2011/026066 2011. 02. 24

US 2004/0215612 A1, 2004. 10. 28, 全文.

(87) PCT国际申请的公布数据
W02011/106520 EN 2011. 09. 01

US 2010/0023878 A1, 2010. 01. 28, 说明书第 4 页第 0050-0055 段、附图 1-30.

(73) 专利权人 爱普莱克斯控股公司
地址 美国加利福尼亚州

审查员 乔元昆

(72) 发明人 O. 迪亚拉梅 D. 米勒 C. 布兰查德
T.C. 多塞 J.M. 苏多尔

(74) 专利代理机构 北京市柳沈律师事务所
11105
代理人 胡琪

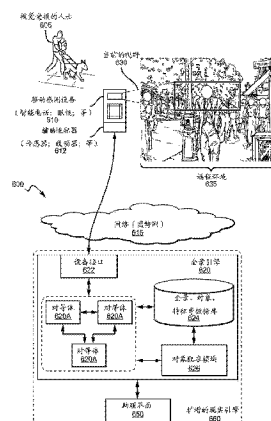
权利要求书2页 说明书24页 附图7页

(54) 发明名称

支持视觉受损的个体的扩增的现实全景

(57) 摘要

提出了一种向远程用户提供实时对象识别的方法和系统。该系统包括包含相机的便携式通信设备、远离该便携式通信设备并可由该便携式通信设备经由网络访问的至少一个客户端-服务器主机、以及可由所述客户端-服务器主机设备访问的识别数据库。驻留在所述客户端-服务器主机设备上的识别应用能够利用该识别数据库来向便携式通信设备的远程用户提供使用便携的通信设备捕获的视觉图像的实时对象识别。在一个实施例中,能看见的助理与便携的通信设备的视觉受损的用户分享扩增的现实全景,其中该全景由来自该设备的传感器数据构造。



1. 一种扩增的现实引擎,包括:

远程设备接口,配置为经由网络从远程移动感测设备接收周围环境传感器数据,该周围环境传感器数据反映视觉受损人士的远程环境并且包括当前视野和设备位置数据;

全景引擎,配置为至少部分地基于所述周围环境传感器数据来构造代表远程环境的 3D 数字模型并包括该环境的一些对象的扩增的现实全景,所述当前视野集成到扩增的现实全景 3D 数字模型中作为设备位置的函数;以及

能看见的助理界面,配置为:

向能看见的助理视觉地展现扩增的现实全景的 3D 数字模型及集成的当前视野,

基于能看见的助理与在 3D 数字模型中并在当前视野之外的至少一个对象之间的交互产生设备命令,

经由远程设备接口,向远程移动感测设备发送能看见的助理选择的传感器控制设备命令,所述设备命令指示所述移动感测设备捕获与在所述扩增的现实全景中并在所述集成的当前视野之外的、能看见的助理选择的对象相关联的额外的传感器数据,以及

通过将额外的对象数据集成到所述 3D 数字模型中来更新所述扩增的现实全景。

2. 根据权利要求 1 所述的引擎,其中当前视野包括由所述远程感测设备捕获的图像数据。

3. 根据权利要求 1 所述的引擎,其中所述扩增的现实全景包括多形态的全景数据。

4. 根据权利要求 3 所述的引擎,其中所述多形态的全景数据包括触觉数据,并且其中设备命令包括与远程环境中所述视觉受损人士附近的真实世界对象相关联的触觉反馈指令,且所述真实世界对象对应于所述扩增的现实全景中的对象。

5. 根据权利要求 1 所述的引擎,其中所述移动感测设备包括智能电话。

6. 根据权利要求 1 所述的引擎,扩增的现实全景包括与视觉受损人士相关联的个人化的全景。

7. 根据权利要求 6 所述的引擎,所述个人化的全景包括在视觉受损人士的管理下的私人周围环境传感器数据。

8. 根据权利要求 7 所述的引擎,其中所述个人化的全景包括作为私人周围环境传感器数据的函数的过滤的部分。

9. 根据权利要求 1 所述的引擎,其中所述扩增的现实全景包括代表历史的周围环境传感器数据的历史部分。

10. 根据权利要求 9 所述的引擎,其中所述扩增的现实全景包括基于当前视野指示扩增的现实全景的历史部分和最近更新的部分之间的区别的加亮的对象。

11. 根据权利要求 1 所述的引擎,其中所述扩增的现实全景包括代表公开地收集的传感器数据的公共部分。

12. 根据权利要求 1 所述的引擎,其中能看见的助理界面还配置为展现指示周围环境传感器数据的获取和当前视野的展现之间的时间差的延迟。

13. 根据权利要求 1 所述的引擎,其中能看见的助理界面还配置为基于下述周围环境传感器数据中的至少两个校正所述扩增的现实全景中当前视野的布置:GPS 位置数据、图像配准数据、加速计数据、磁力计数据、无线信号数据、测距法数据和音频数据。

14. 根据权利要求 1 所述的引擎,其中所述扩增的现实全景包括分配给扩增的现实全

景内的对象的元标签。

15. 根据权利要求 14 所述的引擎,其中所述元标签包括对于能看见的助理可视的、且叠加在扩增的现实全景上的元数据。

16. 根据权利要求 14 所述的引擎,其中所述元标签包括非视觉的元数据。

17. 根据权利要求 1 所述的引擎,其中所述能看见的助理界面还配置为指示相对于当前视野到对象的距离。

18. 根据权利要求 1 所述的引擎,其中所述设备命令包括针对移动感测设备的激活指示对象相对于移动感测设备的位置的致动器的指令。

19. 根据权利要求 1 所述的引擎,其中所述设备命令包括针对移动感测设备的发出指示对象相对于移动感测设备的位置的可听信号的指令。

20. 根据权利要求 1 所述的引擎,其中所述全景引擎包括配置为将所述周围环境传感器数据转换为扩增的现实全景中的对象地标的对象特征配准模块。

21. 根据权利要求 20 所述的引擎,其中所述扩增的现实全景中的对象地标包括图像配准地标。

22. 根据权利要求 21 所述的引擎,其中所述图像配准地标包括尺度不变特征变换(SIFT) 地标。

23. 根据权利要求 1 所述的引擎,其中设备位置数据包括设备方向数据。

支持视觉受损的个体的扩增的现实全景

[0001] 本申请要求于 2011 年 2 月 24 日提交、具有序列号 61/307,675 和于 2010 年 2 月 26 日提交、具有序列号 61/339,071 的美国临时申请的优先权的权益。这里讨论的这些和所有其它外来的材料通过引用整体并入。当并入的参考资料中术语的定义或使用不一致或者与这里提供的该术语的定义相反,则应用这里提供的术语的定义而不应用参考资料中该术语的定义。

技术领域

[0002] 本发明的领域是用于残疾个体的辅助技术。

背景技术

[0003] 不论是称作计算机视觉、机器视觉诸如此类,提供自动的(automated)对象识别的系统传统上一直是在工业和军事应用中颇感兴趣的要实现的课题。提供对象识别的任何系统所遇到的正在发生的挑战之一是由同样的对象、甚至是由从同样的视角(perspective)观看的同样的对象所表示的图像中的可变性,这种可变性是诸如由天气模式、季节转换以及单独一天的过程中日光的偏移产生的光和影的变化的环境因素的结果。为了解决这个和其它重大的技术挑战,提供可靠的对象识别的大多数系统使用一个或多个复杂的算法,该算法需要大量的计算资源用于其执行。因此,实现对象识别的很多传统方法只在相当大的延迟后才能提供感兴趣的对象的识别,需要用于对象图像的处理的昂贵和复杂的计算平台,或可能经受这两个缺点。尽管如此,如在 Fink 等人于 2005 年 1 月 5 日提交的题为“Digital Object Recognition Audio-Assistant for the Visually Impaired”的美国专利申请公开 2005/0208457 中所讨论的,已经针对提供对象识别做出了一些努力。

[0004] 不幸的是,根据传统的技术水平,对象识别系统可以例如向遭受感测损伤的个体(诸如盲人或视觉受损人士)提供特别的益处。原则上,视觉受损的个体的视力可以通过提供对象识别能力功能的电子助理类型设备的对象识别能力而有效地扩增。然而,为了使提供对象识别的电子设备最充分地有益于视觉受损的个体,该设备应处理至少三个标准。这些标准中的一个期望提供电子的视力扩增的该设备是便携的,从而使得视觉受损的个体可以容易地按照期望地运送该设备。这些标准中的一个期望该设备提供灵活且用户友好的界面使得视觉受损的用户能够激活并控制该设备的对象识别功能。此外,为了最充分地使能(empower)视觉受损的个体,同时还保护他们的安全,期望该提供对象识别的设备可靠且实时地进行对象识别,从而使得视觉受损的用户参与(engage)他或她当前环境中的特征。

[0005] 然而,如上面所讨论的,传统的方法用来实现对象识别所需要的计算资源通常是大量的。该计算能力要求可能相当大地超过甚至在相对宽松的性能标准下单个便携设备的资源。如 Kurzweil 等人于 2005 年 4 月 1 日提交的题为“Gesture Processing with Low Resolution Images with High Resolution Processing for Optical Character Recognition for a Reading Machine”的美国专利 7,627,142 中所讨论的,便携设备至多

提供对于特征识别的最小支持。当提供对象识别功能的系统的用户的个人安全和效能要求对象的识别是高度可靠且实时提供时,所需要的计算资源远远超过单个便携电子设备所能够获得的资源。

[0006] 用来帮助视觉受损的用户的一种可能的途径包括与远程助理交互。理想地,远程助理应该可以和视觉受损的用户的整个环境交互。其它人已经朝着使能视觉受损的用户和远程助理之间的交互做出了努力。Eспенlaub 等人于 2007 年 4 月 6 日提交的题为“System and Method for Assisting a Visually Impaired Individual”的美国专利 7864991 讨论了视觉受损的个体无线地向助理发送有关一个情况的视听信息。该助理然后可以用对该情况的解决方案进行回复。遗憾的是,助理缺少对个体的周围环境的完整视野。

[0007] 另一个例子包括 Gallick 于 2006 年 8 月 4 日提交的题为“Communications Device for Visually Impaired Persons”的美国专利申请公开 2008/0043934,其讨论了向视觉受损人士提供具有表面传感器的设备,其中远程助理可以观察该人与该设备的交互。

[0008] 一个更先进的努力包括 Gowda 于 2008 年 2 月 4 日提交的题为“Systems and Method for Tele-presence”的美国专利申请公开 2008/0198222,其使远程助理的概念更进一步。Gowda 指出视觉受损的主体可以与向导连接,其中该向导可使用有关主体的环境的多形态(multi-modal)信息(或许通过将主体指引到对象)来帮助该主体。助理仍旧被限制到由主体提供的视野并且还是缺少获取有关环境的额外数据的能力。

[0009] 有趣的是,只有有限的努力是针对使用扩增的现实系统来帮助视觉受损的个体。仍需要意识到的是远程助理可以与表示视觉受损人士的环境的扩增的现实全景及其对象相交交互。可以基于经由个体的智能电话或周围数据的其他源收集的传感器数据构造扩增的现实全景。该助理可以与对象或扩增的现实全景的其它方面相交交互,并将设备命令发送回个体的智能电话以帮助该个体或获得有关环境的更多的信息。

[0010] 除非上下文指明是相反的,否则这里阐明的所有范围应解释为包含其端点,且开放性的范围应解释为包括商业上实际的值。类似地,除非上下文指明是相反的,否则值的所有列举应该认为包括中间的值。

[0011] 因此,仍存在对于用于视觉受损的个体的远程帮助的需要。

发明内容

[0012] 本发明主题提供其中视觉受损的人士可以从能看见的助理寻求帮助的装置、系统和方法。本发明主题的一个方面包括一种扩增的现实引擎,包括能够构造代表远程环境及其对象的模型的扩增的现实全景的全景引擎。该全景可至少部分地基于从远程视觉受损的人士的智能电话或接近视觉受损人士的其他移动感测设备收集的周围环境传感器数据来构造。优选地,该周围环境传感器数据反映该人士的远程环境,可能包括与图像、视频、音频、加速度、方向、位置、测距法关联的数据或其它传感器数据。能看见的助理使用能看见的助理界面与扩增的现实全景及其对象交互,以向远程的视觉受损人士提供辅助的反馈。该助理界面展现具有集成的远程人士的感测设备的当前视野的扩增的现实全景。人类的或自动的能看见的助理共享远程视觉受损人士的经历并且与扩增的现实全景中的对象交互,使得一个或多个设备命令被发送回该人士的智能电话。设备命令可以指示智能电话或其它移动感测设备捕获与对应的现实世界的对象有关的额外的信息。例如,设备命令可以向视觉

受损人士提供指示接近一个或多个对象的触觉或听觉反馈,或可以控制选择一个或多个适用的对象识别算法或例程以获取额外的数据。

[0013] 在一些实施例中,扩增的现实全景包括多形态的全景数据,其中扩增的现实中的对象包括多个类型的感测数据,可能包括视觉数据、触觉数据、动觉数据、可听到的数据或其它类型的传感器数据。可以用包括不可视数据的元数据对对象加标签。示例的不可视数据包括触觉元数据标签、听觉元数据标签或可由助理影响(leverage)以帮助视觉受损的用户的其它类型的标签。多形态的元数据标签或不可视的标签可以以设备命令的形式发回远程视觉受损人士的设备。因此,视觉受损人士可以基于不可视的反馈导航通过他们自身的个人化环境。

[0014] 本发明主题的各个对象、特征、方面和优点将从以下对优选实施例的详细描述并连同附图而变得显而易见,在附图中类似的数字代表类似的部件。

附图说明

[0015] 图 1 示出根据本发明的一个实施例的用于提供实时的对象识别和增强的视力的系统;

[0016] 图 2 示出根据本发明的另一个实施例的用于提供实时的对象识别和增强的视力的系统;

[0017] 图 3 示出根据本发明的一个实施例的在其上捕获了由图 1 中的识别应用 124 产生的示例显示的可视框(visual frame);

[0018] 图 4 示出根据本发明的一个实施例的用于提供实时的对象识别和增强的视力的系统的功能概览;以及

[0019] 图 5 是根据本发明的一个实施例的表示用于提供实时的对象识别和增强的视力的方法的流程图。

[0020] 图 6 示出能够向视觉受损人士提供增强的视力的另一个系统。

[0021] 图 7 示出配置为展现远程环境的扩增的现实全景的可能的能看见的助理界面。

具体实施方式

[0022] 应当注意,尽管以下的描述被引到基于计算机/服务器的扩增的现实引擎(engine),但认为各种可选的配置也是适合的,且可以采用包括服务器、接口、系统、数据库、引擎、适配器、控制器或其它类型的计算设备的单独或集体操作的各种计算设备。应该理解,计算设备包括配置为执行存储在有形的、非临时性计算机可读存储介质(例如,硬盘驱动器、固态驱动器、RAM、闪存、ROM 等)上的软件指令的处理器。软件指令优选地配置计算设备来提供如下面关于公开的装置所讨论的作用、职责或其它功能。在特别优选的实施例中,各种服务器、系统、数据库或接口使用标准化的协议或算法——可能基于 HTTP、HTTPS、AES、公共-私有密钥交换、万维网服务 API、已知的金融业务协议或其它的电子信息交换方法,来交换数据。优选地,在分组交换网络、因特网、LAN、WAN、VPN 或其它类型的分组交换网络上进行数据交换。

[0023] 应该理解,公开的技术提供了很多有益的技术效果,包括配置远程感测设备以辅助视觉受损人士。例如,远程能看见的助理可以使用扩增的现实引擎以向远程的视觉受损

人士的智能电话发出设备命令。该命令可以指示电话对于视觉受损人士提供指导。

[0024] 如这里所使用的,除非上下文另外地指出,术语“耦接到”旨在包括直接耦接(其中相互耦接的两个元件相互接触)和间接耦接(其中至少一个额外的元件位于两个元件之间)两者。因此,术语“耦接到”和“与...耦接”作为同义使用。

[0025] 本申请针对提供实时的对象识别和视力增强的系统和方法。以下描述包括有关本发明的实现的具体信息。本领域的技术人员会认识到可以以不同于本申请中具体讨论的方式实现本发明。此外,没有讨论本发明的一些具体的细节以避免混淆本发明。本申请中没有描述的具体细节在本领域普通技术人员的知识中。本申请中的附图及其附随的详细描述只针对本发明的示例性实施例。为了保持简明,使用本发明的原理的本发明的其它实施例没有在本申请中具体描述并且也没有通过本附图具体图示。应当记住,除非另外说明,图中类似或对应的元件可以由类似或对应的参考数字指示。

[0026] 本发明人已经意识到提供计算机居间的对象识别的传统方法通常导致不理想的用户体验。此外,本发明人已经认识到技术中对实现作为视觉受损人士增强视力的方法的一部分的细致入微且复杂的对象识别解决方案的急切的需要。本申请公开了旨在提供灵活、强大且用户易应答的解决方案的系统和方法,所述解决方案配置为精确且方便地提供对象、面部、上下文(context)以及环境识别作为扩增视力的完整的方案的一部分。例如在一个实施例中,根据本发明概念的系统和方法可以使能(enable)捕获包括脸部特征的视觉图像并提供具有那些特征的人士的实时识别。此外,在一个实施例中,根据本发明概念的系统和方法可以使能捕获包括环境特征的视觉图像并提供位置的实时识别,以及有可能根据该位置向远程用户提供导航信息。

[0027] 由本申请公开的许多益处之一是用于通过可访问的接口向移动用户提供实时的能看见的辅助、信息以及通信的设备、系统和方法。例如在一个实施例中,可以使用连接到用户个人计算机(PC)的相机使能的便携式通信设备和/或其它集中或分布式计算和通信服务提供扩增的现实和增强的视力服务。

[0028] 例如,这里具体化的系统允许视觉受损的用户使用相机使能的便携式通信设备(诸如移动电话或个人数字助理(PDA))以例如与能看见的助理或者自动的计算机视觉引擎实时地通信,该能看见的助理或者自动的计算机视觉引擎允许它们实时检测、认出并跟踪对于低分辨率视频分析而言过远的对象、地标、特别的标签(tag)、文本、标牌(signage)或条形码。本系统的实施例以非常高的速度提供高分辨率图像识别,所述速度例如大约5至10帧每秒(fps)或更高。根据本发明的一个实施例,通过开发(tap)用户的PC的处理能力使能这样的性能。

[0029] 在一些实施例中,对象或环境识别处理的结果可以使用文本到语音或预录音的标签大声发音(enunciated)。也就是说,可以通过实时提供对象的发音的语音识别来识别出对象。可以使用利用本发明的实施例分布或访问的识别数据库进行处理和识别。那些数据库例如可以包括对大多数用户共同的缺省条目。在一些实施例中,识别数据库可以是开放的,从而使得可以在用户的PC本地的存储资源、或在中央共享的社区数据库中存储并且访问对于用户个人而言重要的额外的信息层。

[0030] 除了发音的语音识别之外,诸如超链接、图像、二维(2D)和三维(3D)图形等的图形元素可以以低延迟和高帧速率实时地附连和配准(register)到视频流中的特征。此外,

使用图形以及触觉和语音使能的触摸屏界面实现的可访问界面,当与语音识别和手势识别结合时,使能宽范围的受到挑战的用户的使用,所述用户诸如视觉受损、年老和其他具有交流和认知障碍的人士。

[0031] 专门的标签可以在系统中预先训练,允许用户简单地将标签贴到没有被该系统另外地识别出的各个对象和地标。这种贴标签的方法可以使用可通过实时的视频分析而识别的特征,诸如识别为对于各种对象识别算法最优的图像,所述对象识别算法例如尺度不变特征变换算法(SIFT)、加速鲁棒特征算法(SURF)或其它较简单的基于颜色或形状的技术。当用户将便携式通信设备相机指向一个或多个场景中的一个或多个对象时,可以由系统实时自动地分析对象以识别一个或多个对象。该系统还可配置为图形地或通过音频或触觉界面在便携式通信设备的显示器中提供有关识别的对象的链接信息或动作选项。基于对例如对象、文本、电话号码、电子邮件、SMS 联系人或街道标牌的识别而生成的链接以及展现为用户先前的输入的那些可以产生更多的用户选项,用于访问额外的信息层或用于初始化通信或基于位置的服务。

[0032] 本发明的实施例实现混合式点对点(P2P)和中央处理架构,由此使能使用超过相机使能的便携式通信设备的限制的处理、能力和存储资源,同时避免与集中主机服务实现相关联的费用。例如,通过传送分层编码的压缩视频和从高分辨率图像内捕获选择的感兴趣的区域用于在用户 PC 或 PC 群上进行处理,视频图像可以接收另外超过便携式通信设备自身的处理能力的分析的深度。以该方式,可以实时分析和识别例如对象、地标、文本、条形码和面部。

[0033] 根据本发明的一个实施例,远程存在和增强的视觉服务允许经过由便携式通信设备捕获的图像或实况视频向远程助理的显示的远程辅助和训练。此外,例如经过 PC 链接到便携式通信设备的远程助理可以使用该便携式通信设备基于到来的图像远程地拍摄高分辨率的图像。此外,使用 PC 的远程助理可以通过使用分布的处理特征和系统提供的分级和分层的编码近乎实时地生成全景,允许基于到来的图像或视频帧的配准(registering)和自动缝合(stitching)来形成逐行地较高分辨率且 α 混合的全景。

[0034] 然后这样的全景能够基于实况到来的视频的位置全景摇摄(pan tilted)和变焦,允许移动用户或其远程助理获得场景的较高分辨率视图,并由此具有增强的视力能力,例如看到更远的距离或甚至阅读那些具有正常视力的人都不能阅读的标牌,或通过访问先前的图像而在暗处都能看见。在远程助理能够是双向音频/视频通信同时例如依靠地理信息系统(GIS)具有对移动用户的实时位置信息的访问时,可使能用于远程助理的实况扩增的现实。

[0035] 个人化的训练允许移动用户使用便携式通信设备或者借助于可以远程地针对用户执行训练的能看见的助理的帮助,将对象训练为数据库的一个或多个系统。这样的数据库的社区分享以及对集中提供的数据库的访问允许在基于便携式通信设备/PC 的系统上创建本地数据库,以使能上述的实时性能。该系统的实施例可配置用于例如通用产品代码(UPC)条形码检测和识别,以访问额外的电子产品目录信息或其它用户生成的信息。该系统的实施例还可以将这样的信息与这里公开的实时识别引擎和分布的结构支持的图像和特征数据库相关联。移动用户可以通过可访问的界面拍摄感兴趣的对象的图像,然后该移动用户或其远程的能看见的助理可以手动定位条形码和/或使用早期视觉特征,以捕获高

分辨率的照片。例如包括 UPC 条形码的区域的感兴趣的区域然后可以被传送到 PC 基站,在 PC 基站该条形码就这样被识别,且在本地或者通过网络访问 UPC 数据库。

[0036] 如上所述,可以通过将合适地配置的相机使能的便携式通信设备连接到用户的或社会上的联网的 PC 而高效且可靠地获得支持实时检测、识别、跟踪以及图形重叠元素(例如超链接)的使用所需的高帧速率视频和快速高分辨率照片分析所需的处理资源。这样的连接可以通过授权这样的资源共享的便携式通信设备、PC 或 PC 群经由基于网际协议(IP)的覆盖网络支持处理提供,其中这样的分布式架构由冗余的集中主机服务的服务器支持。由便携式通信设备和移动互联网设备(MID)或上网本(netbook)共享的本地服务器使能的设备-设备网络可以例如导致针对所描述的分布的实况视频和高分辨率图像识别处理的单独的系统。

[0037] 在一个实施例中,在便携式通信设备和 PC 之间分布的早期视觉系统可以用来分析低分辨率视频和用来识别对象和场景以及很可能包含诸如文本、条形码、数字、颜色等信息的区域。这个早期视觉系统的结果以及例如移动用户选择、移动用户位置以及一天中的时间可以用来有效地控制经由 IP 覆盖网络的图像捕获和传输功能。然后感兴趣的假定区域可以被传达给图像捕获控制以便识别已经接收和分析的视频或较低分辨率图像的那些区域,但是要求来自便携式通信设备的高分辨率图像传送。然后那些高分辨率图像可以按优先级排序用于传递到 PC。解码的图像和视频还可被指引到后端识别和动态数据库产生引擎。基于以上产生的队列而动态重组数据库可以对识别性能具有显著的影响。

[0038] 如上所述,可以使用诸如 SIFTs 和 SURFs 的算法进行来自便携式通信设备的实时视频的分析 and 对象识别,同时进行用于文本、条形码或其它细微特征区域的视频的同时分析。针对诸如光学字符识别(OCR)的技术,诸如文本或条形码的细微特征的检测可以被用于触发感兴趣区域的高分辨率成像的优先级排序。在一个实施例中,结合对象识别分析和特定细微特征分析的结果,从而即使不理想的 OCR 分析也可被用来增强识别性能。

[0039] 本发明人设想了一个维基(wiki)类型的社区数据库,其中用户可以提交他们的个人训练数据库。那些社区数据库可以使用 UPC 和电子产品目录号和信息来规范化,和/或用用户产生和检查过的数据库来加注解。数据库条目可以包括例如视频内容、高分辨率图像或感兴趣的区域、位置、时间信息、文本以及音频。此外,数据库条目可以包括如上所述且将在以下更详细地描述的全景,其已经用摇摄倾斜变焦(PTZ)序列加标签、注释或者叙述过。在本系统的实施例中包括配置有可访问界面的便携式通信设备使得移动用户或其能看见的远程助理可以将一个或多个数据库条目动态地直接下载到其便携式通信设备和基站 PC。

[0040] 可以对于广泛种类的具体使用实现本发明的实施例。例如,医疗保健和紧急护理界面可以通过音频发音的触摸屏及语音和手势识别界面,以及通过与紧急个人应答和远程患者监控服务的兼容性来使能。此外,本发明的实施例可以配置为检测 3G 或 4G 移动网络数据质量以及在可靠的流和潜在有损耗的实时数据传输之间切换。此外,本发明的实施例可配置为通过添加到已有的 PDA、智能电话或 MID 的外围设备,使用多个网络连接(诸如 3G 和/或 4G 及以上、电路交换的全球移动通信系统(GSM)以及公共交换电话网络标准),用于诸如远距离医学和或远程帮助的重要应用中增加的冗余度和可靠性。在一些实施例中,视频内容的使用作为取景器以及远程高分辨率图像捕获(例如,远程摄影师功能),对于诸如药物摄入监控和或远程检察和现场修理的远程存在功能可以特别有价值。可替换地,在和

由远程人事部门使用的多媒体基于位置的服务(LBS)考勤卡系统一起使用时,以安全的方式记录音频/视频内容以及位置数据可以附加价值。

[0041] 图1示出根据本发明的一个实施例的用于提供实时对象识别和增强的视力的系统100。如图1所示,系统100包括便携式通信设备110、客户端-服务器主机设备120和中央主机服务器130。注意,示出便携式通信设备110、客户端-服务器主机设备120和中央主机服务器130的组合是为了概念清楚而决不是意图作为限制。例如,在一个实施例中,系统100可包括便携式设备110和一个或多个客户端-服务器主机设备120,但是省略中央主机服务器130,而在另一个实施例中,系统100可包括便携式设备110和中央主机服务器130,但省略客户端-服务器主机设备120。此外,如现有技术所公知的,便携式通信设备110、客户端-服务器主机设备120和中央主机服务器130中的每个的操作由位于执行存储在每个设备或服务器中的存储器(未示出)中的软件的每个设备或服务器中的微处理器或中央处理单元(未示出)执行。

[0042] 如图1所示,客户端-服务器主机设备120和中央主机服务器130远离便携式设备110。配备有相机112的便携式设备110配置为使用相应的网络通信链路102和104访问客户端服务器主机和/或中央主机服务器130。此外,在存在客户端-服务器主机设备120和中央主机服务器130两者的实施例中,网络通信链路106可以在那些系统元件之间作为通信媒介。在本发明的各个实施例中,网络通信链路102、104和106可对应于同样的或不同的网络上的通信。例如,网络通信链路102可对应于在诸如Wi-Fi或WiMax的局部网络上的通信,网络通信链路104可对应于支持数据频带通信的无线移动载波网络,并且网络通信链路106可对应于在诸如因特网的分组网络上的数据传输。

[0043] 便携式设备110被示出为包括相机112和识别应用114的移动版本。便携式设备110可包括装备有例如智能电话或个人数字助理(PDA)的相机。相机112可包括摄像机和/或照相机,并且能够捕获高分辨率的视频和/或静态图像。根据图1的实施例,识别应用114的移动版本能够例如在HTC Windows移动智能电话上运行,或在其它合适地配置的移动通信设备上运行。这样的设备由在美国和国际上运营的基本所有无线通信提供商提供。识别应用114的无线版本可配置为提供可访问的界面,其使得移动用户能够为了通信和训练的目的而连接到远程助理,或连接到移动用户的基站客户端-服务器主机设备120(诸如移动用户家中的PC),其通常执行训练和实时对象识别。客户端-服务器主机设备120也可以具有可访问的界面,其允许移动用户和远程助理两者容易地管理该移动用户的训练和数据库管理的需要。

[0044] 如图1中进一步示出的,根据本实施例,客户端-服务器主机设备120包括识别应用124和识别数据库126。虽然识别数据库126被示出为驻留在客户端-服务器主机设备120上,但是并非必需是这种情况。相反地,本发明概念可以通过系统100的变型来实现,在该系统100中对应于识别数据库126的识别数据库仅可由客户端-服务器主机设备120访问,诸如可由客户端-服务器主机设备经由网络通信链路106访问的识别数据库136,或通过可由客户端-服务器主机设备120经由P2P网络连接(图1中也未示出)访问的另一个客户端-服务器主机设备(图1中未示出)上驻留的识别数据库126。驻留在客户端-服务器主机设备120上的识别应用114配置为使用识别数据库126和136的两者或其中之一,以向便携式通信设备110的远程移动用户提供对使用便携式通信设备捕获的视觉图像的实

时对象识别。

[0045] 中央主机服务器 130 被示出为包括识别应用 134 的服务器版本、识别数据库 136 以及其它特征。图 1 中还表示出由中央主机服务器 130 提供的服务,并且可包括在客户端-服务器主机设备 120 和识别应用 114 的移动版之间建立通信的安全装置。公开的网络架构的 P2P 传输消除了对繁重带宽主机服务的需要,并且因此允许经济的客户端-服务器配置用于系统 100 的用户。尽管如此,发明人也设想了使用中央主机服务器,诸如中央主机服务器 130,用于在用户自己的个人基站计算机(例如运行识别应用 124 的计算机,诸如客户端-服务器主机设备 120)不可用时访问。

[0046] 识别应用 134 的中央主机服务器版本还可配置为在例如客户端-服务器主机设备 120 上操作,这使能基本自含(self contained)的配置,在该配置中便携式通信设备 110 可连接到运行识别应用 134 的服务器版本和识别应用 124 的上网本或 PC。这可以通过便携式通信设备和 PC 或上网本之间的 P2P Wi-Fi 连接完成。这样的配置允许用户解决移动载波数据频带覆盖问题或用于成本节约目的,但是将需要移动用户携带两个系统。在家庭设置中,这可以是非常实际的系统设置。

[0047] 系统 100 还具体化了个人云计算模型的前提,由此,移动用户开发作为远程识别或训练引擎的他们自己的上网本或 PC 的处理能力。根据本实施例,用户可以选择他们自己的 PC,或家庭成员或者朋友的 PC,作为客户端-服务器,同时还具有对由中央主机服务器 130 提供的集中主机服务选项的访问,中央主机服务器 130 的使用可能与一些对用户附加的财务费用相关。所描述的配置导致提供高用户可用性的灵活的系统。

[0048] 如图 2 所示,本发明的一些实施例使用混合分布式/集中式计算平台,诸如专有的 iVisit™ 网络架构,以使能图 1 的便携式通信设备 110 和客户端-服务器主机设备 120 和/或中央主机服务器 130 之间的交互。图 2 示出根据本发明的一个实施例的系统 200,包括与客户端-服务器主机设备 220 和/或与中央主机服务器 230 通信的便携式通信设备 210a 和 210b。注意,便携式通信设备 210a 和 210b、客户端-服务器主机设备 220 和中央主机服务器 230 分别对应于图 1 中的便携式通信设备 110、客户端-服务器主机设备 120 和中央主机服务器 130。

[0049] 如图 2 所示,在一个实例中,诸如便携式通信设备 210a 的便携式通信设备可与多个客户端-服务器主机设备 220 以及中央主机服务器 230 通信。可替换地,便携式通信设备 210a 可与多个客户端-服务器主机设备 220 通信,但是从中央主机服务器 230 临时断开。在那些情况的任一种中,中央主机服务器 230 和/或多个客户端-服务器设备 220 配置为提供分布式计算平台处理例如使用便携式通信设备 210a 捕获的视觉图像。如图 2 中进一步示出的,在另一个实例中,诸如便携式通信设备 210b 的便携式通信设备可以只与中央主机服务器 230 通信。在该情况下,例如运行识别应用 134 的服务器版本并且使用识别数据库 136 的中央主机服务器 230 可配置为提供计算平台用于处理使用便携式通信设备 210b 捕获的视觉图像。

[0050] 作为对由 iVisit 提供的特征的介绍,iVisit (www.िविजिट.com) 是 IP 视频会议领域的先锋之一,且运营过第一个因特网 IP 视频会议服务之一,至今为止有 20 亿会议记录(minutes)和 4 百万下载。与商业可获得的服务一起,iVisit 目前使用其独有的 P2P 网络架构每个月免费对超过 2 千万会议记录的多方视频会议提供主机服务。iVisit 一直在开发

和运营可缩放的客户端-服务器平台并且对工作在便携式通信设备和 PC 上的 IP 网络提供主机服务。iVisit 已经注册了用机构内自建的(on-premises)服务器软件运营他们自己的服务的超过 50 个客户端-服务器被许可方以及超过 80 万用户。最近,在展示全双工视频通话(即以 14fps 的 160x 120 的分辨率)的 Windows 便携式通信设备上支持配置为在本发明的实施例中的移动客户端的完成的 beta 版。这样的性能已经在不够理想的网络性能条件下、在 Ev-DO Rev. A 之上的 Windows Mobile PDA 上展示。此外,320 x 240 的 Wi-Fi 分辨率可以支持全屏模式用于 PDA 电话上的实况视频会议。

[0051] 再次参考图 1,客户端-服务器主机设备 120 和便携式通信设备 110 可配置为提供集成的多方音频/视频会议、演示和桌面共享,基于位置的服务,按键通话,消息传送和更多。配置了 Windows Mobile、iPhone、PC 和 Mac 的 beta 版的客户端正在展示出用于实况和存储及转发通信的较高性能,同时避免对无线网络或设备的不利影响。在较高性能的便携式通信设备上能够支持 24 小时的待机时间和超过两小时的激活的(active)视频通话。混合的集中/分布式介质传输、自适应带宽管理、分层多比特率视频和音频编码的实现还使得本发明的实施例为针对 3G 和 4G 应用服务及以上的有希望的平台。简而言之,本方法的可缩放性允许这里公开的实时对象识别和视力增强功能变得可实现。本发明人相信这里公开的发明提供了目前能够在固定移动平台上以接近 10fps 的帧速率进行实时对象识别的唯一方案。

[0052] 除了支持固定的带宽之外,本发明的实施例支持在 Wi-Fi、WiMAX 和任何 3G 移动网络上的实况视频。对 WiMax 的支持将允许到达乡村的和 service 匮乏的地区,因为这是在乡村设置中理想的宽带方案显示采用。如果可获得允许在 10kbps 的数量级的数据速率的合理一致连通性,则在 1X 网络上也可发生异步的视频消息传送。更高的带宽可以用来提高视频质量直到达到便携式通信设备的处理限制,或用来支持使用公开的分布式介质传输的增加了的用户数目。本发明的实施例还考虑到相当大数目的用户可能在 PC 客户端上,并且将允许基于他们的处理能力和支持高带宽的能力,在这些 PC 用户之间的通信具有较高的质量。例如,可能只在具有 ATOM/SnapDragon、Core Duo 或类似的处理器的 PC、上网本或 MID 上使能大且高清晰度(HD)的视频格式。

[0053] 可以在快速和高效的用户数据报协议上(UDP)构建适合用在本发明的实施例中的音频和视频传输。根据本发明的实施例,客户端的带宽需求可以基于相机、光照、运动和视频窗口尺寸而自动调节以最优带宽使用。系统 100 的基于客户端和主机服务器的服务可以以简单的基础设施扩张扩大到允许例如从好几万并发用户到几百万并发用户。此外,在本系统的实施例上发生的通信可以被加密且与健康保险便利和责任法案(HIPPA)相适应。只有授权访问专用域的用户可以和在该域基站服务器上的其它用户通信。在一些实施例中,验证使用高级加密标准(AES)加密、基于类似 Kerberos 的票据(ticket)。

[0054] 由本发明的实施例使能的多方视频通话例如允许针对语音和听力受损人士的咨询(consultations)、翻译或解释服务。用户选项菜单下简单的“添加到通话”按键可以允许用户邀请额外的用户参加通话。“参加会议”功能起到类似于呼入(call-in)会议桥的作用。这些 P2P 视频会议服务可以在产品发布时作为对用户的额外的向上销售,或者作为基本封装的一部分得到支持。

[0055] 本发明的实施例允许用户在通话和消息传送期间“快照和发送”高分辨率的照片。

在这些图像的传输期间,视频传输会受到影响。然而,“快照和发送”功能使能可以在通话期间被发送或者可以作为带有音频/视频注释的照片在消息和记录期间被发送的非常高质量的图像的分发。此外,本发明的一些实施例包括集成的全球定位系统(GPS)和映射服务。这些选项使得能够将移动用户的 GPS 位置传送给例如授权的组、联系人列表或特定个体。实况或记录的 GPS 位置可以上传并显示在地图上,其可以进而使能各种各样的其它基于位置的服务。这样基于连接的位置的服务可以被用来使能例如“跟着我”或“朋友发现器”类型的服务或用于企业车队管理和远程劳动力管理应用。

[0056] 如上所讨论的,本发明的一些实施例包括配置用于便于例如新手(novice)、年长者或感测受损的用户(诸如视觉受损的远程用户)的易访问性的特征。这些包括添加新的可访问的触摸屏界面,新的可访问的便携式通信设备状态指示(诸如电力、网络状态、到来的通话或信息)。对于年长者和/或新手视觉受损的用户,可以预编程用于一个或多个远程能看见的助理的联系信息,从而使得系统可以响应于按下单个按键或其它输入而自动发送信息。对于更高级的用户,以用于用户的音频菜单的形式念出选项网格(grid)的方向键和可访问的触摸屏界面允许通过分级音频菜单选择来访问所有的系统功能。记录的人类声音可以在嘈杂环境中用于这些发音,具有极大的成功和对于移动用户的易理解性。

[0057] 在一些实施例中,与合适的文本到语音引擎(例如 Accapela)结合的键盘接口可以被用来使系统更加可访问。非特定语者人名拨号(SIND)和语音命令也可用来使能语音驱动接口。经由例如 Nokia 视频连通性电缆(CA-75U)或无线 LAN/UpnP 到兼容的 TV 的直接连接可以使能数字放大器应用。

[0058] 在本系统的一些实施例中使用的识别应用 124 将允许在训练过程期间通过远程助理的相当大量的控制。这样的控制包含例如以下特征:例如在训练期间选择移动相机的分辨率的远程能力、擦除检测到的特征的能力、通告有关检测到和识别出的对象的范围和方向信息的能力、以及使用一种语调快速地念出之前检测到的对象不再在视野中的能力。该系统还配置为通知移动用户关于光的条件,以及允许远程助理控制在大多数相机使能的便携式通信设备上可获得的照明特征,例如诸如发光二极管(LED)相机闪光灯或“手电筒”应用。

[0059] 由于一个或多个能看见的助理的远程训练很可能是配置系统用于由便携式通信设备 110 的视觉受损的移动用户的独立使用的手段,因此训练数据库的同步是重要的。加元标签(meta-tagging)的方法允许每个数据库条目具有图像的多个实例,每个具有用于训练的范围输入。因此,可能扩展系统的范围,并且与其它情况相比,从各个距离处的对象得到更准确的范围估计。高分辨率图像可以被用于识别在更大距离处的更小的对象。市场上现在有 5 兆像素相机便携式通信设备,如果用户可以等待传输数据,则该便携式通信设备将允许从更远距离处检测和识别对象。用于家中使用的 Wi-Fi 配置可以使这种模式更加实际。

[0060] 图 3 示出根据本发明的一个实施例的在其上捕获了由图 1 中的识别应用 124 产生的示例显示的可视框(visual frame),其包括适合由具有基本电脑技能的远程助理使用的可访问的用户界面。界面 324 可以被认为是能看见的助理界面。基于按下双触按键,移动用户可以利用音频和单向视频连接连接到远程助理,允许该远程助理充当远程摄影师来训练该系统。例如,远程助理首先可以使用在小观看窗口 322 中的来自便携式通信设备 110

的实况视频,以音频命令指导便携式通信设备 110 的移动用户使感兴趣的对象进入到视野中。一旦感兴趣的对象在全视野(full view)中,远程助理可以按下捕获图像按键 323a 或 323b 之一以记录视频的帧,或者请求高分辨率的图像,例如捕获的视觉图像 325。然后远程助理可以在文本描述域 328 中输入对象 ID,其可以在系统处于识别模式中时念出该对象。

[0061] 本发明的一些实施例实现例如诸如 SIFTs 的对象识别算法,其与在移动机器人中使用以支持导航定位、映射以及视觉伺服的那些类似。通常,对象识别算法在被应用到平面的、有纹理的对象时表现最佳。该算法还可靠地识别由平坦的、有纹理的结构构成的三维对象,或由略微弯曲的成分构成的那些。对象识别算法的优点在于其可以提供在光照条件、遮蔽(occlusions)和方向剧烈变化的现实环境中的可靠识别。然而,诸如人脸的三维可变形对象通常并不是以健壮(robust)的方式来处理。

[0062] 很多对象识别算法能够在一对一验证场景和在小数据库中实现高识别性能。这样的一对一场景还更容易地允许多个图像被用于匹配和作为训练集,允许系统相对于对象的光照和方向的变化变得更加健壮。然而,在试图识别例如受到遮蔽的对象时,当与非常大的数据库比较、且当数据库包括非常例如类似的对象(例如,无糖可乐瓶和经典可乐瓶)时,那些同样的算法会降至大约 80% 到 95% 的识别率。本发明概念包含对上述缺陷的多种补救方法。

[0063] 一种这样的补救方法包括图像处理算法的集成,诸如例如将对象识别算法与 OCR 融合。无糖可乐和经典可乐之间的区分、或建筑物或街道标牌的识别展示可能具有与其它对象非常相似的图案,或者对于对象识别算法可能不表现出足够的区别特征的对象。很多忍受类似性问题或缺乏足够的区别特征的对象将包括文本。因此,以上描述的很多挑战可以使用各种融合策略来解决。

[0064] OCR 典型地要求横过要读取的文本的大约 3 兆像素图像或大约 100-200dpi。本系统可以配置为用于实况视频模式以支持对象识别,并且自动切换到 3 兆像素图像捕获模式以获取用于 OCR 的所要求的分辨率。可以基于例如基于估计的对象识别的范围测量来调整便携式通信设备相机的数字变焦。

[0065] 除了使用 OCR 用于读取各种对象标签上的文本,可以使用 UPC 条形码作为另一种验证手段。可以在实况视频分析模式时检测出产品上的条形码,然后系统可以切换到静态图像模式以捕获较高分辨率的静态图像。高分辨率的静态图像可以使能 UPC 条形码的读取,用于对于 UPC 数据库进行验证。此外,在一对一验证场景中,使用多个图像用于训练是可行的,由此使能横过光照条件和 / 或方向的范围的对象识别。对于某一固定建造的对象,诸如例如建筑物入口,有可能在一天的不同时间从不同的方向捕获训练图像,代表能遇到的不一样但是可重复的光照条件。还应当注意,基于焦点组交互,即使识别率低于 90% 也能对视觉受损的移动用户有重要的价值。

[0066] 人类具有大约 135x200 度的视野,但是典型的紧凑相机只具有 35x50 度的视野。因此,本发明的实施例包括从实况视频或较高分辨率的图像提供自动全景产生的特征,可以对诸如远程助理的远程客户端 - 服务器主机设备用户具有重要的价值。例如,全景特征可以被用于克服对于远程助理的隧道视觉问题,以及用于克服对于视觉受损的移动用户的由于相机指向(pointing)问题造成的标识检测和识别的图像裁剪问题。

[0067] 用于产生全景的一种方案使用自动 - 缝合(stitch)技术,该技术目前通过使用普

通的相机捕获大量图像并且将各个图像缝合到一起形成具有大得多的视野(例如高达 360 度的视野)的合成图像而起作用。自动缝合的结果可以被视作允许移动用户或远程助理从任何方向看的计算机图形模型。可替换地,图像可以映射到球体或圆柱形的表面以给出 360 度照片。

[0068] 在本发明的一些实施例中,客户端-服务器主机设备识别应用 124 配置为以例如大约 5fps、大约 8fps 或大约 10fps 的速率接收无损视频帧。那些视频帧可以预期是适当次序的。因此,可以实现配置用来进行图像配准(imageregistration)的自动-缝合软件模型,以按实况图像序列被接收的样子处理所述实况图像序列。至少一个自动-缝合实现允许离散的场景改变为被识别为新的全景。本发明的实施例中包括的客户端-服务器主机设备识别应用 124 的一些版本可以配置为当检测到这样的场景改变时自动初始化新的全景。此外,可以允许远程助理例如手动地或基于移动用户的位置在全景之间来回导航。

[0069] 图像的 SIFT 配准比多-带混合耗时少。因此,在一些实施例中,首先显示未混合的结果和不完整的全景、然后当额外的图像到达时对它们进行填充被证明是有利的。例如,可通过 SIFT 配准模型分析实况视频数据,提供可以在系统认为便携式通信设备相机所指向的全景部分之上显示的图形重叠。基于来自能看见的助理的远程请求、或响应于识别感兴趣的区域的位置的自动图像剖析,本系统可以拍摄全景的任何部分的较高分辨率的图像。可以使用分层编码以允许对于感兴趣的区域发送较高分辨率图像的额外的层,减少系统的数据传输简档(profile)。

[0070] 图像分析算法已经在使用 SIFT 方法可靠地识别位置具体特征和确定基于图像的地标的方向和距离中示出巨大的希望。SIFT 方法可以用于使局部不变的特征相关,以在任意的旋转、缩放、亮度和对比度改变及其它变换下有效地匹配杂乱图像的小的部分。图像典型地分割为小的重叠部分,每个重叠部分以对可能的变换不变的方式描述。然后,每个部分可以被单独地匹配,且匹配片重新集合。可以以小于 1 秒执行描述的处理序列,即使在将图像与大数据库匹配时。

[0071] SIFT 算法还可应用于室内设置中移动用户的全球定位问题。例如,可以随时间在数据库中捕获 SIFT 地标,导致产生环境的 3D 地图,允许系统使用 3D 地标用于定位。可以使用测距法来减少对 3D 全球定位结果的搜索,但是系统完全能够独自找到基于位置的 SIFT 特征。关于本系统的实施例,发明人设想视觉受损的移动用户可能需要作为验证手段来沿着他们意欲的路线访问地标,或者他们可能需要识别和定位终点目的地。可以使用非常稀疏的地图来满足那些需要。此外,通过实现诸如例如路线限制、大致 GPS 位置的估计、或可从无线网络获得的基于位置服务(LBS)的成流线型技术,可以减少相关地标数据库的搜索空间,消除对里程计的需要。

[0072] 此外,2D 地图对某些视觉受损的移动用户应用可以是足够的。例如,地标可以用允许 2D 地图和定位的 GIS 数据来贴标签。可替换地,这样的地标可以与 GIS 数据库不相关,而是相对于路线地标和目的地的 1D 表示报告行进。SIFT 地标即使没有用 GIS 数据来贴标签,也可以提供例如对于到街区上的房屋的终端导引有用的相关范围和方向信息。SIFT 特征相对于噪音和遮蔽最为健壮,但是也有特别容易受到光照、反射的变化莫测的影响的地标,或否则缺乏足够的 SIFT 特征而不独特的地标。在那些实例中,可以使用文本标牌检测和 OCR 来扩增地标或场景识别。

[0073] 如之前提到的,文本和标牌识别是本发明的一些实施例的特征。然而,对于视觉受损的移动用户,不加区别的自动的文本或标牌识别产生了信息过载的重大风险。与可以阻挡不相关的视觉图像以集中在与他们的需要或兴趣一致的文本或标牌上的能看见的用户不同,视觉受损的用户不能进行这样的初始区分。因此,本系统的实施例包括对标牌的“根据请求的(on demand)”识别以及某些导航帮助和安全特征,例如诸如信号灯和人行道激活按键。

[0074] 由于环境条件的相当大的可变性以及“根据请求的”文本自身的特征的可变性,因此在自然设置中自动检测并且读取“根据请求的”特征在传统上是很困难的。本系统的实施例包括使用强分类符的可调级联实现的“根据请求的”识别功能,并且用视频流实时操作,且在高分辨率静态图像上具有非常高的准确度。例如,实时分析 160x120, 320x240 或 640x480 的视频流的客户端-服务器主机设备识别应用 124 可以将可能的感兴趣的文本区域扩展到超出初始检测到的区域,然后初始化 1 兆像素或更高分辨率的静止捕获和传输,其可被用于 OCR 或其它增强的检测和识别算法。

[0075] 在其中感兴趣的文本区域扩展超出图像的边界的情况下,检测器应该向移动用户提供方向性扫描指导,允许另一个高分辨率图像被捕获和传输到客户端-服务器主机设备识别应用 124 且缝合到相邻区域的图像,由此克服无意地裁剪的图像的问题。这种方法的一个目的在于使能对于视觉受损的移动用户的自动系统辅助,以帮助确保扫描适当的感兴趣区域并且没有向 OCR 引擎提交不完整的(例如,过度裁剪的)文本区域。为此,可以将客户端-服务器主机设备识别应用 124 配置为确定感兴趣的文本标识的方向,并为移动用户提供指导以改变他们相对于标识或文本区域的位置。

[0076] 此外,客户端-服务器主机设备识别应用 124 可以确定成功识别文本区域、远处的地标或标识所需的最低分辨率图像。可以以分层的方式编码这样的感兴趣的区域,以允许通过例如基于客户端-服务器主机设备的远程助理的更快的递增的分辨率增强。随后可以将这些感兴趣的区域二进制化并传递到准备好的 OCR 引擎(诸如例如 ABBYY 商业引擎)。可以分析结果以确保他们构成了完整的单词或词组,并且没有意义的结果可以这样被识别出并被丢弃。

[0077] 注意,将需要训练和测试数据用于“根据请求的”文本和标牌识别。该训练数据可以包括收集的代理(proxy)的数据集、模拟视频流和图像、以及示例的缝合的全景,并且可包括视频、图像和/或全景的手动标记用于期望的文本区域的识别。训练可以包括例如用于大训练集的组合的主动学习,并且可以使用额外的界面以允许基于 PC/ 上网本的远程助理与移动用户合作来捕获并标记数据。

[0078] 在一个实施例中,可以以 C++ 来实现,对性能优化。在本发明的一些实施例中实现的识别算法通过将高分辨率兆像素图像分割为碎片并以不同的比例尺分析它们,来对高分辨率兆像素图像进行工作。因此,那些实施例被配置为处理实况视频流分辨率和较高的兆像素分辨率两者。

[0079] 以下概述本发明的系统的实施例的一些特征、性能能力和优点。本系统的实施例支持从相机使能的便携式通信设备 110 (例如,智能电话,PDA)到远程客户端-服务器主机设备 120 的 30fps 的 160x120、15fps 的 320x240 以及 5fps 的 640x480 的视频传输速率。可以以下述速率支持例如经由 Wi-Fi 的到客户端-服务器主机设备 120 的高分辨率图像传

输:假设 3.5G 移动网络上 400kbps 的上行带宽,所述速率对应于 1 兆像素以小于约 2 秒传输,并且 5 兆像素以小于约 8 秒。此外,在自动全景创建模式中,传输速率可以对应于 1 兆像素大约 36 秒,并且 5 兆像素大约 3 分钟。基于实况视频 1 兆以上像素全景的近乎实时的低分辨率全景创建可以在 3.5G 网络上以从仅仅是大约 2 秒到几分钟的范围的时间间隔来产生。从实况、低分辨率全景到 1 兆以上像素全景的转变可以作为背景操作逐渐地进行,随着时间向自动的视力增强系统或基于客户端-服务器主机设备的远程助理提供对提高的分辨率的获取(access)。

[0080] 对于自动的模式,系统可以配置为使得只有感兴趣的区域才会基于实况全景的先前的 SIFT 算法分析而被提取和传输。例如,本发明的实施例可以配置为使用 1.5GHz Pentium 处理器用于相对于数百个条目的数据集的 SIFT 分析,以获得 20fps 的 160x120、10fps 的 320x240 和 5fps 的 640x480。对于远处的对象,需要更高的分辨率,这将取决于地标的范围而使系统执行变慢。可以以大约 2 到 15fps 的检测速率和大约 1 到 2fps 的识别速率,基于实况视频图像进行文本和标牌检测。可以以大约 1 到 15fps 的速率从 PC/ 上网本向便携式通信设备发送感兴趣区域命令。可以在 Wi-Fi 上以大约 2 秒或在 3.5G 网络上以大约 36 秒进行 1 兆像素图像的捕获和传输。感兴趣的兆像素区域裁剪可以导致数据大小和传输速度的 5 到 10 倍(fold)的提高。OCR 可以以大约 1 到 2 秒进行,加上执行标识内容的文本到语音的发音所需要的时间。

[0081] 假定如上所述的系统视频传输和后端图像识别速度,接近的地标和对象的 SIFT 特征即使在低分辨率也可以被识别,且因此可以被实时识别。SIFT 可以既识别地标也确定其范围和方向。本发明的实施例中实现的训练数据集可以配置为使用多个范围和方向表示地标。然而,远处的对象会要求较高分辨率的图像且会缺少压缩伪像。因此,那些远处的对象通常在 3.5G 移动网络上传输较慢。给定移动用户遇到的典型的距离和文本大小,文本标牌的成功的 OCR 几乎总是将要求兆像素图像分辨率。客户端-服务器主机设备配置可以通过使能使用从便携式通信设备到客户端-服务器主机设备识别应用 124 的 Wi-Fi 连接来克服移动网络延迟,但是其它方法可以用于智能多-分辨率视频和场景捕获、编码和传输,以加速(speed)识别和在 3.5G 移动网络上的数据传输。

[0082] 如上所述,可以由便携式通信设备提供各种分辨率的实时视频。可以由自动视力增强系统、由基于客户端-服务器主机设备的远程助理、或由移动用户或便携式通信设备 110 分别请求实时视频。便携式通信设备的移动用户、基于客户端-服务器主机设备(例如,PC/ 上网本)的远程助理、或视力增强系统也可以启动全景创建,全景创建可以显示实况视频以及单独地使用之前描述的 SIFT 自动-缝合方法开始在全景中进行贴片(tile)。可以调整实况视频的位置/视角,使得其相对于其周围的全景保持位于中心。自动缝合 SIFT 匹配和配准阶段通常不要求大量的处理时间,且假设视频图像在从便携式通信设备接收到时可以是排序的,在客户端-服务器主机设备 120 上,全景创建可以是近乎实时的。

[0083] 混合和颜色校正会耗时较长且通常将在背景中发生。在移动用户或基于客户端-服务器主机设备的远程助理选择高分辨率(例如,兆像素)的全景时,系统可被配置为显示实况视频全景,同时指导移动用户扫描他们的周围,在该扫描期间捕获在选择的点处的移动用户或基于客户端服务器设备的助理的选择的高分辨率图像。这些高分辨率图像可以以金字塔(pyramid)方式编码,从而可以在背景中发送附加的流,由此针对基于客户端-服

服务器主机设备的远程助理使能分辨率的逐渐增强。在自动视力增强模式中,可以使用例如识别数据库 126/136 和为图像剖析训练的检测器以及用于接近地标识别的同样的低分辨率 SIFT 方法在远离便携式通信设备 110 的移动用户的客户端-服务器主机设备 120 上实时分析这些实况视频。该分析的结果可被用来确定可以以可获得的分辨率实时识别出的地标,同时文本标牌检测器还可以在感兴趣的文本区域在视野中时实时提示用户。该移动用户随后可以被文本检测器提示以调整便携式通信设备相机的目标(aim),并且视力增强系统可从被确定为最可能避免截断的标识或地标的视角拍摄高分辨率图像。

[0084] 各种感兴趣区域扩展策略可以被用于基于实况视频的分析 and 构造的全景向移动用户提供相机瞄准提示。随后该移动用户可被训练以稳定地握住便携式通信设备相机用于视力增强系统拍摄兆像素图像。这些高分辨率图像可再次以金字塔的方式编码,从而附加的流可以在背景中发送并被处理以逐渐增强分辨率用于 OCR 识别后端。在一些实施例中,可以使用较低分辨率的检测器来确定识别所识别出的感兴趣的区域所需的最可能的分辨率。然而,根据移动用户和/或基于客户端-服务器主机设备的远程助理,自动缺省设置可被覆盖,使得用户能够选择适合于他们选择的观看距离的分辨率。

[0085] 图 4 示出根据本发明的一个实施例的、所描述的视力增强系统的功能性概述。视力增强系统 400 包括由便携式通信设备 410 提供的功能、由客户端-服务器主机设备 420 提供的功能、以及那些系统节点之间的相互关系。返回参考图 1,由图 4 中的便携式通信设备 410 和客户端-服务器主机设备 420 提供的功能可以被视为分别对应于图 1 中的便携式通信设备 110 上的识别应用 114 的无线版本和客户端-服务器主机设备 120 上的识别应用 124 的操作。以上概括的方法被设计为对应于人的周边视觉和中央视觉,并且因此使能用于视觉受损的那些神经生理学(neurophysiological)系统的扩增。例如,所描述的低分辨率实况视频分析可以被视为对应于能看见的人士使用周边视觉辨认出大致形状和地标的的能力,而需要较高敏锐度的中央视觉来读取和识别具有细微的细节的文本、远处的对象或区域。在本发明的实施例中,可以在将资源交给更耗时的高分辨率识别之前使用实况视频或低分辨率全景进行感兴趣区域的检测。

[0086] 根据本发明的一些实施例,可以实现空间和时间分层的视频和静止图像编码和解码,以增强例如诸如 H263+ 视频编码解码器的视频编码解码器的性能。可以实现金字塔方式编码的逐行分辨率算法,以便最佳地支持宽带和拨号连接的不用种类的混合。在点对点(point-to-point)设置中,例如,可以连续地调整媒体传输速率,以最佳地使用点对点路径上可获得的能力。该方法允许系统传递最佳用户体验,而不会有对共享该路径的其它通信量的过分的影响。当存在多个接收者时,几乎可以肯定在任意给定的时刻每个接收者可以接收的数据量将变化。这可能是由于在较大群体会议设置中的宽带对拨号用户的不同种类混合导致的,或仅仅是由于网络拥塞和/或延迟导致的。如果发送者要将相同的数据发送给所有客户端节点,则该发送者通常必须在以下两者中做出选择:以最低能力接收节点作为目标,由此使配备较好的接收者的体验降低;或在知道到较低配备的接收者的路径将很可能过载的情况下以较高速率发送。

[0087] 为了解决这一问题,本发明的实施例配置为在多个层中编码视频,其可以被有选择地转发以产生不同速率的数据流,同时保持解码的视频的完整性。这样的分层可以通过对 H. 263+ 编码解码器的修改来获得,该修改允许从单个流中选择不同帧速率。本方法被配

置为支持在大约 6 比 1(6:1)比率(例如,如果视频以 60kbps 编码,则其可以以从 10kbps 到 60kbps 的范围的数据率向接收者提供)内变化的数据率。因此,不同的接收者可以以不同的帧速率接收具有相同的空间质量的视频。还可以实现类似的编码机制以支持不同的空间层。因此,例如,一个接收者可以以 160x120 的分辨率接收视频,而另一个接收允许 320x240 分辨率的额外的增强层。

[0088] 本发明的实施例包括新的复制和转发协议(RFP),其提供了路由能力的几个显著的发展。在通过例如因特网对中央主机服务器的并发访问是得不到的或不期望的情况下分布式处理和对 PC/ 上网本客户端-服务器资源的访问需要使能自包含的群,其中视频和图像可以例如诸如在无线 LAN 上在设备之间传输。为完成此,使用 P2P 媒体传输的自包含的客户端-服务器架构可以被有效地使用,以便(1)独立地对待每个媒体资源,使得对于群的不同成员可在网络中的不同点处产生数据复制,以及(2)使能数据复制的多个点的级联,以创建可变深度的分布树。

[0089] 利用上述的两个发展,本系统架构的一个实施例配置为使得每个媒体发源(origination)节点与复制和转发节点位于同一地点。事实上,每个客户端(例如,移动客户端或 PC/ 上网本客户端-服务器)可以对于在客户端产生的媒体充当媒体分布服务器。该方法以下述的方式支持对称的 P2P 分布模型:每当这样的复制的可替换点经由分布的路由树变得可用,数据复制的点就可以容易地从发源源节点离开。

[0090] 除了这些路由能力的改变,系统 RFP 的实施例被设计为支持在现实世界部署中必需的传输层服务。那些可以包括逐跳流控制/ 拥塞避免,以确保例如没有以比下行流节点可以成功接收更高的速率向该下行流节点发送数据。此外,系统 RFP 的实施例包括用于必须可靠地传递的数据的确认和重新发送的设施,而无需要求从发源源节点重新发送。因此,本系统 RFP 的实施例可以被看做传统应用级分布式服务器和网络级多播解决方案的混合。类似 IP 多播,公开的 RFP 可以允许创建具有任意数目的复制点的分布树。然而与 IP 多播不同的是,公开的 RFP 还可以解决在 1 对任意 N 的路由的上下文环境中难以解决的关键传输层问题。利用目前部署和测试的基本的传输能力,本发明人在这公开使用位于媒体发源客户端(例如,源节点)之外的 RFP 节点的分布方案的三个额外的实施例。

[0091] 一个这样的实施例的特征可以在于“会议服务器”模型,其中具有高速网络连接的位于中心的 RFP 节点对于星型拓扑中的会议群的所有成员提供复制。第二实施例的特征可以在于“本地代理/ 高速缓存”模型,其中位于接收者群附近的专用 RFP 节点向去往该群的所有媒体提供复制服务。例如,位于校园或在 LAN 上的 RFP 节点可以向所有本地接收者提供复制服务。第三实施例采用动态 P2P 模型的形式,其中端用户对于接收同样流的其他客户端执行复制和转发服务。

[0092] 图 5 示出根据本发明的一个实施例的、描述提供实时对象识别和增强的视力的方法的步骤的流程图 500。已经从流程图 500 略去了对本领域的普通技术人员显而易见的某些细节和特征。例如,如现有技术中已知的,一个步骤可以包括一个或多个子步骤或可以涉及专门的装备或材料。尽管流程图 500 中指示的步骤 510 到 550 足以描述本方法的一个实施例,但其它实施例可以使用与流程图 500 中所示的那些不同的步骤,或可包括更多或更少的步骤。

[0093] 结合图 1 参照流程图 500 的步骤 510,步骤 510 包括通过包括相机的便携式通信设

备捕获对象的视觉图像。可以使用包括相机 112 的便携式通信设备 110 执行步骤 510。作为具体的实施例,让我们考虑便携式通信设备 110 为包括能够捕获高分辨率图像的摄像机的智能电话。在该例中,步骤 510 可以对应于智能电话的用户使用高分辨率摄像机拍摄视频。

[0094] 流程图 500 的方法继续到步骤 520,其包括将视觉图像经由网络发送给至少一个客户端-服务器主机设备。步骤 520 可以通过便携式通信设备 110 使用网络通信链路 102 发送捕获的视觉图像到客户端-服务器主机设备 120 来执行。在一些实施例中,如例如由图 2 所示,便携式通信设备 110 可以例如经由 P2P 网络链接到多个客户端-服务器主机设备。

[0095] 现在参考图 5 中的步骤 530,流程图 500 的步骤 530 包括使用客户端-服务器主机设备上的识别应用处理视觉图像来产生对象识别数据。在一个实施例中,可由客户端-服务器主机设备 120 上的识别应用 124 执行步骤 530。此外,参照图 2,在一个实施例中,每一个都运行识别应用 124 的多个客户端-服务器主机设备 220 可充当用于步骤 530 中执行的处理的分布式计算平台。可替换地,在一个实施例中,可通过中央主机服务器 130 使用服务器版识别应用 134 来执行该处理。如之前提到的,可以使用 SIFT 算法、SURF 算法和 OCR 的一个或多个来执行步骤 530 的处理。

[0096] 前进到流程图 500 的步骤 540,步骤 540 包括通过便携式通信设备经由网络从至少一个客户端-服务器主机设备接收对象识别数据。如可从图 1 看到的,步骤 540 可对应于通过便携式通信设备 110 经由网络通信链路 102 接收由识别应用 124 产生的对象识别数据。

[0097] 在流程图 500 的步骤 550 中,对象实时地被便携式通信设备 110 的远程用户识别。使用便携式通信设备 110 执行步骤 550。在一个实施例中,便携式通信设备 110 可包括识别应用 114 的移动版本,其可包括配置为实时提供对象的发音语音识别、和/或使能由视觉受损的远程用户使用当前方法的界面。在一个实施例中,视觉图像包括环境特征并且识别对象包括位置识别。在后一个实施例中,本方法还可包括根据识别的位置向远程用户提供导航信息。

[0098] 在图 6 中,类似于参照图 1 和图 2 描述的系统,系统 600 通过与扩增的现实引擎 660 的交互向视觉受损的人士 605 提供增强的感测能力。扩增的现实引擎 660 构造包括其对象的远程环境 635 的扩增的现实全景,其经由助理界面 650 直接展现给能看见的助理。不仅使视觉受损的人士 605 与扩增的现实全景交互,而是能看见的助理也直接与扩增的现实全景交互,以便如以下讨论中所述地辅助视觉受损的人士 605。

[0099] 以下的讨论是从能看见的助理的视角来展现的。因此,视觉受损的人士 605 和远程环境 635 被认为相对于能看见的助理是远程的。术语“远程”被委婉地用来指示能看见的助理与视觉受损的人士 605 物理地分开足以使得不能实行物理帮助的距离(典型地为大于约 50 米)。例如,视觉受损的人士 605 可位于与助理界面 650 不同的大陆上。因此,能看见的助理可以与视觉受损的人士 605 地理上分隔开。

[0100] 视觉受损的人士 605 使用移动感测设备 610 经由一个或多个传感器来捕获周围环境传感器数据。如之前讨论的,移动感测设备 610 可以包括智能电话或其它类型的便携式设备。其它便携式设备的例子可包括眼镜、专用设备、或配置为经由网络 615 获取和发送数

据的其他计算设备。移动感测设备 610 包括至少一个能够获取传感器数据的传感器(优选地为相机)。移动感测设备 610 可以包括其它类型的传感器,包括加速计、扩音器、磁力计、生物传感器、条形码读取器、RFID 读取器、里程计、GPS、收音机或其它类型的传感器。在一些实施例中,传感器还可以包括一个或多个互连的生物传感器的个人区域传感器网络,所述生物传感器配置为获得有关视觉受损的人士 605 的身体健康的数据;例如血压、神经活动、心率、体温、呼吸率、灌注数据,或其它健康度量。移动感测设备 610 从其传感器中的每一个获取数据,并准备用于经由网络 615 传输到扩增的现实引擎 660 的传感器数据。

[0101] 在一些实施例中,移动感测设备 610 可以配置为经由辅助的适配器 612 而获得超过其意欲的目的的能力。辅助的适配器 612 经由一个或多个耦接接口(例如,蓝牙或 802.15x、USB、无线 USB、火线(Firewire)、PCIe、Light Peak、RS-232、Zigbee、eSATA、以太网等),有线或无线地与移动感测设备 610 耦接。适配器 612 可以包括不在移动感测设备 610 中存在的额外的传感器。例如,缺少 RFID 读取器的 iPhone™。适配器 612 可以包括配置为经由 USB 连接与 iPhone 耦接的 RFID 读取器,其中 RFID 读取器可以捕获无线 RFID 数据,例如 IDBLUE™ 读取器(见 URL www.idblue.com)可以适配为与所公开的主题一起使用。此外,适配器 612 可以提供其它类型的增强的能力,可能包括:宽带通信(例如,WiMAX、WiGIG、ad-hoc 联网、UWB、WirelessHD 等)、回声定位能力、触觉反馈致动器(例如,振动器、盲文界面等)、增强的音频输出/输入(例如立体声、过滤器等),或其它增强的能力。

[0102] 有可能与适配器 612 或其它外部传感器合作的移动感测设备 610 优选地配置为获取代表远程环境 635 的周围环境传感器数据,或更具体地,当前视野(FoV) 630。鉴于移动感测设备 610 可以配置为从其传感器连续地获取数据而不需要来自视觉受损的人士 605 的有指导的动作,传感器数据可以被认为是周围环境传感器数据。甚至可以逐个传感器地主动收集或被动收集周围环境传感器数据。主动收集的传感器数据被认为包括在视觉受损人士 605 的命令下或在能看见的助理的命令下收集的数据。被动收集的传感器数据包括简单地在传感器激活时收集的数据。例如,移动感测设备 610 可以使所有传感器在“收集”模式下,在该模式中移动感测设备 610 获得周围环境传感器数据而不考虑远程环境 635 中的情况。

[0103] 代表远程环境 635 的周围环境传感器数据还可以通过除了移动感测设备 610 或适配器 612 之外的其它数据源收集。例如,公开地可访问的实况相机源(feed)也可以被访问,其中扩增的现实引擎 660 将这样的“外部”周围环境数据集成到扩增的现实全景的构造中。系统 600 内的其它数据源可以包括政府数据库(例如地理勘察等),军队或私人车辆(例如,无人驾驶飞机、飞机、汽车等),市政传感器(例如警用无线电、实况公路相机源、消防站等)、安全机构、卫星、Google® StreetView™、或公开或私人的周围环境传感器数据的其它源。

[0104] 周围传感器数据经由网络 615 发送到扩增的现实引擎 660,其中扩增的现实引擎 660 可以包括远程设备接口 622、全景引擎 620、全景数据库 624 或对象配准模块 626。应该理解,网络 615 可以包括有线或无线的一个或多个联网基础设施,包括因特网、蜂窝网络、WAN、LAN、VPN、PAN 或其它类型的网络。远程设备接口 622 可以配置为根据各种期望的方法从移动感测设备 610 接收周围环境传感器数据。在一些实施例中,远程设备接口 622 可以包括提供移动感测设备 610 可访问的 URL 或网络服务 API 的网络服务器。在其它实施例中,远程设备接口 622 可包括支持一个或多个私有协议的专用通信链路。例如,移动感测设备

610 可以使安装的应用(例如,图 1 的识别应用 114)被配置为与作为远程设备接口 622(例如图 1 的识别应用 124 或 134)操作的补充应用直接通信。

[0105] 可能记得,公开的基础设施可以以点对点的方式工作(见图 2)。类似地,全景引擎 620 可以包括经由网络彼此耦接的多个组件。例如,对等体 620A 可以代表一个或多个在网络 615 上布置的分离的计算设备,可能包括远离能看见的助理的、或者甚至相对于视觉受损人士 605 本地或远程的计算机。此外,对等体 620A 可以或许基于由 Google、Salesforce®、Microsoft®、Amazon®或其它服务提供的基于云的系统,跨越云计算架构的一个或多个节点分布。每个对等体 620A 至少在一定程度上可提供对在捕获的周围环境传感器数据中的对象识别的支持。例如,一个或多个对象配准模块 626 可以在对等体 620A 之间分布,其中每个对等体 620A 如前所述具有不同的对象识别或配准职责。例如,通过对等体 620A 的通信路径可以构造为其中每个对等体 620A 根据金字塔编码/解码方案(见图 4)以不同的图像分辨率提供配准信息。在其它实施例中,全景引擎 620 可以代表单个计算设备,该计算设备作为向远程视觉受损的人提供服务的基于网络的服务器而操作。

[0106] 全景引擎 620 从移动感测设备 610 获得周围环境传感器数据,优选地,其中传感器数据包括与当前视野(FoV) 630 有关的信息。FoV 630 代表针对部分远程环境 635 的当前的数据收集。术语“当前”是用来传达 FoV 630 代表遭受延迟的传感器数据的基本实时的表示的概念。被认为落入所述当前概念的可接受的延迟可以小于 1 分钟,更优选地小于 10 秒,更优选地小于 1 秒,或甚至更优选地小于 0.25 秒。

[0107] 可以通过从与移动感测设备 610 相关联的传感器获得的图像数据(例如静态的、视频等)和位置数据表示 FoV 630。全景引擎 620 将可能包括方向数据的图像数据和位置数据转化为呈现的 FoV,用于经由助理界面 650 向能见的助理展现。此外,全景引擎 620 配置为构造扩增的现实全景,该扩增的现实全景至少部分基于周围环境传感器数据构造代表远程环境 635 及其关联的对象的模型。该扩增的现实全景还可以包括集成的当前 FoV 630。应该理解,如之前讨论的,全景引擎 620 还可将除了视觉数据之外其它类型的周围环境数据并入扩增的现实全景,从而扩增的现实全景可以包括多形态的数据。

[0108] 当全景引擎 620 收集或否则获取周围环境数据时,全景引擎 620 将数据缝合在一起以创建包括视觉数据、音频数据、触觉数据、动觉数据、元数据数据或其它类型的数据的多形态全景。可以使用包括识别对象特征的各种适合的技术(可能通过 SIFT 配准)将视觉数据缝合在一起以组合图像。一种可以适配为和本发明主题一起使用的用于从图像数据构造全景的适合的方法包括在 Baudish 等人于 2005 年 7 月 28 日提交的题为“Real-Time Preview for Panoramic Images”的美国专利 7424218 中描述的那些。

[0109] 当构造扩增的现实全景时,全景引擎 620 还可以将不同类型的数据缝合在一起。例如,可以从 Google StreetView 收集街道的基本图像,同时可以从自服务的订户使用的一个或多个移动感测设备 610 接收的图像数据收集街道中的对象。一些对象可以是永久的(例如,灯柱)而其它对象可以是临时的(例如左下角的报纸机)。这样的对象可以和其它数据叠加以形成扩增的现实全景。此外,周围环境传感器数据中的音频数据可以和对象绑定在一起。考虑驶过的汽车或卡车。可以通过比较由音频或图像数据(例如位置、方向、大小、形状、制造、模型等)代表的与对象关联的普通属性或特征而将来自车辆的声音绑定到车辆的图像。

[0110] 全景引擎 620 还优选地配置为识别远程环境 635 的传感器范围内的对象。可以由不同的形态或形态的组合来识别对象,所述形态包括图像数据、音频数据、无线数据(例如, Wi-Fi、RFID 等)或其它类型的数据。例如,可以或许基于在 Boncyk 等人于 2001 年 11 月 5 日提交的题为“Image Capture and Identification System and Process”的美国专利 7016532 中描述的技术,通过 SIFT 特征确定图像数据中的对象。不管如何识别对象,全景对象可以和全景数据库 624 中的其它全景参数一起存储。对象可以作为包括属性的可管理数据对象存储。属性可以包括对象特征、名字、元数据、时间戳、位置、呈现指示或其它信息。

[0111] 应该注意,扩增的现实全景可以包括时间信息,或可被认为是动态结构。全景的部分可以包括代表历史的周围环境传感器数据的历史部分。历史部分可以在收集当前周围环境传感器数据时更新。鉴于全景具有历史内容,敏锐的读者将理解,扩增的现实全景可以被时移以冻结或者甚至回放远程环境 635 中视觉受损人士 605 的经历,或回放远程环境 635 的部分。可以将扩增的现实全景作为远程环境 635 的虚拟“黑盒子”扩增的现实。

[0112] 如果期望或需要,全景引擎 620 还可以过滤构造的扩增的现实全景中的内容。在一些场景中,周围环境传感器数据可以被认为是私人的而不是公开的。私人数据代表被认为是在视觉受损人士 605、或其它采用合适的验证、授权或许可级别的授权个体的管理或控制下的数据。鉴于周围环境数据可以包括私人数据,扩增的现实全景可以包括与视觉受损人士 605 或其它授权的能看见的助理关联的个人化全景。如果需要,扩增的现实全景可以包括作为私人传感器数据的函数的过滤的部分。例如,当视觉受损人士 605 正在与 ATM 机的键盘交互时,全景引擎 620 可以过滤、抖动或否则模糊能看见的助理对键盘的可视性。

[0113] 虽然扩增的现实全景被构造用于经由助理界面 650 向能看见的助理展现,但是扩增的现实全景也可以被认为是和视觉受损人士 605 共享的扩增的现实全景。全景引擎 620 可以以指示移动感测设备 610 采取合适的动作的设备命令的形式将对象信息提供回移动感测设备 610。动作可以包括通知视觉受损人士 605 接近对象(例如,听觉或触觉反馈)、调整传感器获取速率、过滤传感器数据、或否则从全景引擎 620 控制远程移动感测设备 610 以展现扩增的现实全景信息。在这样的场景下,能看见的助理和视觉受损的人士 605 两者共享构造的扩增的现实全景。额外的动作还可以包括指示移动感测设备 610 与现实世界的对象(例如,售卖机、医疗设备、亭子、打印机等)交互。

[0114] 类似于图 3,图 7 展现了能看见的助理界面 750,通过该界面,全景引擎将构造的扩增的现实全景 770 展现给能看见的助理。在示出的例子中,助理界面 750 图示为可能运行在台式计算机上的 2D 桌面应用。尽管如此,在其它实施例中,界面 750 可以经由眼镜或其它合适的观看设备(提供 4π 球面度观看的头盔或房间)在网页内作为 3D 界面或其它期望的界面展现在移动设备上。能看见的助理使用助理界面 750 以向远程视觉受损人士提供自动或手动援助。

[0115] 虽然扩增的现实全景 770 展现为视觉受损人士目前发现自身的远程环境的平面的 2D 图像,但是应当理解,扩增的现实全景 770 代表远程环境的数字模型且包括该环境中对象的模型。扩增的现实全景 770 中展现的对象被认为是允许助理如所期望的更新对象信息的交互的对象。此外,如以上间接提到的,扩增的现实全景 770 可以被认为是环境的 3D 数字模型,其还可以包括时间分量。因此,扩增的现实全景 770 可以被认为是远程环境的 4D

模型。

[0116] 扩增的现实全景 770 代表从一个或多个周围环境数据源接收的、缝合在一起的信息。扩增的现实全景 770 可以包括当前、更新的部分以及反映之前接收的周围环境传感器数据的历史部分两者。优选地,扩增的现实全景 770 包括集成的当前 FoV 730,其可能如所指示的叠加在扩增的现实全景 770 上,其中 FoV 730 实质上对应于由远程视觉受损人士使用的移动感测设备捕获的当前的 FoV。

[0117] 如之前所述,在从远程环境获取传感器数据和展现扩增的现实全景的部分之间可以存在延迟。可以对于全景的一个或多个部分展现延迟,甚至以不同等级的粒度。在例子中,对于 FoV 730 展现延迟。然而,可以对于扩增的现实全景 770 的当前部分、扩增的现实全景 770 的历史部分、或扩增的现实全景 770 中的对象展现延迟。提供延迟信息被认为是有利的,因为能看见的助理可以理解模型的各部分有多么陈旧或最新,这在帮助远程环境中的视觉受损的人士时可以有用的。

[0118] 能看见的助理可以或许通过一个或多个全景命令 753 如所期望地与扩增的现实全景 770 交互。鉴于扩增的现实全景 770 代表比仅仅是图像要复杂的远程环境及其对象的模型,助理可以与视觉受损人士分离地在该扩增的现实全景 770 周围导航。例如,助理可以变焦放大扩增的现实全景 770 以便形象化地走在视觉受损的人士前面(假定扩增的现实全景 770 包括基于额外的传感器数据的部分),以确定道路是否没有障碍。其它全景命令 753 可以包括全景摇摄(pan)、倾斜、旋转、过滤、时移或者另外地调整扩增的现实全景 770 的展现。

[0119] 扩增的现实全景 770 包括可能如所示的以列表形式展现的一个或多个扩增的现实对象 752。扩增的现实对象可以或许通过 FoV 730 中的那些对象和 FoV 730 外的对象如所期望地分隔开。根据这样的方案来分离对象被认为是有利的,因为该组织向能看见的助理提供对于扩增的现实全景 770 中的那些对象很可能比其它对象更加时间相关的了解。FoV 730 之外的扩增的现实对象 752 (例如报纸售卖机)可能是基于旧数据的陈旧的。

[0120] 扩增的现实全景 770 还可以展现与扩增的现实对象 752 相关的额外的信息。每个扩增的现实对象 752 还可如所期望的用相关联的属性或元数据来展现。例如,标签 761 被夸饰地示出,以代表分配给对象的属性或元数据。可以如所期望的根据很多不同的方案来展现对象信息。示例对象信息包括对象名字、对象位置、对象更新的时间戳、相关特征或对象地标(例如, SIFT 配准、边缘、角落等)、元数据分配、对象尺寸、颜色或可以绑定到扩增的现实对象 752 的其它类型的信息。

[0121] 负责创建扩增的现实全景 770 的全景引擎还可以用扩增的现实对象 752 来自动填充(populate)扩增的现实全景 770 并且自动填充其对象属性。由于扩增的现实对象 752 通过他们的特征(例如 SIFT 配准地标、位置、颜色,等)来识别,因此引擎可以将属性绑定到对象。在示出的例子中,被自动填充的一种类型的属性包括从 FoV 730 的原点到各个对象的相对距离。取决于全景引擎多么近期地收集周围环境数据,某些距离可能比其它距离更精确。例如,如果由于收集的数据的低延迟, FoV730 中的扩增的现实对象 752 是动态移动的对象,则其可能具有更精确的距离,而 FoV 730 之外的对象可能有相关联的陈旧数据,其很可能导致不那么精确的距离估计。

[0122] 应该记住,人类的能看见的助理还可以手动管理扩增的现实对象 752。例如,助理

界面 750 还可以向人类的能看见的助理提供一个或多个对象命令,允许该助理管理扩增的现实对象 752。设想的对象命令可以包括添加、删除、分配元数据、更新、选择、过滤、保存或另外的管理对象。考虑一个场景,其中灯柱尚未被识别出或者还没有被自动填充在扩增的现实全景 770 中。能看见的助理可以使用鼠标指针追踪灯柱的轮廓,并且指示全景引擎创建或添加该灯柱作为对象。该助理可以进一步如所期望地向对象分配一个或多个元标签,或全景引擎还可以基于将来自助理的输入与限定的对象的特征或配准地标相关联而向对象(例如,位置、颜色、形状等)分配元标签。另一个场景可以包括能看见的助理在扩增的现实全景 770 中放置虚拟的道路点对象,其随后可以被发送到视觉受损人士的设备。该人士可以使用来自道路点对象的音频或触觉反馈(即,设备命令)以导航对应的现实世界。

[0123] FoV 730 可以集成在扩增的现实全景 770 中,作为在周围环境传感器数据中接收的移动感测设备位置信息的函数(function)。在更优选的实施例中,全景引擎将 FoV 730 叠加在扩增的现实全景 770 上,以确保能看见的助理理解远程移动感测设备当前相对于扩增的现实全景 770 中的扩增的现实对象 752 位于哪里。可以从在周围环境传感器数据中接收的对应的数据确定 FoV730 的位置和方向。在一些实施例中,该位置数据可以包括 GPS 坐标,并且方向数据可以从加速计数据或磁力计数据获得。可以通过组合多个形态的数据以校正、调整或细化其位置而细化 FoV 730 的放置。组合两个或更多个形态被认为提供更精确的放置。例如,对象特征(例如,对象地标、SIFT 配准地标等)可以与 GPS 数据组合以通过相对于对象地标的三角测量来校正 FoV730 的位置。此外,可能包括幅度、频率或相位信息的音频数据(例如,回声位置,多普勒等)可以与图像数据甚至加速计数据组合,以确定对象或 FoV730 的位置。加速计数据可以被积分(integrate)一次以确定速度,或被积分两次以确定覆盖的距离,然后和 GPS 信息或其它数据类型相结合以减少确定位置或方向时的误差。简而言之,可以基于至少两个传感器数据类型校正 FoV730 的放置。传感器数据类型可以包括 GPS 位置、图像配准数据、加速计数据、磁力计数据、无线数据(例如,Wi-Fi 定位、RFID 数据、无线三角测量等)、测距法数据、音频数据(例如,声音、回声定位等)、或其它类型的数据。应该注意,当能看见的助理导航通过扩增的现实全景 770 时,即使助理转动他们的视野,FoV 730 也可以相对于扩增的现实全景 770 保持其正确的位置或方向。因此,设想助理界面 750 也可以如图所示独立地展现 FoV 730。

[0124] 扩增的现实对象 752 可以如由叠加的标签 761 夸饰地(euphuistically)表示的在扩增的现实全景 770 中被加亮。可以通过各种方法实现加亮扩增的现实对象 752。在一些实施例中,对象包括加亮的轮廓、图标或图示扩增的现实对象 752 存在于扩增的现实全景 770 中的其它指示符。特别优选的指示符包括指示扩增的现实全景 770 的历史部分和扩增的现实全景 770 的最近的更新之间的区别的加亮显示。在远程视觉受损人士进入其中扩增的现实对象 752 可能移动或改变的动态环境时,这样的方法是有用的。例如,右下角的报纸售卖机可能是新放置在街道上的,如标签 761 的展现所示。

[0125] 如上所简述的,能看见的助理还可将元标签、标签 761 例如作为元数据经由标签命令 754 分配给在扩增的现实全景 770 中的扩增的现实对象 752。标签可以包括可视标签或不可视标签。可视标签可以被表示为图标或其它对于能看见的助理可视的指示符,并且可以提供该对象的特性的列表。还应当理解,不可视标签可以包括听觉或甚至触觉元数据,其可以用于将对象识别给远程视觉受损人士。设想的标签命令可以包括很多不同的标签管

理动作,包括创建、删除、修改、移动、保存、过滤或另外地管理对象标签。分配元标签给对象向对于具体客户端个性化扩增的现实全景 770 提供了很多的灵活性。

[0126] 当能看见的助理或视觉受损人士与共享的扩增的现实全景 770 交互时,扩增的现实对象 752 可以如上所述动态地更新。应该记得,很多扩增的现实对象 752 可以被认为是公共对象,其可以被并入到来自服务的其它客户端或订户的其它的扩增的现实全景。因此,服从隐私担忧,扩增的现实全景 770 或其扩增的现实对象 752 可以在多个使用者之间共享。此外,还可以考虑扩增的现实全景 770 包括反映扩增的现实全景 770 的共享或公共可获得或可更新方面的 Wiki 部分。示例的像 Wiki 的部分的例子可以包括注解、对象、修订或其它用户提交的扩增的现实内容。

[0127] 为了说明的目的,扩增的现实全景 770 展现如由扩增的现实全景 770 中的人的空白的脸所指示的已经过滤的部分,其中至少对于该示例,脸部数据被认为是私人数据。即使获得了私人图像数据,全景引擎也可以认出私人数据的特征(例如,脸部、安全项等)并且出于隐私担忧而移除或否则模糊该私人数据。扩增的现实全景 770 可以如期望的或如配置的包括其它类型的过滤部分。过滤部分的示例使用可以包括模糊住宅环境、模糊具体个体、模糊安全输入、或者模糊被认为私密的其他项目。还可以对其它类型的数据应用迷惑化(obfuscation),所述其它类型的数据包括音频数据、位置数据、目的地、移动速率或其它信息。全景引擎可以将对象特征(例如 SIFT 配准、音频签名等)与对象相关联。当特征满足过滤标准时,对象可在扩增的现实全景中被模糊。

[0128] 能看见的助理具有用来帮助远程视觉受损人士的多个可获得的选项。一种选项简单地包括经由智能电话与该人士建立语音通信链路。该能看见的助理可以讨论该人士的当前情况并提供口头帮助。更优选的方法包括使用助理界面 750,以发送一个或多个设备命令 751 给远程移动感测设备以控制该设备的操作。设备命令 751 可包括指导移动感测设备捕获与扩增的现实全景 770 中扩增的现实对象 752 或要识别的对象相关联的额外的对象信息的宽范围的指令。

[0129] 取决于助理的需求或期望,设备命令 751 可以采用很多不同的形式。示例命令可以包括对于视频改变数据获取帧速率、调整捕获图像的分辨率、对于音频捕获改变采样率、调整交换的数据的压缩或加密参数、交替远程感测设备和全景引擎之间的通信路径、指示设备触发一个或多个致动器(受力活塞、振动器、Braille 垫等)、发出代表对象接近的可听到的信号、或其它控制命令。因此,能看见的助理可以控制远程感测设备以增强扩增的现实全景 770 并且帮助远程视觉受损的人士。考虑其中视觉受损的人士希望购买报纸的场景。能看见的助理可以向设备发出设备命令 751,使得设备在视觉受损的人士接近报纸售卖机时振动或发出声音。

[0130] 如关于图 1 所讨论的,可以在远程感测设备或附加的(add-on)适配器中部署对象识别技术的方面。鉴于基于设备的识别应用可以接近设备驻留或甚至在设备中,因此设备命令 751 还可包括朝着识别应用定向以控制或管理该应用的操作的高级指令。例如,助理可以指示设备切换(toggle)更适于捕获有关远程环境或远程环境中的对象的额外信息的一个或多个识别算法的操作。一种算法可能更适合于 OCR,而非脸部识别或对象识别。考虑其中远程视觉受损人士与钱打交道的场景。能看见的助理可以或许基于地理位置(例如,国家)而非需要额外的处理器能力的通用对象识别算法来指示远程设备使用专用于识别汇率

的识别算法。这样的方法提供与现实世界对象的优化的交互,其中对于当前的环境,所选择的算法可以比其它算法更有效地使用。

[0131] 虽然以上的讨论提到助理作为人类操作者,但是应该理解,全景引擎或甚至助理界面 750 的部分还可作为自动的能看见的助理来操作。在远程视觉受损人士和扩增的现实全景之间的很多交互可以是自动的。如之前所讨论的,当以自动的方式操作时,助理界面 750 可以向视觉受损人士发送反馈,指示到扩增的现实全景 770 中的已知或已发现的对象的相对距离。在这样的实施例中,助理界面 750 可以作为全景引擎中的模块操作,其中该模块作为专家系统操作,该专家系统经由一个或多个应用程序接口回过头来向视觉受损人士提供建议。此外,视觉受损的人士或人类的能看见的助理可以基于视觉受损的人士通过现实世界的移动而以编程来配置自动的响应或动作,其基于取决于扩增的现实全景 770 中的对象特征(例如,SIFT 配准、音频签名、纹理等)的标准触发动作。

[0132] 因此,本应用公开了旨在提供灵活、强大和用户响应的解决方案的系统和方法,该解决方案被配置为作为集成的方法的一部分准确且方便地提供对象、脸部、上下文或环境识别,以特别在共享的扩增的现实全景中增强视力。由本申请公开的许多益处之一是用于向移动用户提供实时能看见的助理、信息或通过可访问的接口的通信的设备、系统和方法。例如,可以使用连接到用户的个人计算机(PC)和/或其它集中式或分布式计算和通信服务的相机使能或传感器使能的便携式通信设备提供扩增的现实或增强的视力服务。

[0133] 对本领域技术人员应当显而易见的是,除了那些已经描述的之外的很多更多的修改是可能,而不会偏离这里的发明概念。因此,本发明性主题除了在所附权利要求的范围内之外不被限制。此外,在解释说明书和权利要求两者时,所有术语应当按照与上下文一致的最宽的可能方式解释。特别的,术语“包括”应该被解释为以非排他的方式提到元件、组件或步骤,指示所提到的元件、组件或步骤可以存在或使用或其它没有明确涉及的元件、组件或步骤组合。在说明书权利要求提到从包括 A、B、C……和 N 的组中选择的某事物的至少一个时,该文本应该被解释为只要求来自该组的一个元素,不是 A 和 N,或 B 和 N 等。

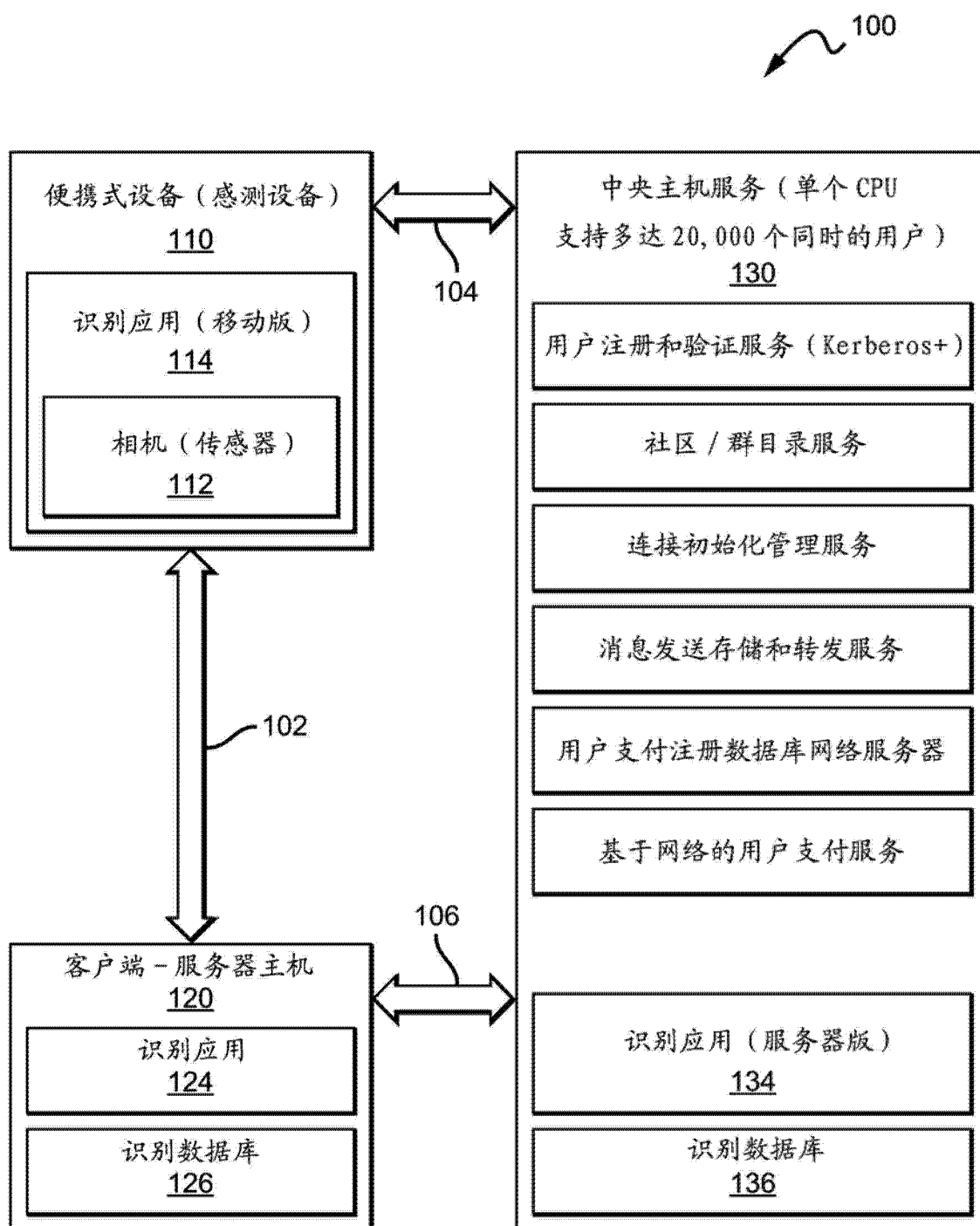


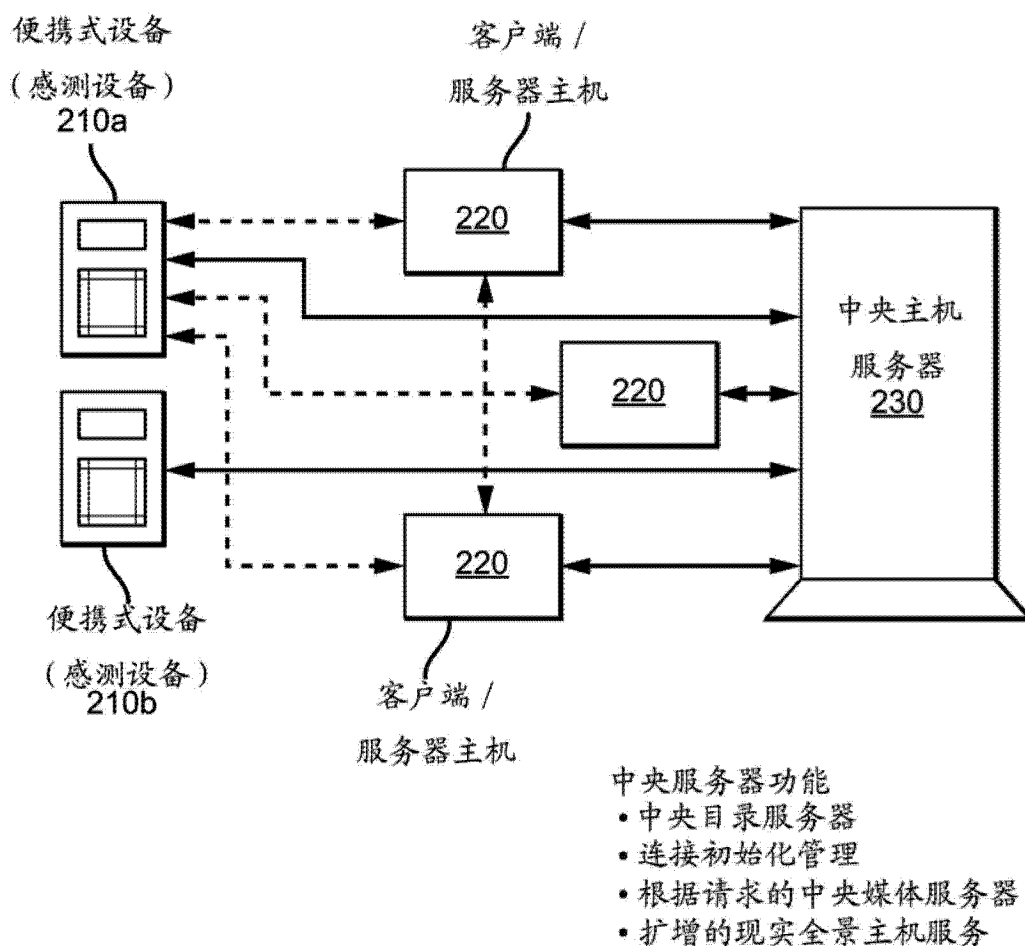
图 1

混合的分布式 / 集中式计算平台

点对点媒体服务

- 可缩放的多方通信
- 异步 P2P 文件共享
- 根据请求的中央媒体服务器
- 共享的扩增的现实全景

200



- 高度可缩放，实况多方连接
- 将宽带文件共享益处释放给私人社交网络，其中成员可以自己组织和联合控制存储；且千兆字节休眠数据传输 - 能力形成安全的私人网络用于通信、文件传输、协作、共享的扩增的现实全景等
- 低维护费用

图 2

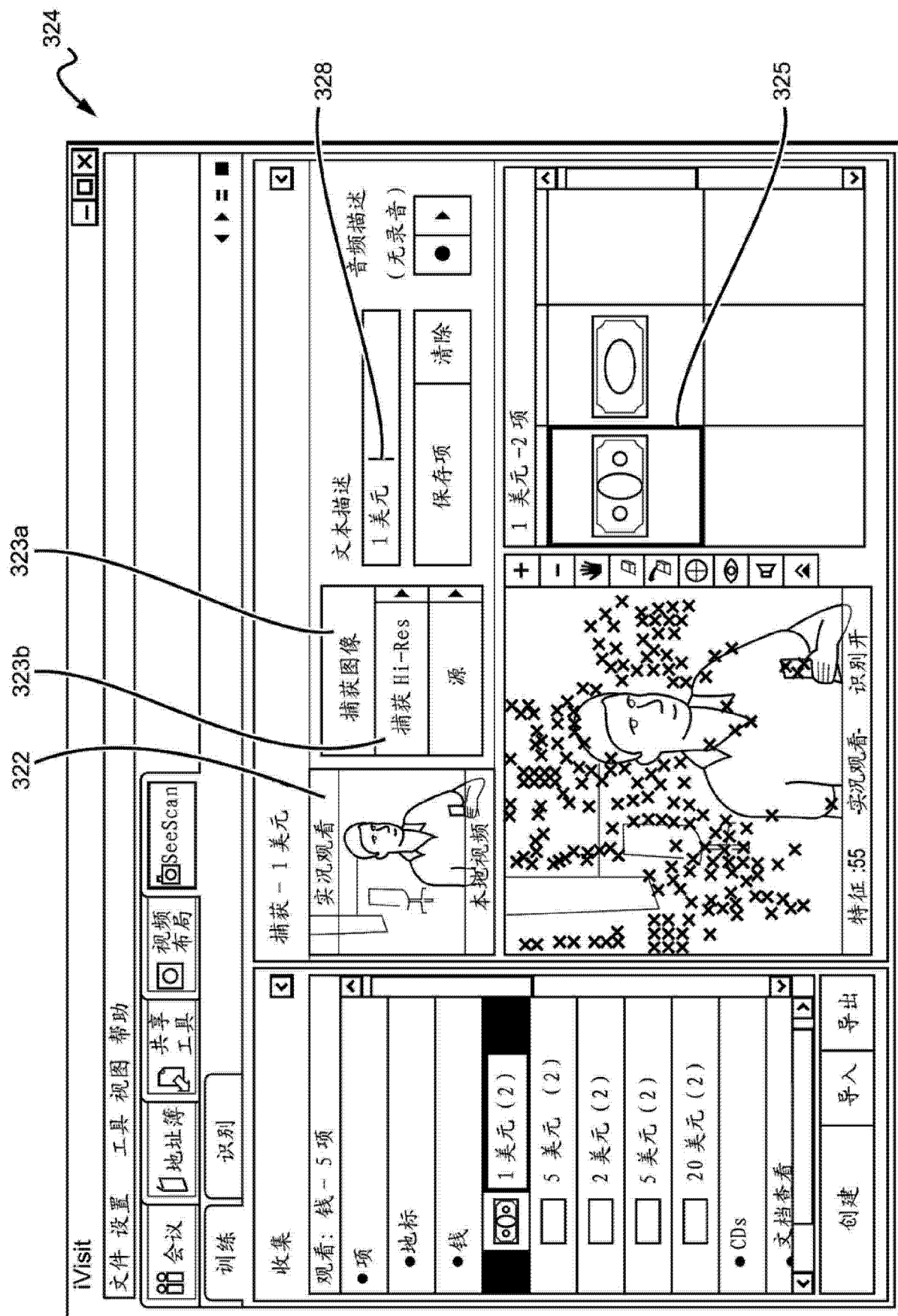


图 3

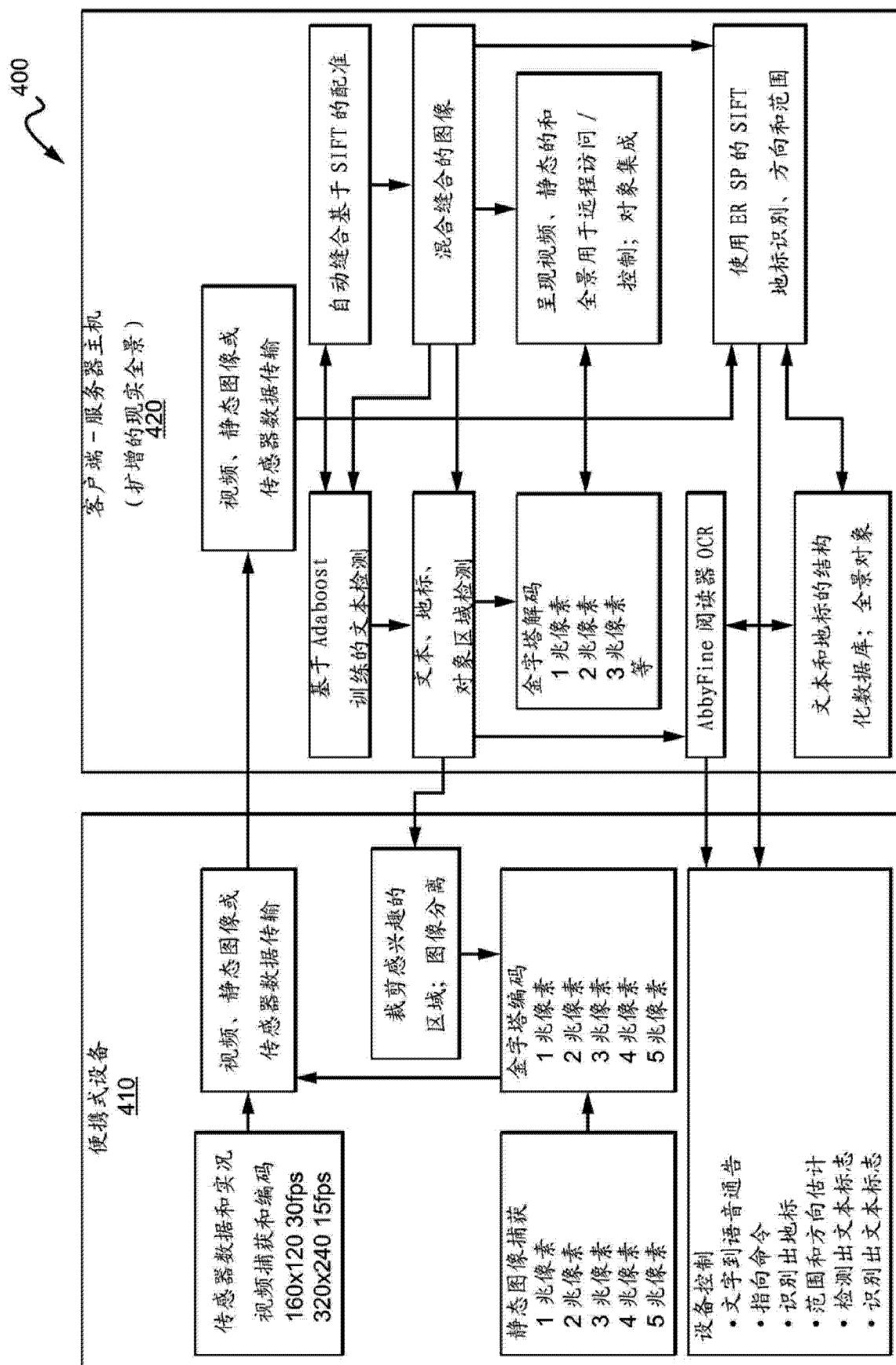


图 4

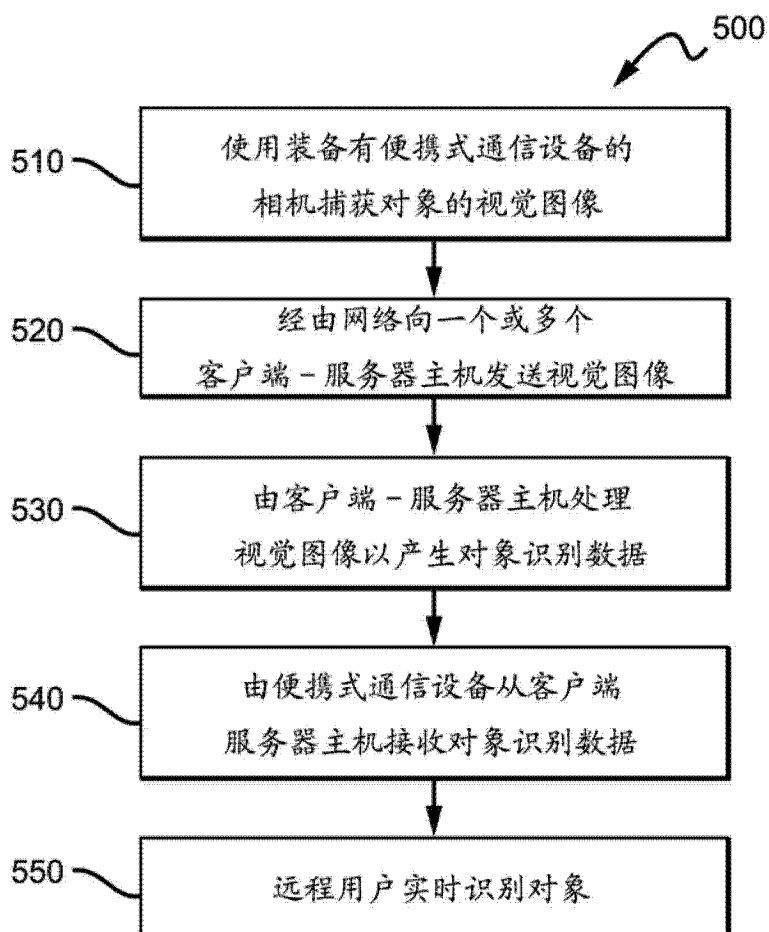


图 5

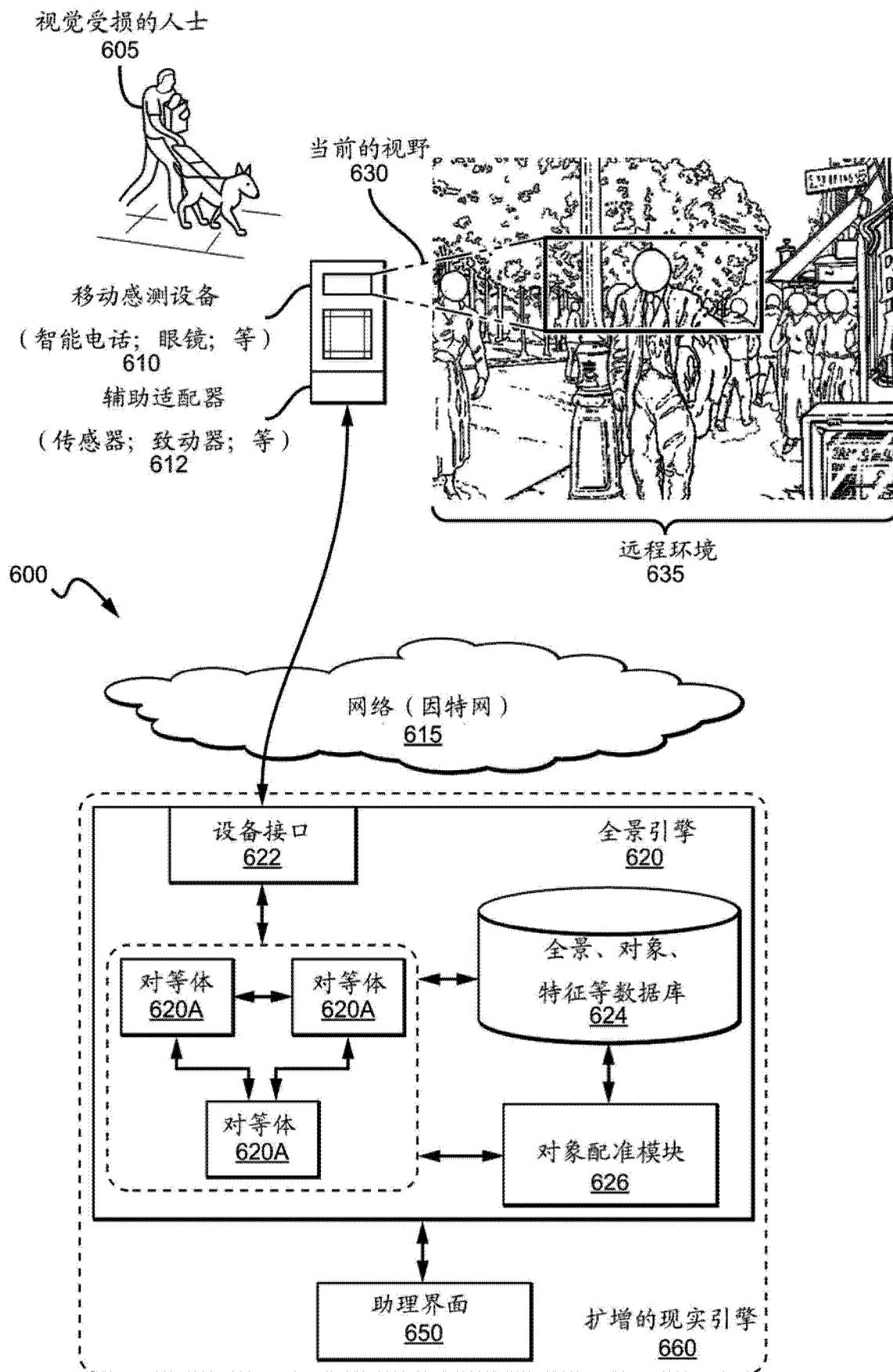


图 6



助理界面
750

图 7