(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102893327 A
(43)申请公布日 2013.01.23

(21)申请号 201180024228.9
(22)申请日 2011.03.18
(30)优先权数据
61/315,475 2010.03.19 US
61/318,217 2010.03.26 US
12/797,503 2010.06.09 US
(85)PCT申请进入国家阶段日
2012.11.15
(86)PCT申请的申请数据
PCT/US2011/029038 2011.03.18
(87)PCT申请的公布数据
WO2011/116309 EN 2011.09.22
(71)申请人 数字标记公司
地址 美国俄勒冈州
(72)发明人 G.B. 罗兹 T.F. 罗德里格斯
G.B. 肖 B.L. 戴维斯 J.V. 阿勒
W.Y. 康威尔
(74)专利代理机构 北京尚诚知识产权代理有限公司 11322
代理人 龙淳
(51)Int.Cl.
G10L 15/24(2013.01)
G10L 21/02(2013.01)

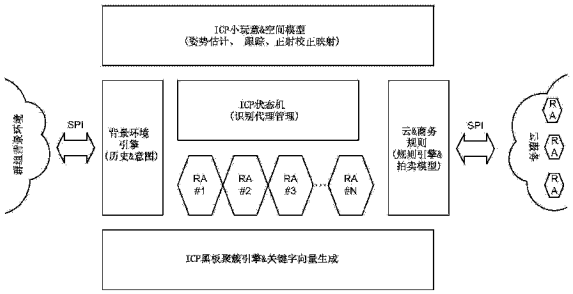
权利要求书 8 页 说明书 100 页 附图 26 页

(54)发明名称

直觉计算方法和系统

(57)摘要

智能电话感测来自用户环境的音频、图像、和/或其他刺激，并且自发地行动从而满足推断出的或预见到的用户需求。在一个方面中，所详述的技术涉及对手机的摄像机观察到的景象进行基于手机的认知。应用于所述景象的图像处理任务可以参考资源成本、资源限制、其他刺激信息(例如音频)、任务可替代性等因素从各种备选者中选择。手机可以取决于图像处理任务进行的成功程度或者基于用户对图像处理任务的明显兴趣而对所述任务应用更多或更少的资源。在一些方案中，数据可以提交给云进行分析或进行搜集。适当的装置响应的认知和识别可以由间接信息(诸如背景环境)辅助。也详述了大量其他特征和方案。



CN 102893327 A

1. 一种方法，其采用具有配置成执行所述方法的一个或更多动作的处理器的便携式用户装置，所述装置还包括接收音频的至少一个麦克风，所述方法包括以下动作：

把与所述麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与所述音频数据相对应的识别出的用户讲话数据；以及

通过参考所述识别出的用户讲话数据，推断一个或更多信号处理操作，或者推断用于信号处理操作的参数，以便与所述麦克风所接收的音频相关地被应用。

2. 如权利要求 1 所述的方法，还包括：对所述麦克风所接收的音频执行所述推断出的信号处理操作。

3. 如权利要求 2 所述的方法，还包括：与所述推断出的信号处理操作的执行相关地在所述便携式装置的屏幕上显示与所述执行有关的小玩意，其中所述小玩意的外观从第一状态改变为第二状态以指示所述信号处理的过程。

4. 如权利要求 1 所述的方法，其中所述推断动作包括：把识别出的用户讲话数据应用于数据结构，并获得与其相对应的指令或参数数据。

5. 如权利要求 1 所述的方法，其中所述信号处理操作包括音频均衡化功能。

6. 如权利要求 1 所述的方法，其中所述参数与对所述音频进行采样或再采样的采样频率有关。

7. 如权利要求 1 所述的方法，其中所述参数与关于所述音频将要查阅的远程数据库的识别有关。

8. 如权利要求 1 所述的方法，其中所述信号处理操作与将要应用于所述音频的内容识别处理有关。

9. 如权利要求 8 所述的方法，其中所述信号处理操作与将要应用于所述音频的基于水印的内容识别处理有关。

10. 如权利要求 8 所述的方法，其中所述信号处理操作与将要应用于所述音频的基于指纹的内容识别处理有关。

11. 如权利要求 1 所述的方法，其中所述识别出的讲话数据使用户环境中的对象得以识别，并且所述方法包括基于所述得以识别的对象来推断所述一个或更多信号处理操作、或者参数。

12. 如权利要求 1 所述的方法，包括：将所述音频数据应用于所述便携式用户装置中的语音识别模块。

13. 如权利要求 1 所述的方法，其中所述识别出的用户讲话数据包括以下否定词：没有、不、和忽略，

并且所述方法包括基于所述否定词来改变信号处理。

14. 如权利要求 1 所述的方法，其中所述推断动作还部分地基于背景环境信息。

15. 一种方法，其采用具有配置成执行所述方法的一个或更多动作的处理器的便携式用户装置，所述装置还至少包括用于从用户环境中分别接收第一种类型和不同的第二种类型的刺激的第一和第二传感器，所述第一传感器包括用于感测音频刺激的麦克风，所述方法包括以下动作：

把与所述麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块；

从所述语音识别模块接收与动词相对应的识别出的动词数据；

通过参考所述识别出的动词数据，确定第一种和第二种刺激类型中的哪种刺激类型是用户所感兴趣的；

从所述语音识别模块接收与用户环境中的对象相对应的识别出的名词数据；以及

通过参考所述识别出的名词数据，确定一个或更多信号处理操作或用于信号处理操作的参数，以便与所确定的类型的刺激相关地被应用。

16. 如权利要求 15 所述的方法，其中所述动词数据包括与以下动词相对应的数据：看、注视、察看、观看、和阅读。

17. 如权利要求 15 所述的方法，其中所述动词数据包括与以下动词相对应的数据：收听、和听。

18. 如权利要求 15 所述的方法，其中所述名词数据包括与以下名词相对应的数据：报纸、书、杂志、海报、文本、印刷品、票券、盒子、包裹、纸箱、包装纸、产品、条型码、水印、照片、人、男人、男孩、女人、女孩、人们、显示器、屏幕、监视器、视频、电影、电视、无线电广播、iPhone、iPad、和 Kindle。

19. 如权利要求 15 所述的方法，包括通过参考所述识别出的动词数据来确定视觉刺激是用户所感兴趣的，并且所述方法包括确定将要应用于所述视觉刺激的图像处理的类型。

20. 如权利要求 19 所述的方法，其中所述图像处理的类型包括数字水印解码。

21. 如权利要求 19 所述的方法，其中所述图像处理的类型包括图像指纹识别。

22. 如权利要求 19 所述的方法，其中所述图像处理的类型包括光学字符识别。

23. 如权利要求 19 所述的方法，其中所述图像处理的类型包括条型码读取。

24. 如权利要求 15 所述的方法，包括：

通过参考所述识别出的动词数据来确定视觉刺激是用户所感兴趣的；和

通过参考所述识别出的名词数据来确定将要把过滤功能应用于视觉刺激。

25. 如权利要求 15 所述的方法，包括：

通过参考所述识别出的动词数据来确定视觉刺激是用户所感兴趣的；和

通过参考所述识别出的名词数据来确定将要把光聚焦功能应用于视觉刺激。

26. 如权利要求 15 所述的方法，其中所述识别出的用户讲话数据包括以下否定词：没有、不、和忽略。

27. 一种方法，其采用具有配置成执行所述方法的一个或更多动作的处理器的便携式用户装置，所述装置还至少包括用于分别接收第一种类型和不同的第二种类型的刺激的第一和第二传感器，所述方法包括以下动作：

在所述装置处接收帮助识别用户环境中的用户感兴趣的对象的非触觉用户输入；以及

通过参考指示感兴趣的对象的所述输入，配置相关的传感器数据处理系统来提取与所述对象相关的信息。

28. 如权利要求 27 所述的方法，包括：接收指示用户感兴趣的所述对象的用户讲话输入。

29. 如权利要求 27 所述的方法，其中所述配置动作包括建立用于处理与相关传感器有关的数据的参数。

30. 一种方法，其采用具有配置成执行所述方法的一个或更多动作的处理器的便携式用户装置，所述装置还至少包括用于从用户环境中分别接收第一种类型和不同的第二种类

型的刺激的第一和第二传感器，所述第一传感器包括用于感测音频刺激的麦克风，所述方法包括以下动作：

把与所述麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与所述音频数据相对应的识别出的用户讲话数据；以及

通过参考所述识别出的用户讲话数据来建立参数，所述参数至少部分地定义将要与第二种类型的刺激相关地被应用的处理。

31. 一种包含非暂时性软件指令的计算机可读物理存储介质，其使用户装置处理器通过所述软件指令被编程为：

把与麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与所述音频数据相对应的识别出的用户讲话数据；以及

通过参考所述识别出的用户讲话数据来建立参数，所述参数至少部分地定义将要与第二种类型的刺激相关地被应用的处理。

32. 如权利要求 31 所述的计算机可读物理存储介质，还包括使所述处理器根据所述建立的参数来处理所述第二种类型的刺激的指令。

33. 一种智能电话装置，其具有接收音频的至少一个麦克风、并且具有配置成执行以下动作的处理器：

把与所述麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与所述音频数据相对应的识别出的用户讲话数据；

通过参考所述识别出的用户讲话数据来建立参数，所述参数至少部分地定义将要与第二种类型的刺激相关地被应用的处理；以及

根据所述建立的参数来处理所述第二种类型的刺激。

34. 一种方法，其采用具有多个传感器、一处理器、和一存储器的便携式装置，所述处理器在以传感器数据作为输入并产生输出的多个识别代理服务的执行过程中被使用，所述存储器包含黑板数据结构，所述方法包括：取决于(a)一服务在性质上是否是商用的、和 / 或(b)从与所述服务相关的外部提供商提供的信任标志是否满足标准，而向所述服务授予在所述黑板数据结构中发布、编辑或删除数据的特权。

35. 如权利要求 34 所述的方法，其中所述黑板数据结构被配置为包含多个虚拟页面的维基系统，所述多个虚拟页面之间有链接，并且不同的识别代理服务能够将数据发布到所述多个虚拟页面。

36. 一种便携式装置，其具有图像和音频传感器、处理器和存储器，所述存储器存储使所述装置执行以下动作的指令：

处理图像数据以产生对象识别数据；

处理音频数据以产生识别出的讲话数据；以及

与产生所述识别出的讲话数据的过程中的模糊话语的解析相关地使用所述对象识别数据。

37. 一种便携式装置，其具有位置和音频传感器、处理器和存储器，所述存储器存储使所述装置执行以下动作的指令：

通过参考来自所述位置传感器的数据，来获得关于装置位置的位置描述信息；

处理音频数据以产生识别出的讲话数据；以及

与产生所述识别出的讲话数据的过程中的模糊话语的解析相关地使用所述位置描述信息。

38. 一种方法，包括以下动作：

分析所接收的图像数据以确定视彩度度量或对比度度量；以及

在决定多个不同的图像识别处理中的哪一个应该应用于移动电话摄像机所捕获的图像数据、或者多个不同的图像识别处理应该以什么顺序应用于移动电话摄像机所捕获的图像数据的过程中使用所确定的度量，以便从所述移动电话向用户呈现根据不同类型的图像得到的信息。

39. 如权利要求 38 所述的方法，包括：根据所述决定动作来应用图像识别处理。

40. 如权利要求 38 所述的方法，包括：作为所述决定的结果而应用条型码读取功能、光学字符识别功能、面部识别功能、和 / 或水印解码功能。

41. 如权利要求 38 所述的方法，包括：作为所述决定的结果而应用条型码读取功能。

42. 如权利要求 38 所述的方法，包括：作为所述决定的结果而应用光学字符识别功能。

43. 如权利要求 38 所述的方法，包括：作为所述决定的结果而应用面部识别功能。

44. 如权利要求 38 所述的方法，包括：作为所述决定的结果而应用水印解码功能。

45. 如权利要求 38 所述的方法，包括：从移动电话装置的摄像机系统接收所述图像数据。

46. 如权利要求 38 所述的方法，还包括：在决定不调用所述多个图像识别处理中的哪一个的过程中使用所确定的度量。

47. 一种移动电话，其包括处理器和存储器，所述存储器包含使所述处理器执行如权利要求 38 所述的方法的非暂时性软件指令。

48. 一种存储有非暂时性软件指令的计算机可读存储介质，所述指令使移动电话处理器由此被编程为：

分析所接收的图像数据以确定颜色饱和度度量或对比度度量；以及

在决定多个不同的图像识别处理中的哪一个应该被所述移动电话调用、或者多个不同的图像识别处理应该以什么顺序被所述移动电话调用的过程中使用所确定的度量。

49. 一种方法，包括：

分析所接收的图像数据以确定颜色饱和度度量；

将所确定的度量与阈值进行比较；

如果所确定的度量低于所述阈值，则应用来自第一组处理的一个或更多识别处理；以及

如果所确定的度量高于所述阈值，则应用来自与所述第一组处理不同的第二组处理的一个或更多识别处理。

50. 如权利要求 49 所述的方法，还包括：如果所确定的度量低于所述阈值，则在应用来自所述第一组处理的一个或更多识别处理之后，应用来自所述第二组处理的识别处理。

51. 如权利要求 49 所述的方法，其中，所述第一组和第二组中的一组包括条型码读取处理，并且所述第一组和第二组中的另一组包括面部识别处理。

52. 如权利要求 49 所述的方法，其中，所述第一组和第二组中的一组包括条型码读取处理，并且所述第一组和第二组中的另一组包括对象识别处理。

53. 如权利要求 49 所述的方法，其中，所述第一组和第二组中的一组包括 OCR 处理，并且所述第一组和第二组中的另一组包括面部识别处理。

54. 如权利要求 49 所述的方法，其中，所述第一组和第二组中的一组包括 OCR 处理，并且所述第一组和第二组中的另一组包括对象识别处理。

55. 一种方法，包括：

分析第一组图像数据以计算颜色饱和度度量；

将所计算的颜色饱和度度量作为输入应用于基于规则的处理，以确定应该应用多个不同的识别处理中的哪一个或者应该以什么顺序应用多个不同的识别处理；以及

将所确定的识别处理应用于一组图像数据。

56. 如权利要求 55 所述的方法，包括：将所确定的识别处理应用于第一组图像数据。

57. 如权利要求 55 所述的方法，包括：将所确定的识别处理应用于与所述第一组图像数据不同的第二组图像数据。

58. 一种基于传感器的沿一路线的人力导航方法，包括以下动作：

确定去往目的地的路线；

使用由用户携带的电子装置中的一个或更多传感器来感测用户沿着所确定的路线的进展；以及

向用户提供反馈以帮助导航；

其中所述反馈包括由滴答声构成的样式，所述样式随着用户朝向所述目的地前进而变得更频繁。

59. 如权利要求 58 所述的方法，其中所述反馈包括振动反馈。

60. 如权利要求 58 所述的方法，包括：根据用户所面向的方向来变化所述反馈，以帮助用户确定行进的方向。

61. 如权利要求 58 所述的方法，包括：当用户处于静止状态时增大所述反馈的幅度，或者当用户在移动时减小所述反馈的幅度。

62. 如权利要求 58 所述的方法，其中所述一个或更多传感器包括磁力计，所述磁力计产生指示其方向的输出数据，其中所述磁力计能够指示由于用户以一取向携带所述装置而偏离用户所面向的方向的方向，并且其中所述方法包括对所述偏离进行补偿。

63. 一种操作处理图像数据的配备有摄像机的便携式装置的方法，所述装置由用户携带，所述方法包括以下动作：

执行初始的一组多个不同的图像处理操作；以及

无需明确的用户命令，在环境准许的限度内调用额外的图像处理操作；

其中所述装置自发地行动从而满足推断出的或预见到的用户需求。

64. 如权利要求 63 所述的方法，包括：存储或者安排存储由一个或更多所述图像处理操作产生的数据对象，并把与所述数据对象有关的语义声明发送到远程的链接数据登记库。

65. 如权利要求 63 所述的方法，包括：辨别所述图像数据所表示的场景内的一个或更多视觉特征，并在所述装置的屏幕上在与所述场景内的所述视觉特征相对应的位置处呈现视觉小玩意。

66. 如权利要求 65 所述的方法，其中所述小玩意呈非矩形形状。

67. 如权利要求 65 所述的方法，包括：感测用户在所述装置屏幕上关于一个或更多小玩意的用户手势，并基于所述用户手势来采取行动。

68. 如权利要求 67 所述的方法，其中所述动作包括以下动作中的至少一个：

(a) 将更多或更少的处理资源分配给与一小玩意相关联的功能，所述功能在感测所述用户手势之前已经被启动；

(b) 缩减与一小玩意相关联的处理，并存储与其有关的信息，使得用户偏好或行为模式能够被辨别；

(c) 在远程处理系统中继续进行相关处理的同时，至少暂时地缩减与所述装置上的一小玩意相关联的处理；

(d) 编辑图像以排除一个或更多特征；

(e) 改变呈现在所述装置屏幕上的图像数据中的一个或更多特征的投影；和

(f) 定义多个小玩意所表示的实体之间的社会关系。

69. 如权利要求 65 所述的方法，包括：以透视图方式扭曲所呈现的小玩意中的至少一个小玩意以便与在所述场景中辨别出的表面特征相对应。

70. 如权利要求 65 所述的方法，包括：当一个或更多所述图像处理操作朝向诸如识别或鉴别出所述场景中的特征之类的期望结果进展时，改变所呈现的小玩意中的一个小玩意的亮度、形状或尺寸。

71. 如权利要求 63 所述的方法，其中所述调用动作包括：基于包括以下因素中的至少一个因素的环境来调用额外的图像处理操作：

(a) 位置；

(b) 日时；

(c) 与一个或更多人的接近度；

(d) 基于所述初始的一组图像处理操作的输出；或者

(e) 用户行为的统计模型。

72. 如权利要求 63 所述的方法，包括：根据包括来自一个或更多所述图像处理操作的结果的数据来推断关于用户期望的交互类型的信息，并且基于所述信息来调用额外的图像处理操作。

73. 如权利要求 63 所述的方法，还包括：将数据发送给远程系统，使得所述远程系统能够执行一个或更多与所述装置相同的图像处理操作。

74. 如权利要求 63 所述的方法，其中所述装置自发地行动从而确定由所述装置的摄像机成像的一组硬币的价值。

75. 如权利要求 63 所述的方法，包括：基于指示一个或更多以下信息的数据，从较大的第二组可能的图像处理操作中选择将要执行的第一组额外的图像处理操作：

(a) 装置资源利用率；

(b) 与不同的可能操作相关联的资源需求；和

(c) 不同的可能操作之间的对应度。

76. 如权利要求 63 所述的方法，包括：辨别由所述图像数据表示的场景内的一个或更多视觉特征，并且把和每个这样的特征相关的数据与对应的标识符相关联地存储，其中所述标识符基于以下信息中的至少两个：

(a) 会话 ID ;

(b) 明确的对象 ID ;和

(c) 根据所述特征或根据相关环境取得的数据。

77. 如权利要求 63 所述的方法,包括 :使用所述装置中的非图像传感器系统来产生非图像信息,并且将这样的信息用于以下目的中的至少一个 :

(a) 影响对图像处理操作的选择 ;和

(b) 在关于所述图像数据的两个或更多候选结论之间消除歧义 ;

其中所述非图像传感器系统包括地理位置系统、音频传感器、温度传感器、磁场传感器、运动传感器、和嗅觉传感器中的至少一种。

78. 如权利要求 63 所述的方法,还包括 :将所述图像数据或来自一个或更多所述图像处理操作的数据中的至少某一部分传送给远程计算机系统,使得所述远程计算机系统能够继续先前由所述装置执行的图像处理,试图搜集所述装置在其处理中未能辨别出的信息。

79. 一种操作配备有磁性传感器的智能电话的方法,所述方法的特征在于,感测由零售环境中的多个电磁发射器发射的磁性信号,并且基于磁性信号向用户提供导航或产品信息。

80. 一种方法,包括以下动作 :

在操作的第一阶段,从用户的环境捕获一图像序列 ;

处理所述序列以识别所述序列中的特征并鉴别相关信息,所述处理至少部分地由用户携带的便携式装置执行 ;以及

在跟随在所述第一阶段之后的操作的第二阶段中,使用与所述便携式装置相关联的输出装置将所述相关信息呈现给用户。

81. 如权利要求 80 所述的方法,包括以下动作中的至少一个 :

(a) 识别所述序列的后面部分中的在所述序列的前面部分中无法识别出的图像特征,并且使用来自所述后面部分的所述识别结果来识别所述前面部分中的特征 ;和

(b) 对用户手势做出响应,从而向前或向后前进通过用至少一部分所述相关信息注释的所述序列的至少一部分。

82. 一种链接数据方法,其特征在于,限制用户访问关于一物理对象的声明的能力或者限制用户做出关于一物理对象的声明的能力,除非所述用户与所述对象或者与先前做出这样的声明的另一用户具有可证明的关系。

83. 如权利要求 82 所述的方法,其中所述可证明的关系是,按照所述用户携带的智能电话装置中的传感器系统所产生的数据指示出的那样,所述用户存在于所述物理对象的一定距离内。

84. 一种链接数据方法,其特征在于,基于由用户携带的传感器产生的数据检查运动信息,并且在所述运动信息指示出所述用户以限定的方式移动的情况下,限制所述用户访问与一物理对象相关的声明的能力或者限制所述用户做出与一物理对象相关的声明的能力。

85. 如权利要求 84 所述的方法,其中所述限定的方式包括以超过阈值的速度运动。

86. 一种处理装置,其包括处理器、存储器、触摸屏、位置确定模块和至少一个音频或图像传感器,所述存储器存储将所述处理器配置成在所述触摸屏上呈现用户界面的指令,所述用户界面的第一部分呈现来自所述传感器的信息,并且同时,所述用户界面的第二部分

呈现与所述装置的位置相关的信息。

87. 如权利要求 86 所述的装置，其中与所述装置的位置相关的所述信息包括描绘附近区域的地图，并且其中所述指令将所述处理器配置成在所述地图上呈现指示用户的历史动作的推针。

88. 一种处理装置，其包括处理器、存储器、屏幕和图像传感器，所述存储器存储将所述处理器配置成在所述触摸屏上呈现与由所述图像传感器感测的图像相对应的数据的指令，所述处理器还在所述触摸屏上呈现雷达扫描线的扫掠效果以指示处理图像数据的过程中的装置活动。

89. 如权利要求 88 所述的装置，其中所述指令将所述处理器配置成跟随着进行扫掠的雷达扫描线来呈现关于由所述图像传感器成像的对象的取向的标志。

90. 如权利要求 89 所述的装置，其中所述标志指示感测到的图像数据中的数字水印的取向。

91. 一种进行声源定位的方法，包括以下动作 ：

利用环境内的多个无线电话对环境音频进行采样 ；

将由第一电话感测的音频信息发送给第二电话 ；

辨别使所述第一电话的位置与第二位置相关的位置数据 ；和

在所述第二电话中，处理所述位置数据、从所述第一电话接收的音频信息、以及由所述第二电话采样的音频，以辨别相对于所述第二电话的声源方向。

92. 如权利要求 91 所述的方法，其中所述发送动作包括 ：发送尚未被窗口频域压缩变得时间模糊的信息。

直觉计算方法和系统

[0001] 相关申请数据

[0002] 在美国，本申请要求 2010 年 3 月 26 日提交的申请 61/318,217 和 2010 年 3 月 19 日提交的 61/315,475 的优先权，并且是 2010 年 6 月 9 日提交的申请 12/797,503 的继续申请。

[0003] 本说明书涉及对本受让人先前的专利和专利申请中详述的技术的扩展和改进，其中上述先前的专利和专利申请包括：专利 6,947,571；以及 2010 年 3 月 3 日提交的申请 12/716,908（公开号为 20100228632）；2010 年 2 月 24 日提交的 12/712,176；2010 年 1 月 28 日提交的 12/695,903（公开号为 20100222102）；2009 年 8 月 19 日提交的 PCT 申请 PCT/US09/54358（公开号为 WO2010022185）；2009 年 6 月 24 日提交的 12/490,980（公开号为 20100205628）；2009 年 6 月 12 日提交的 12/484,115（公开号为 20100048242）；和 2008 年 11 月 14 日提交的 12/271,772（公开号为 20100119208）。读者被假定熟悉上述公开内容。

[0004] 来自这些刚刚引用的文献的原理和教导可以应用于这里详述的方案的背景中，反之亦然。（通过引用将这些先前的专利和申请的全部公开内容结合在本文中。）

技术领域

[0005] 本说明书涉及各种技术；大部分涉及使智能电话和其他移动装置能够对用户的环境做出响应（例如通过充当直觉的视听装置）的方案。

背景技术

[0006] 手机已经从专用的通信工具发展成多功能的计算机平台。“有一个应用软件可以装”是为人们所熟悉的口头禅。

[0007] 超过二十万个应用软件可用于智能电话，从而提供种类极多的服务。然而，这些服务中的每一个都必须由用户特意地识别并启动。

[0008] 从可以回溯到二十多年前的普适计算（ubiquitous computing）的视角来看，这是极其悲哀的。在上述普适计算中，计算机需要我们更少地去关注它，而不是更多地去关注它。真正“智能”的电话应该是自主地采取行动来实现推断出的或预期到的用户期望。

[0009] 沿着这一方向向前跃进的一步将会是，为手机配备使其成为智能视听装置的技术，从而监视用户的环境并且响应于视觉和 / 或其他刺激而自动选择并采取操作。

[0010] 在实现这样的装置的过程中存在着许多挑战。这些挑战包括：理解对装置输入的刺激所表示的含义的技术，基于该理解来推断用户的期望的技术，以及在满足这些期望的过程中与用户进行交互的技术。这些挑战中可能最大的挑战是上述第一个挑战，它基本上是机器认知方面的长期存在的问题。

[0011] 考虑手机摄像机。对于每个所拍摄的帧，手机摄像机输出大约一百万个数字（像素值）。这些数字表示汽车、条形码、用户的孩子、或者一百万个其他东西之一吗？

[0012] 假定该问题具有一个直接的解决方案。将这些像素传送给“云”并使大量匿名计算机将每种已知的图像识别算法应用于该数据，直到其中一种图像识别算法最终识别出所描

绘的对象。(一种特定的方法是,将未知的图像与发布到基于万维网的公共照片储存库(如Flickr 和 Facebook)中的数十亿图像中的每一个进行比较。在找到最相似的发布照片之后,可以记录与该匹配照片相关联的描述性词语或“元数据”,并将其用作用于识别未知图像的主题的描述符。)在消耗了几天或几个月的云计算能力(和数兆瓦的电力)之后,答案得以产生。

[0013] 然而,这样的解决方案无论在时间方面还是在资源方面都是不实际的。

[0014] 稍微更实际一点的方法是将图像发布给众包(crowd-sourcing)服务,如 Amazon 的Mechanical Turk。该服务把图像提交给一个或更多人类审阅者,这一个或更多人类审阅者将描述性词语提供回给该服务,随后这些描述性词语被传送回给装置。当其他解决方案证明无效时,这是可能的替代方案,尽管时间延迟在许多情况下过长。

发明内容

[0015] 在一个方面中,本说明书涉及可以用来更好地解决认知问题的技术。在一个实施例中,应用图像处理方案来相继地获得更多且更好的关于所输入的刺激的信息。图像内容的大概意思可以在一秒钟内获得。更多信息可以在两秒钟之后获得。利用进一步的处理,更加精炼的评估可以在三或四秒钟之后获得,等等。该处理可以通过用户不需要这样的处理继续进行的(明确的、暗示的或推断的)指示而在任何一点被中断。

[0016] 如果这样的处理不能产生迅速的令人满意的结果并且用户继续对图像的主题感兴趣(或者如果用户没有相反的指示),那么可以将图像提交给云进行更加彻底且冗长的分析。书签或其它指示符可以存储在智能电话上,从而允许用户复核并了解由远程服务所做的这种进一步分析的结果。或者如果这种进一步的分析得出了可引起行动得以采取的结论,那么可以提醒用户。

[0017] 对适当的装置响应的认知和识别可以由附属信息(如背景环境)来辅助。如果智能电话从所存储的概况信息(profile information)得知用户是 35 岁的男性,并且从 GPS 数据和相关地图信息得知用户位于波特兰的星巴克咖啡店,并且从时间和天气信息得知现在是工作日的昏暗且下雪的早上,并且从装置的历史中检索出在先前几次造访该位置时用户采用手机的电子钱包购买了咖啡和报纸、并使用手机的浏览器浏览了报导橄榄球比赛结果的网站,那么智能电话的任务就得到相当大的简化。可能的输入刺激不再有无限大的范围。而是,输入的景象和声音很可能是在昏暗且下雪的早上在咖啡店中通常会遇到的那类景象和声音(或者相反地说,不可能是例如在东京的阳光充足的公园中遇到的景象和声音)。响应于这样的景象和声音的适当的可能动作也不再有无限大的范围。而是,候选动作很可能是与波特兰的在上班途中的 35 岁对橄榄球感兴趣的喝咖啡的用户相关的动作(或者相反地说,不可能是与例如东京的坐在公园中的老年妇女相关的动作)。

[0018] 通常,最重要的背景环境信息是位置。第二最相关的背景环境信息通常是动作的历史(通过以往的各星期、季节等的当前这一天来获悉)。同样重要的是关于用户的社交群体或用户的人口统计群体中的其他人在类似的情况下所做的事的信息。(如果在Macys 百货商场的特定位置驻足的最后九个十几岁的女孩都拍摄了走廊端显示器上的一双靴子的图像、并且全都对了解价格感兴趣、并且她们中的两个人还对了解存货中有哪些尺码感兴趣,那么在该位置驻足的第十个十几岁的女孩所拍摄的图像很可能也是同一双靴子的图像,并

且该用户很可能也对了解价格感兴趣，或许也对存货中有哪些尺码感兴趣。）基于这样的附属信息，智能电话可以加载适合于在统计上可能出现的刺激的识别软件，并且可以准备采取在统计上与响应相关的动作。

[0019] 在一个特定实施例中，智能电话可以具有可利用的数百个备选的软件代理，这些软件代理中的每一个都能够执行多种不同的功能，每种功能在例如响应时间、CPU 利用率、内存利用率和 / 或其他相关限制方面都有不同的“成本”。于是手机可以进行规划练习（planningexercise），例如限定出由各种可利用的代理和功能构成的 N 叉树、并沿路径航行穿过该树以辨别出如何以最低成本执行期望的操作组合。

[0020] 有时，规划练习可能无法找到适合的解决方案，或者可能会发现其成本令人望而却步。在这种情况下，手机可以决定不进行某些操作——至少在当前时刻不进行。手机可以不进行任何关于该任务的进一步处理，或者在使解决方案变得实际可行的附加信息变得可获得的情况下手机可以在过一会之后再试一次。或者，手机可以简单地将数据提交给云以便通过更有能力的云资源来处理，或者手机可以存储输入的刺激以便在之后再访问并有可能进行处理。

[0021] 系统的处理（例如，图像处理）中的很大一部分实质上可能是投机性的——是带着某处理可能在当前背景环境下有用的期待尝试的。根据本技术的另一方面，根据各种因素来对这些处理分配更多或更少的资源。一个因素是成功率。如果一个处理看上去似乎能产生积极的结果，那么可以给它分配更多资源（例如，内存、网络带宽等），并且可以允许它继续进入进一步的操作阶段。如果一个处理的结果看上去似乎是令人沮丧的，那么可以给它分配更少的资源，或者将其完全停止。另一个因素是用户对特定处理的结果感兴趣与否，这也可以类似地影响是否允许一个处理继续进行以及允许该处理用哪些资源继续进行。（用户兴趣可以例如通过用户触摸屏幕上的某一位置来表达 / 明示，或者可以根据用户的动作或背景环境来推断，例如根据用户移动摄像机从而将特定对象重新定位在图像帧的中心的动作来推断。用户兴趣的缺乏可以类似地通过用户的动作来表达，或者根据用户的动作或用户动作的缺少来推断。）另一因素是处理的结果对正被分配更多资源或更少资源的另一处理的重要性。

[0022] 一旦已经实现了认知（例如，一旦已经识别图像的主题），那么手机处理器或云资源就可以建议应该提供给用户的适当响应。如果描绘的主题是条形码，那么可以指示一个响应（例如，查找产品信息）。如果描绘的主题是家庭成员，那么可以指示不同的响应（例如，发布到在线相册上）。然而，有时，适当的响应不是立即就明显易见。如果描绘的主题是街道景象或停车计时器，那该怎么办？再一次，附属信息源（如背景环境和来自自然语言处理的信息）可以应用于该问题以帮助确定适当的响应。

[0023] 智能电话的传感器被不断地供以刺激（由麦克风感测的声音、由图像传感器感测的光、由加速计和陀螺仪感测的运动、由磁力计感测的磁场、由热敏电阻器感测的周围温度、等等）。一些刺激可能是重要的。大多数刺激是噪声并且最好被忽略。当然，手机具有各种有限的资源，例如 CPU、电池、无线带宽、金钱预算等。

[0024] 因此，在另一方面中，本技术涉及确定要处理密集的一堆数据中的哪些，并且涉及使在平台的约束下进行视觉搜索的数据处理方案与系统的其他需求平衡。

[0025] 在另一方面中，本技术涉及例如与视觉对象（或音频流）相一致地在移动装置屏幕

上呈现“小玩意(bauble)”。用户对小玩意的选择(例如通过轻拍触摸屏)导致与对象相关的体验。小玩意可以随着装置逐渐了解更多或者获得更多关于对象的信息而在明确性或尺寸方面进化。

[0026] 在早期的实现方案中,所描述的这类系统将是相对基础的,并且不会展现出较多的洞察力。然而,通过将数据细流(或洪流)(连同以这些数据为基础的关于用户动作的信息一起)馈送回给云进行存档和分析,这些初期系统可以建立借以构建模板和其他训练模型的数据基础——使这些系统的后代能够在被供以刺激时具有高度的直觉性和响应性。

[0027] 如接下来将变得明显的那样,本说明书也详述了大量的其他发明特征和组合。

[0028] 尽管主要在视觉搜索的背景环境下进行描述,但应理解的是,这里详述的原理也适用于其他背景环境(如来自其他传感器或传感器的组合的刺激的处理)。许多详述的原理具有宽得多的适用性。

[0029] 类似地,尽管下面的描述集中讨论几个示例性实施例,但应理解的是,这些发明原理不限于以这些特定形式实现。因此,例如,尽管具体提到了一些细节(如黑板数据结构、状态机构造、识别代理、延迟执行(lazy execution)、等等),但是它们中的任何一个都可能是不需要的(除非由所发布的权利要求特别指定)。

附图说明

[0030] 图 1 用架构图示出采用本技术的某些方面的实施例。

[0031] 图 2 是示出使本地装置涉及云处理的图。

[0032] 图 3 用不同的功能方面(按照系统模块和数据结构)对认知处理的各特征进行映射。

[0033] 图 4 示出空间组织和理解的不同水平。

[0034] 图 5、5A 和 6 示出可在构成服务决定的过程中使用的数据结构。

[0035] 图 7 和 8 示出根据人工智能已知的并且在本技术的某些实施例中采用的规划模型的一些方面。

[0036] 图 9 标识出可由操作系统执行的四个级别的并行处理。

[0037] 图 10 对说明性实现方案进一步详述这四个级别的处理。

[0038] 图 11 示出在辨别用户意图的过程中涉及的某些方面。

[0039] 图 12 描绘出可在某些实现方案中使用的循环处理方案。

[0040] 图 13 是图 12 方案的另一视图。

[0041] 图 14 是描绘系统操作的某些方面的概念图。

[0042] 图 15 和 16 分别示出与识别代理和资源跟踪相关的数据。

[0043] 图 17 示出可用来帮助机器理解观察空间的图形目标。

[0044] 图 18 示出基于音频的实现方案的一些方面。

[0045] 图 19 和 19A 示出各种可能的用户界面特征。

[0046] 图 20A 和 20B 示出使用经阈值处理的斑点进行对象分割的方法。

[0047] 图 21 和 22 示出其他示例性用户界面特征。

[0048] 图 23A 和 23B 示出用户界面中的雷达特征。

[0049] 图 24 用来详述其他用户界面技术。

[0050] 图 25-30 示出与传感器相关系统的声明配置方案相关联的特征。

具体实施方式

[0051] 在许多方面中，本公开的主题可以被认为是对允许用户使用计算机装置与用户的环境交互而言有用的技术。这一宽广的范围使得所公开的技术非常适合于不计其数的应用。

[0052] 由于本公开中详述的主题的范围和多样性极大，所以很难实现有条理的介绍。明显的是，下面呈现的许多主题章节既以其他章节为基础，又是其他章节的基础。因而，不可避免地，各章节是按照有点任意的顺序呈现的。应认识到的是，来自每个章节的一般原理和特定细节也可以在其他章节中得到应用。为了防止本公开的长度膨胀失控(简明总是有益的，特别是在专利说明书中)，不同章节的特征的各种置换和组合并没有无遗漏地详述。本发明人意图明确地教导这些组合 / 置换，只是实践性要求把所详述的合成方案留给根据这些教导最终实现本系统的那些人来决定。

[0053] 还应注意的是，这里详述的技术建立在前面引用的专利申请中所公开的技术上并对其进行了扩展。因此请读者参考那些详述了申请人期望本技术被应用于的方案并且在技术上对本公开进行了补充的文献。

[0054] 认知，非居间化(disintermediated)搜索

[0055] 移动装置(如手机)正在成为认知工具，而不仅仅是通信工具。在一个方面中，认知可以被认为是向一个人告知这个人所处的环境的活动。认知动作可以包括 ：

[0056] • 基于传感输入来感知各种特征 ；

[0057] • 感知各种形式(例如，确定协调地结合起来的结构)；

[0058] • 关联，如确定外部结构和关系 ；

[0059] • 定义各种问题 ；

[0060] • 定义问题解决状态(例如，它是文本 ：我可以做什么？ A. 读取它)；

[0061] • 确定解决方案选项 ；

[0062] • 启动动作和响应 ；

[0063] • 识别通常是确定适当响应的过程中的第一个基本步骤。

[0064] 视听移动装置是辅助进行向一个人告知其所处环境的过程中所涉及的那些处理的工具。

[0065] 移动装置以惊人的速率激增。许多国家(包括芬兰、瑞典、挪威、俄罗斯、意大利和英国)据传道具有的手机多于人口。根据 GSM 联盟，当前有近似四十亿个 GSM 和 3G 手机在使用中。国际电信联盟估算在 2009 年末会有 49 亿移动蜂窝用户。升级周期是如此之短，以致于平均每 24 个月就要更换一次装置。

[0066] 因此，移动装置已经是巨大投资的焦点。行业巨头(如 Google、Microsoft、Apple 和 Nokia)已经认识到巨大的市场取决于扩展这些装置的功能性，并且已经在研究和开发中投资了相当大的款项。在付出这样普遍且强烈的努力后，行业巨头仍未能开发出这里详述的技术，这着实证明了这里详述的技术的创造性。

[0067] “非居间化搜索”(如视觉查询)被相信是对于即将来临的各代移动装置而言最引人注目的应用之一。

[0068] 在一个方面中，非居间化搜索可以被认为是减少（乃至消除）人类在启动搜索的过程中的任务的搜索。例如，智能电话可以始终分析视觉环境，并且不用特意询问就提供解释和相关信息。

[0069] 在另一方面中，非居间化搜索可以被认为是超越 Google 的下一步。Google 构建了统一的大规模系统来把关于公共万维网的全部文本信息组织起来。但是视觉世界太大且太复杂，以致于甚至是 Google 都无法控制。一定会牵扯到无数参与者——每个参与者起着专门的作用，一些作用较大，一些作用较小。将不会存在“一个搜索引擎能支配他们全部”。（考虑到潜在地会牵扯到无数的参与者，或许备选的绰号将是“超居间化搜索（hyperintermediated search）”。）

[0070] 根据下面的讨论而明显可知的是，本发明人相信视觉搜索具体地在其某些方面中是极端复杂的，并且需要装置和云在高度交互的移动屏幕用户界面的支持下进行紧密的协作以产生令人满意的体验。用户的引导和交互至少在最初对搜索结果的有用性是起根本作用的。在本地装置上，关键的挑战是如何将稀少的 CPU/ 内存 / 通道 / 功率资源分配给令人眼花缭乱的一批需求。在云端，基于拍卖的服务模型预期会出现从而推动技术的发展。最初，非居间化搜索会以封闭系统的形式被商业化，但是要走向繁荣，非居间化搜索将要通过可扩展的开放平台来实现。最终，最成功的技术将会是被用来向用户提供最高价值的那些技术。

[0071] 架构图

[0072] 图 1 用直觉计算平台或 ICP 的架构图示出采用本技术的某些原理的实施例。（应该认识到的是，将功能划分成多个块是有点任意的。实际的实现方案可能并不遵循这里描绘和描述的特定结构。）

[0073] ICP 小玩意 & 空间模型组件（ICP Baubles & Spatial Modelcomponent）处理涉及观察空间、显示、及其关系的任务。一些相关功能包括与把小玩意叠盖到视觉景象上的过程有关的姿势估计、跟踪、和正射校正映射（ortho-rectified mapping）。

[0074] 在一个方面中，小玩意可以被认为是与所拍摄图像的特征相关联地显示在屏幕上的增强现实图标。这些小玩意可以具有交互性和用户调谐性（即，不同的小玩意可以出现在不同用户的屏幕上，从而察看同一景象）。

[0075] 在一些方案中，小玩意显现出来指示系统最先隐约识别出的东西。当系统开始辨别出在显示器上的某一位置处存在着用户潜在感兴趣的某个东西（视觉特征）时，系统呈现小玩意。随着系统推断出更多有关该特征的信息，小玩意的尺寸、形状、颜色或亮度可以发生变化，从而使得其更加突出和 / 或使得其提供的信息更加丰富。如果用户轻拍小玩意从而表示对该视觉特征感兴趣，那么系统的资源管理器（例如，ICP 状态机）可以不均衡地对该图像特征的分析处理拨划比其他图像区域更多的处理资源。（关于用户这一轻拍动作的信息也与关于该特征或该小玩意的信息一起存储在数据存储器中，使得用户对该特征的兴趣可以在下一次被更快速地识别或自动地识别。）

[0076] 当小玩意第一次出现时，关于该视觉特征，可能除了它看上去似乎构成视觉上分立的实体（例如，明亮的斑点，或具有边缘轮廓的某个东西）以外什么都不知道。在该理解水平上，可以显示一般的小玩意（或许被称为“原型小玩意（proto-bauble）”）如小星形或圆形。随着更多关于该特征的信息得以推断出来（它看上去似乎是面部或条形码或树叶），可

以显示使加深的理解得到反映的小玩意图形。

[0077] 小玩意可以在性质上是商用的。在一些环境中，在显示屏上可能会泛滥着不同的小玩意来竞争用户的关注。为了解决该问题，可以存在用户可设定的控制(视觉冗长控制)，其能够调节在屏幕上呈现多少信息。附加地或者备选地，可以提供一种控制，其允许用户建立商用小玩意与非商用小玩意的最大比率。(如同 Google 那样，从长期来看，从系统收集原始数据会证明比向用户呈现广告更有价值。)

[0078] 合乎期望的是，被选择进行显示的小玩意是基于各种维度的当前背景环境确定的对用户而言最有价值的那些小玩意。在一些情况下，商用和非商用小玩意都可以基于在云中进行的拍卖处理来选择。最终被显示的小玩意的名单可以受用户影响。用户与之交互的那些小玩意成为明显的受偏爱者并且更可能在未来被显示；用户反复忽略或摒弃的那些小玩意不会被再次显示。

[0079] 可以提供另一 GUI 控制来指示用户的当前兴趣(例如，观光、购物、远足、社交、航行、吃饭等)，并且可以相应地调谐小玩意的呈现。

[0080] 在一些方面中，(左侧具有音量旋钮并且右侧具有调谐旋钮的) 旧式汽车收音机的类似物是适合的。音量旋钮对应于用户可设定的对屏幕繁忙度(视觉冗长度)的控制。调谐旋钮对应于单独地或者联合地指示出当前与用户相关的内容类型的传感器、存储数据和用户输入(例如用户的可能意图)。

[0081] 图示的 ICP 小玩意 & 空间模型组件可以借用或者基于发挥相关功能的现有软件工具来构建。一个现有软件工具是 ARToolKit——起因于华盛顿大学的人机界面技术实验室的研究而产生的可免费获得的一套软件(hitl<dot>Washington<dot>edu/artoolkit/)，现在由西雅图的 ARToolworks 公司(artoolworks<dot>com)进一步开发。另一套相关工具是 MV 工具——一种流行的机器视觉函数库。

[0082] 图 1 仅示出一些识别代理(RA)；可以存在几十或几百个识别代理。识别代理包括基于传感器数据(例如，像素)和 / 或派生物(例如，“关键字向量”数据，参见 US20100048242、WO10022185) 来执行特征和形式提取并帮助进行关联和识别的组件。它们通常帮助识别可用信息、并从可用信息中提取含义。在一个方面中，一些 RA 可以类推为是专门搜索引擎。一个可以搜索条形码，一个可以搜索面部，等等。(RA 也可以是其他类型，例如在不同处理任务的服务中处理音频信息、提供 GPS 和磁力计数据等。)

[0083] RA 可以基于会话和环境的需要而在本地执行处理，在远程执行处理，或者在本地和远程都执行处理。RA 可以根据装置 / 云协商的行业规则而在远程被加载和操作。RA 通常取来自一共享数据结构(即下面讨论的 ICP 黑板) 的关键字向量数据作为输入。RA 可以提供基本服务，这些基本服务由 ICP 状态机根据解答树而被组合起来。

[0084] 如同小玩意那样，可以存在涉及 RA 的竞争。即，相互重叠的功能性可以由来自几个不同提供商的几个不同 RA 提供。在特定背景环境中在特定装置上使用哪个 RA 的选择可以随用户的选择、第三方评论、成本、系统限制、输出数据的可再利用性、和 / 或其他标准而变。最终，达尔文筛选可能会发生，使得最好地满足用户需求的那些 RA 成为主流。

[0085] 智能电话供应商可以在最初为该智能电话提供一组默认的 RA。一些供应商可以保持对 RA 选择的控制(围墙花园式方法)，而一些供应商可以鼓励用户发现不同的 RA。在线市场(如苹果应用软件商店)可以发展成充当 RA 市场。为不同的客户群和需求服务的 RA

包可能会出现，例如一些 RA 包能帮助视力有限的人(例如，载有视力帮助 RA 如文本到语音识别)，一些 RA 包能设法满足期望最简单的用户界面的那些人的需要(例如，大按钮控制，非行话图注)；一些 RA 包能设法满足户外爱好者的需要(例如，包括鸟鸣声识别 RA、树叶识别 RA)；一些 RA 包能设法满足世界旅行者的需要(例如，包括语言翻译功能和基于位置的旅行者服务)，等等。系统可以提供一菜单，借助该菜单用户可以使装置在不同的时刻加载不同的 RA 包。

[0086] 一些或全部 RA 可以取决于具体情况而将功能性推送给云。例如，如果可利用去往云的快速数据连接、并且装置的电池接近耗尽(或者如果用户正在玩消耗装置的大部分 CPU/GPU 资源的游戏)，那么本地 RA 可以仅在本地完成一小部分任务(例如，仅进行管理)，并将其余的任务发给云中的对应部分以便在那里执行。

[0087] 如其他地方详述的那样，可由 RA 利用的处理器时间和其他资源可以以动态方式控制——将更多的资源分配给看上去似乎值得该待遇的那些 RA。ICP 状态机的分配器组件可以专心于这种照管。ICP 状态机也可以管理在本地 RA 组件和云中的对应部分之间进行的 RA 操作分配。

[0088] ICP 状态机可以采用以安卓开源操作系统(例如，developer<dot>android<dot>com/guide/topics/fundamentals.html) 以及 iPhone 和 Symbian SDK 为模型设计的一些方面。

[0089] 图 1 中的右边是云 & 商务规则组件，其充当对云相关处理的接口。它也可以执行对云拍卖的管理——确定由多个云服务提供商中的哪一个来执行某些任务。它通过服务提供商接口(SPI)与云进行通信，其中服务提供商接口基本上可以利用任何通信通道和协议。

[0090] 尽管特定的规则将是不同的，但是可以用作本架构的该方面的模型的示例性基于规则的系统包括：电影实验室内容规则和权利方案(例如 movielabs<dot>com/CRR/) 和 CNRI 处理系统(例如 handle<dot>net/)。

[0091] 图 1 中的左边是背景环境引擎，其提供并处理由系统使用的背景环境信息(例如，当前位置是哪里？用户在上一分钟执行了什么动作？用户在上一小时执行了什么动作？等等)。背景环境组件可以跨越接口链接到远程数据。远程数据可以包括任何外部信息，例如有关活动、同等群体(peer)、社交网络、消费的内容、地理的信息——可以使本用户与其他人联系起来的任何信息(如相似的度假目的地)。(如果装置包括音乐识别代理，那么它可以查阅用户的 Facebook 朋友的播放列表。装置可以使用该信息来精炼用户所听的音乐的模型——还考虑例如关于用户预订的在线广播电台的认识等。)

[0092] 背景环境引擎和云 & 商务规则组件可以具有残留在云侧的对应部分。即，该功能性可以是分布式的，一部分在本地，并且在云中有一对应部分。

[0093] 基于云的交互可以利用关于 Google 的应用软件引擎(App Engine)(例如，code<dot>Google<dot>com/appengine/) 和 Amazon 的弹性计算云(Elastic Compute Cloud)(例如，aws<dot>amazon<dot>com/ec2/) 进行的相关云计算已经公开的许多工具和软件。

[0094] 图 1 的底部是黑板和聚类引擎(Blackboard and Clustering Engine)。

[0095] 黑板可以服务于各种功能，包括充当共享数据储存库、以及充当各处理间的通信手段，从而允许多个识别代理观察和贡献特征对象(例如，关键字向量)并相互合作。黑板可

以充当用于系统的数据模型，例如保持视觉表现以帮助进行跨越多个识别代理的特征提取和关联，为时间特征 / 形式提取提供缓存和支持，以及提供存储管理和垃圾桶服务。黑板也可以充当特征类工厂（feature class factory），并提供特征对象例示（创建和损毁、访问控制、通知、以关键字向量形式串行化、等等）。

[0096] 黑板功能性可以利用开源黑板软件 GBBopen（gbbopen<dot>org）。在 Java 虚拟机上运行（并且支持用 JavaScript 编写脚本）的另一开源实现方案是黑板事件处理器（Blackboard Event Processor）（code<dot>Google<dot>com/p/blackboardeventprocessor/）。

[0097] 黑板构造是由 Daniel Corkill 推广的。参看例如 Corkill 的“Collaborating Software － Blackboard and Multi-Agent Systems & theFuture”（Proceedings of the International Lisp Conference, 2003）。然而，本技术的实现方案不需要该概念的任何特殊形式。

[0098] 聚类引擎使多项内容数据（例如，像素）在例如关键字向量（keyvector）中成群到一起。在一个方面中，关键字向量可以大致类推为是文本关键字的视听对应物——输入到处理中以便获得相关结果的一群元素。

[0099] 聚类可以由根据图像数据生成新特征（即，可被表示为点、向量、图像区域等的列表的特征）的低级别处理执行。（识别操作通常寻找相关特征的聚簇，因为这些聚簇潜在地表示感兴趣的对象。）这些特征可以发布到黑板。（可形成识别代理的一部分的更高级别处理也可以生成感兴趣的新特征或对象，并且也将其发布到黑板。）

[0100] 再一次，前面提到的 ARToolKit 可以为该功能性的某一方面提供基础。

[0101] 上文的各方面在本说明书的下面及其他章节中将进一步详述。

[0102] 本地装置 & 云处理

[0103] 如图 2 概念性示出的那样，非居间化搜索应该依靠本地装置和云的强度 / 属性。（云“管道”也作为因素计入该混合物中（例如通过包括带宽和成本在内的限制）。）

[0104] 功能性在本地装置和云之间的特定分配随着实现方案的不同而变化。在一个特定实现方案中，功能性被划分如下：

[0105] 本地功能性：

[0106] • 背景环境：

[0107] －用户身份、偏好、历史

[0108] －背景环境元数据处理（例如，我是谁？我现在面向什么方向？）

[0109] • UI：

[0110] －在屏幕上呈递 & 反馈（触摸、按钮、音频、接近、等等）

[0111] • 大体定向：

[0112] －全局采样；在不进行很多分析的情况下进行分类

[0113] －数据对齐（data alignment）和特征提取

[0114] －特征的枚举拼凑物（enumerated patchwork）

[0115] －帧间采集；时间特征的序列

[0116] • 云会话（Cloud Session）管理：

[0117] －对识别代理的登记、关联 & 双向会话操作

[0118] • 识别代理管理：

[0119] —类似于具有特定功能性的动态链接库(DLL)——识别特定身份和形式

[0120] —资源状态和检测状态可伸缩性

[0121] —由识别代理提供的服务的组成

[0122] —开发和许可平台

[0123] 云角色可以包括：例如，

[0124] • 与所涉及的云端服务通信

[0125] • 管理和执行关于服务的拍卖(和/或审核装置上的拍卖)

[0126] • 例如通过提供与七条身份法则(seven laws of identity，参看微软公司的Kim Cameron)相关联的服务，来提供/支持用户和对象的身份：

[0127] —用户控制和同意。技术身份系统必须仅在用户同意的情况下才揭示标识该用户的信息。

[0128] —对于所限定的用途做最小的公开。公开最小量的标识信息并最佳地限制其使用的解决方案是最稳定的长期解决方案。

[0129] —正当的参与者。数字身份系统必须设计成使得标识信息的公开仅被局限于在给定的身份关系中具有必要和正当地位的参与者。

[0130] —有方向的身份。通用身份系统必须既支持供公共实体使用的“无定向”标识符、又支持供私人实体使用的“单向”标识符，从而在防止相关句柄(correlation handle)被不必要地释放的同时便于身份的发现。

[0131] —运营者和技术的多元化。通用身份系统必须使多个身份提供商运行的多种身份技术通道化并使所述多种身份技术能够相互配合。

[0132] —人类整合。通用身份元系统必须将人类用户定义为通过明确的人类/机器通信机制而被整合的分布式系统的组成部分，从而提供对身份攻击的防护。

[0133] —跨越背景环境的一致体验。统一的身份元系统必须在通过多个运营者和技术实现背景环境的分割的同时，保证其用户享受简单一致的体验。

[0134] • 创建和实施域的构造

[0135] —开帐单、地理、装置、内容

[0136] • 执行和控制用户启动的会话内的识别代理

[0137] • 管理远程识别代理(例如，材料供应、认证、撤销、等等)

[0138] • 照管商务规则和会话管理等

[0139] 云不仅使非居间化搜索变得容易，而且常常也是搜索的目的地(除了诸如OCR之类的仅基于传感器数据通常就能够提供结果的情况以外)；

[0140] 这里详述的技术从包括以下来源的各种来源吸取启发：

[0141] • 生物学：类似于人类视觉系统&高级认知模型

[0142] • 信号处理：传感器融合

[0143] • 计算机视觉：图像处理操作(空间&频率域)

[0144] • 计算机科学：服务的组成&资源管理，并行计算

[0145] • 机器人学：用于自主交互的软件模型(PLAN、Gazebo等)

[0146] • AI：匹配/盘算/执行模型，黑板、规划模型等

[0147] • 经济学 :拍卖模型(次高价中标(Second Price Wins) …)

[0148] • DRM :权利表达语言 & 商务规则引擎

[0149] • 人类因素 :UI,增强现实,

[0150] • 移动价值链结构(Mobile Value Chain Structure):风险承担者,商务模型,政策,等等

[0151] • 行为科学 :社交网络,众包 / 大众分类法(folksonomy)

[0152] • 传感器设计 :磁力计、近程传感器、GPS、音频、光学(景深延伸等)

[0153] 图 3 用不同的功能方面(按照系统模块和数据结构)对说明性认知处理的各种特征进行映射。因此,例如,直觉计算平台(ICP,IntuitiveComputing Platform)背景环境引擎把关联、问题解决状态、确定解决方案、启动动作 / 响应、和管理这些认知处理应用于系统的背景环境方面。换句话说,ICP 背景环境引擎尝试基于历史等来确定用户的意图,并使用这样的信息来告知系统操作的各方面。同样,ICP 小玩意 & 空间模型组件在向用户呈现信息和从用户接收输入这些方面进行许多相同的处理。

[0154] ICP 黑板和关键字向量是与系统的定向方面相关联地使用的数据结构。

[0155] ICP 状态机 & 识别代理管理与识别代理共同照管识别处理以及与识别相关联的服务的组成。状态机通常是实时操作系统。(这些处理也涉及例如 ICP 黑板和关键字向量。)

[0156] 云管理 & 商务规则处理云登记、关联和会话操作——在识别代理和其他系统组件与云之间提供接口。

[0157] 支持小玩意的本地功能性

[0158] 与小玩意相关的一个或更多软件组件所提供的功能中的一些可以包括以下功能 :

[0159] • 理解用户的概况、用户的一般兴趣、用户在其当前背景环境内的当前特定兴趣。

[0160] • 对用户输入做出响应。

[0161] • 使用所选的来自全局图像处理库的模块来对流式形式的多个帧的重叠景象区域进行空间解析和“对象识别(object-ify)”

[0162] • 把呈分层结构的多层符号(像素分析结果、ID、属性等)附加到原型区域上 ;将其打包成原型查询的“关键字向量”。

[0163] • 基于用户设定的视觉冗长水平和全局景象理解,设立小玩意原始显示功能 / 正射投影。

[0164] • 将关键字向量路由到适当的本地 / 云地址

[0165] • 使所附加的“完整背景环境”元数据来自列在顶部的路由对象。

[0166] • 如果路由到本地地址,则处理该关键字向量并产生查询结果。

[0167] • 收集关键字向量查询结果并使适当的小玩意在用户屏幕上活跃 / 把适当的小玩意位块传送(blit)到用户屏幕

[0168] • 小玩意可以是“完全且充分地可引起行动得以采取”,或者可以示出“临时状态”并因此期待用户交互以便进行更深的查询钻研或查询精炼。

[0169] 直觉计算平台(ICP)小玩意

[0170] 在云中进行提供服务和高价值的小玩意结果这一方面的竞争应该激励供应商变得优异并取得商业成功。建立具有基线品质的非商用服务的云拍卖地点可以帮助激励该市

场。

[0171] 用户想要(并且应该会需要)最高品质和最相关的小玩意,使商业入侵程度随用户意图和实际查询而变地得到调节。

[0172] 在对立面上,把屏幕作为不动产购买的购买者可以分成两类:愿意提供非商用小玩意和会话的那些购买者(例如,带着争取客户以打造品牌的目标),以及想要"有资格"拥有作为不动产的屏幕(例如,按照将会看到该屏幕的用户的人口统计状况的形式)并且仅仅对这些屏幕所代表的商业机会投标的那些购买者。

[0173] 当然,在把自己的"关键字、拍卖处理、赞助的超链接呈现"货币化方面,Google 已经建立了巨大的产业。然而,对于视觉搜索,单个实体似乎不太可能会相似地支配该处理的所有方面。而是,似乎可能的是,处于中间层的公司将辅助进行用户查询 / 屏幕不动产购买者的匹配。

[0174] 用户界面可以包括一种控制,借助该控制用户可以摒弃不感兴趣的小玩意——从屏幕上将其去除(并且终止专用于发现与该视觉特征相关的进一步信息的任何正在进行的识别代理处理)。关于被摒弃的小玩意的信息可以记录到数据存储库中,并用于扩充用户的概况信息。如果用户摒弃关于星巴克咖啡店和独立咖啡店的小玩意,那么系统可以推断出用户对所有咖啡店都缺乏兴趣。如果用户仅摒弃了关于星巴克咖啡店的小玩意,那么可以辨别出更窄的用户兴趣缺乏范围。将来进行的小玩意的显示可以查阅数据存储库;早先被摒弃(或者反复被摒弃)的小玩意通常不会被再次显示。

[0175] 类似地,如果用户轻拍小玩意从而表示出兴趣,那么该类型或该类别的小玩意(例如,星巴克、或咖啡店)可以在将来在评估(在许多候选者当中)要显示哪些小玩意时被赋予较高的分数。

[0176] 关于用户与小玩意间的交互的历史信息可以与当前背景环境信息结合使用。例如,如果用户在下午而不是在上午摒弃了与咖啡店相关的小玩意,那么系统可以在上午继续呈现与咖啡相关的小玩意。

[0177] 视觉查询问题固有的复杂性意味着,许多小玩意将属于临时的或原型小玩意那一类——邀请并引导用户提供人类级别的过滤、交互、引导和导航以便更深入地进行查询处理。在某一景象上进行的小玩意显示的进展因此可以随实时人类输入以及其他因素而变。

[0178] 当用户轻拍小玩意或以其他方式表达出对小玩意感兴趣时,该动作通常会启动与该小玩意的主题相关的会话。会话的细节将取决于特定的小玩意。一些会话可以在性质上是商用的(例如,轻拍星巴克小玩意可以获得星巴克产品的优惠一美元的电子赠券)。一些会话可以是提供消息的(例如,轻拍与雕像相关联的小玩意可以导致关于该雕像或雕刻家的 Wikipedia 条目的呈现)。表示识别出所拍摄的图像中的面部的小玩意可以导致各种操作(例如,呈现来自社交网络(如 LinkedIn)的有关这个人的概况;把该照片的带有关于面部的注释的副本发布到识别出的这个人的 Facebook 页面或发布到该用户的 Facebook 页面,等等)。有时,轻拍小玩意会唤来由几个操作构成的菜单,用户可以从该菜单中选择期望的动作。

[0179] 轻拍小玩意表示该小玩意所对应的种类胜过了其他小玩意。如果轻拍的小玩意在性质上是商用的,那么该小玩意赢得了对用户的关注以及对观看者的屏幕这一不动产的暂时利用的竞争。在一些情况下,可以做出相关的支付——或许支付给用户,或许支付给另一

方(例如保证其“赢得”客户的那一实体)。

[0180] 轻拍的小玩意还表示对偏好的表决——可能达尔文同意该小玩意优于其他小玩意。除了影响对在将来呈现给该用户的待显示小玩意的选择之外,这样的确认也会影响对显示给其他用户的小玩意的选择。这有希望把小玩意提供商引导到朝向优异用户服务迈进的良性循环。(如果只有用户喜爱的广告能够获得正在进行的播放时间,那么有多少当前的电视广告会幸存?)

[0181] 如所示的那样,给定的图像景象可以为许多小玩意(常常是屏幕可以有效包含的更多小玩意)的显示提供机会。把该可能性范围缩小到易管理的集合的处理可以从用户开始。

[0182] 可以采用各种不同的用户输入,从前面提到的冗长控制开始,所述冗长控制仅对用户希望屏幕被叠盖有小玩意的频繁度设定基线。其他控制可以指示当前偏好、以及商用小玩意与非商用小玩意的指定混合比例。

[0183] 另一维度的控制是用户在屏幕的特定区域中的实时兴趣表达,例如指示关于用户想要获得更多了解的事物的特征或者指示用户想要以其他方式进行交互的特征。该兴趣可以通过轻拍叠盖在这些特征上的原型小玩意来指示,尽管也可能不需要原型小玩意(例如,用户可以简单地轻拍屏幕的未显出差别的区域以便将处理器的注意力集中到图像帧的该部分上)。

[0184] 另外的用户输入是与背景环境有关的——包括在其他地方详述的许多种信息(例如,计算背景环境、物理背景环境、用户背景环境、实体背景环境、时间背景环境和历史背景环境)。

[0185] 馈送给小玩意选择处理的外部数据可以包括与第三方交互相关的信息——其他人选择与什么小玩意进行交互?赋予该因素的权重可以取决于其他用户和本用户之间的距离度量、以及其他用户的背景环境和本背景环境之间的距离。例如,本用户的社交朋友在相似背景环境情况下的动作所表达的小玩意偏好可以被赋予比陌生人在不同情况下的动作所表达的小玩意偏好大得多的权重。

[0186] 另一外部因素可以是商业考虑因素,例如,第三方愿意支付多少(并且可能的话愿意支付给谁)来暂时地租借作为不动产的少量用户屏幕?如上所述,这样的问题可以作为因素计入基于云的拍卖方案中。拍卖也可以考虑特定小玩意对于其他用户的流行度。在实现本处理的该方面时,可以参考 Google 的用于在线拍卖作为不动产的广告的技术(参看例如 Levy 的 Secret of Googlenomics:Data-Fueled Recipe BrewsProfitability,Wired Magazine,2009 年 5 月 22 日)——广义的次价拍卖的一种变型。本申请人在已公开的 PCT 申请 WO2010022185 中详述了基于云的拍卖方案。

[0187] (简要地,这种基于云的模型的假定是,这些模型类似于以点击率(CTR)为基础的广告模型:各实体将支付数量不定的金钱和/或赞助服务,以确保其服务被使用、和/或确保其小玩意会出现在用户的屏幕上。合乎期望的是,对于由商用和非商用识别代理(例如,已经将 Starbucks 标志预缓存的标志识别代理)提供的识别服务,存在着一个动态的市场。也可以从搜索告知型广告(search-informed advertising)吸取经验——平衡是在交易上赢利的同时向用户提供价值。

[0188] 通常,这些拍卖中的挑战不处于拍卖的进行中,而是在于适当地处理所涉及的变

量的数目。这些变量包括：

[0189] • 用户概况(例如，基于例如通过浏览器世界中的 cookies 而已知的信息——供应商想要花费多少钱来放置一个小玩意)

[0190] • 成本(带宽、计算和机会成本是多少？)；以及

[0191] •装置能力(既包括静态方面，例如硬件供给——闪存？ GPU？ 等等，又包括动态状态方面，例如用户当前位置处的信道带宽、装置的功率状态、内存利用率、等等)

[0192] (在一些实现方案中，小玩意推销商可以根据用户使用的装置类型的指示而更努力地将小玩意放置在富有的用户的屏幕上。具有最新最昂贵类型的装置的用户或者使用昂贵的数据服务的用户可以比具有陈旧装置的用户或者使用后沿数据服务的用户值得受到更多的商业关注。由用户暴露的或者可根据环境推断出的其他概况数据可以类似地由第三方在决定哪些屏幕是其小玩意的最佳目标时使用。)

[0193] 在一个特定实现方案中，可以把几个小玩意(例如，1-8 个)分配给商业宣传(例如，通过类似 Google 的拍卖程序确定，并服从于用户对商用小玩意与非商用小玩意的比例的调谐)，并且一些小玩意可以基于非商业因素(如前面提到的那些)来选择。这些后一种小玩意可以按照基于规则的方式来选择，例如应用对前面提到的不同因素施加权重的算法以便对每个小玩意获得一个分数。然后对相互竞争的各分数进行排序，并把分数最高的 N 个小玩意呈现在屏幕上(其中 N 可以由用户使用冗长控制来设定)。

[0194] 在另一实现方案中，并不先验地分配商用小玩意。而是，按照类似于非商用小玩意的方式对这些小玩意进行评分(通常使用不同的标准，但是按比例缩放到相似的分数范围)。然后呈现分数最高的 N 个小玩意——它们可能全部是商用的、全部是非商用的、或者是混合的。

[0195] 在另一实现方案中，商用小玩意与非商用小玩意的混合比例是随用户的预订服务而变的。对支付介绍性比率的处于入门级别的用户呈现在尺寸和 / 或数量方面较大的商用小玩意。对于为了获得优质服务而付款给服务提供商的用户，向他们呈现较小和 / 或较少的商用小玩意，或者向他们赋予一定的自由来设定他们自己的关于商用小玩意的显示的参数。

[0196] 表示小玩意的图形标志可以在视觉上设计成适合于指示其特征关联，并且可以包括动画元素来吸引用户的注意。小玩意提供商可以向系统提供一定尺寸范围内的标志，从而允许系统在用户放大所显示图像的该区域或者表达对这种小玩意的潜在兴趣的情况下，增大小玩意的尺寸和分辨率。在一些情况下，系统必须充当警察——决定不呈现所提供的小玩意，例如因为该小玩意的尺寸超过由所存储的规则建立的尺寸、因为该小玩意的外观被认为是淫秽的、等等。(系统可以自动地将小玩意按比例缩小至适合的尺寸，并且用一般标志(如星形标志)替换不适合的或不可用的标志。)

[0197] 除了与从图像中辨别出的视觉特征相关地呈现小玩意以外，还可以以其它方式呈现小玩意。例如，可以呈现小玩意来指示装置知道其地理位置，或者指示装置知道其用户的身份。各种操作反馈因此可以提供给用户，而不管图像内容如何。除了特定特征识别以外，一些图像反馈也可以经由小玩意提供，例如所拍摄的图像满足基线质量标准(如焦点或对比度)。

[0198] 每个小玩意可以包含少量的映射表现，或者每个小玩意可以用基本图元的集合来

限定。通常，在平面图中限定小玩意标志。软件的空间模型组件可以根据所拍摄图像内的辨别出的表面来将小玩意标志的投影映射到屏幕上，例如对于与倾斜地观察的店面相关联的小玩意，其看上去似乎是倾斜的并且或许在透视图中是扭曲的。这些问题将在下面的章节中进一步讨论。

[0199] 空间模型 / 引擎

[0200] 在建立愉快的用户体验的过程中，把 3D 世界令人满意地投影并显示到 2D 屏幕上是很重要的。因此，优选系统包括服务于这些目的的软件组件(有各种称谓，例如空间模型或空间引擎)。

[0201] 在 2D 屏幕中再现 3D 世界开始于理解关于 3D 世界的某些事情。对于未加处理的一帧像素(缺乏任何地理位置数据或其他空间理解)，从哪里开始？如何辨别对象并加以分类？如何跟踪图像场景的移动，使得小玩意能够相应地被重新布置？幸运地是，这些问题已经在许多情况下被面对了许多次。机器视觉和视频运动编码是许多领域中提供了有用的现有技术的两个领域，假定本领域技术人员熟悉这些现有技术，并且本领域技术人员可以从这些现有技术中吸取与本申请有关的经验。

[0202] 借助基本原理：

[0203] • 摄像机和显示屏是典型的 2D 空间结构

[0204] • 摄像机通过 3D 世界到 2D 平面的空间投影来工作。

[0205] • 小玩意和原型小玩意被“客观地体现”到空间框架内。

[0206] 下面是把空间理解编码成正交处理流以及背景环境条目和属性条目的提议。该提议利用三个“空间水平”(空间理解的阶段)的构造。

[0207] 空间水平 1 包括基本景象分析和解析。将像素聚簇成多个初始群组。对于所拍摄的作为不动产的景象以及作为不动产的显示屏，存在着一些基本理解。关于跨越多个帧的作为不动产的景象构成的流，也存在着一些基本认识。

[0208] 在几何学上，空间水平 1 存在于简单 2D 平面的背景环境中。空间水平 1 操作包括生成从像素数据辨别出的 2D 对象的列表。由 OpenCV 视觉库(下面讨论)执行的基本操作属于该分析领域。智能电话的本地软件可以流畅处理空间水平 1 操作，并且丰富的 2D 对象列表可以在本地产生。

[0209] 空间水平 2 是过渡性的，从而懂得空间水平 1 的 2D 基元的一些含义，但是还没有达到空间水平 3 的完整 3D 理解。该级别的分析包括设法使不同的空间水平 1 基元相互联系(辨别在 2D 背景环境中对象如何相互联系)并且寻找达到 3D 理解的线索的任务。该级别的分析所包括的操作诸如识别对象群(例如，不同的边缘形成了限定出一形状的轮廓)、注意到各种样式(如沿着一条线的多个对象)、以及辨别“世界空间线索”(如消失点、地平线、和“向上 / 向下”的概念)。也可以揭示出“更近 / 更远”的概念。(例如，面部通常具有已知的尺寸。如果一组基本特征看起来很可能表示面部，并且该组基本特征在高度为 480 像素的场景中只有 40 像素的高度，那么相比于高度为 400 像素的面部像素集合，可以对高度为 40 像素的该组基本特征采集到“更远”属性。)

[0210] 将杂乱的空间水平 1 基元精炼 / 合成为对象相关实体的更短的、更有意义的列表。

[0211] 空间水平 2 可以把类似 GIS 的组织形式施加于景象和景象序列，例如向每个识别出的聚簇、对象、或感兴趣区域分配其自己的逻辑数据层(这些数据层可能有重叠区域)。每

个层可以具有相关的元数据存储库。在该水平上，可以辨别出帧与帧之间的对象连续性。

[0212] 在几何学上，空间水平 2 承认所捕获的像素数据是 3D 世界到 2D 图像帧的摄像机投影。先前辨别出的基元和对象并不被认为是现实的完全表征，而仅是一个视角。对象是在借以观察这些对象的摄像机镜头的背景环境中被考虑的。镜头位置建立据以理解像素数据的视角。

[0213] 空间水平 2 操作通常倾向于比空间水平 1 操作更加依赖云处理。

[0214] 在示例性实施例中，软件的空间模型组件是通用的——将像素数据精炼成更有用的形式。然后不同的识别代理可以在执行它们各自任务的过程中从该公共的精炼数据池中吸取数据，而不是各自进行它们自己版本的这种处理。然而，在决定哪些操作具有这样的通用性使得这些操作理所当然地以这种公共方式被执行、以及决定哪些操作应该被移交给各个单独的识别代理来仅根据需要执行的过程中，必须划一条线。（尽管如此，由识别代理产生的结果也可以例如通过黑板来共享。）可以任意地划出上述的线；设计者具有决定哪些操作落入该条线的哪一侧的自由。有时，该条线可以在智能电话操作期间动态移动（例如在识别代理请求进一步的公共服务支持的情况下）。

[0215] 空间水平 3 操作是基于 3D 的。不管数据是否揭示出完整的 3D 关系（通常不会这样），所进行的分析都是基于这样的前提：像素表示 3D 世界。这种理解对于某些对象识别处理是有用的、甚至是不可缺少的。

[0216] 空间水平 3 因此建立在先前的理解水平上，向外延伸到世界相关性。用户被理解为是具有给定投影和时空轨道的世界模型内的观察者。可以应用把景象映射到世界和把世界映射到景象的变换方程，使得系统理解它处于空间中的哪里和对象处于空间中的哪里，并且具有关于各事物间如何发生联系的某种框架。这些分析阶段从游戏工业和增强现实引擎的工作中吸取经验。

[0217] 与和空间水平 1 相关联的操作（以及和空间水平 2 相关联的一些操作）不同，与空间水平 3 相关联的操作通常是如此的专门化以致于它们并非例行地对输入数据执行（至少对于当前技术而言并非例行地对输入数据执行）。而是，这些任务被留给可能会需要特定 3D 信息的特定识别任务。

[0218] 一些识别代理可以构建用户环境的虚拟模型，并用 3D 背景环境中感测出的对象填充该模型。例如车辆驾驶监视器可以从用户汽车的挡风玻璃向外看，从而注意到与交通安全相关的条目和动作。识别代理可以维持交通环境及其内部的动作的 3D 模型。识别代理可以注意到用户的妻子（由另一软件代理识别出来，该软件代理将识别结果发布到黑板）正驾驶她的红色斯巴鲁车通过红灯（在用户的视角中）。支持这种功能性的 3D 建模当然是可能的，但是不是智能电话的一般服务例行执行的那种操作。

[0219] 这些方面中的一些在图 4 中示出，图 4 概念性地示出空间理解从空间水平 1、到 2、到 3 的逐渐增大的复杂性。

[0220] 在一种说明性应用中，不同的软件组件负责辨别与不同的空间水平相关联的不同类型的信息。例如，聚簇引擎被用于产生一些空间水平 1 的理解。

[0221] 聚簇是指把一群（通常是连续的）像素识别为相互关联的处理。这种关联可以是例如在颜色或纹理方面相似。或者这种关联可以是一个流中的相似性（例如，相似的面部像素图案跨越静态背景从一帧移位到另一帧）。

[0222] 在一个方案中，在系统已经识别出一个像素聚簇之后，系统分配将要与该聚簇相关联的符号(例如，就像 ID 号那样简单)。在进一步管理和分析该聚簇方面，这是有用的(并且在例如数据链接方案中也是有用的)。可以将原型小玩意分配给该聚簇，并且参考标识符号来跟踪该原型小玩意。起因于系统所执行的解析和定向操作而产生的、使聚簇的位置与摄像机在 2D 和 3D 中的位置相关的信息，可以参考该聚簇的符号来组织。类似地，起因于与该聚簇相关联的图像处理操作而产生的数据可以参考该聚簇的符号来识别。同样地，用户的轻拍可以与该符号相关联地记入日志。这种把符号用作借以存储和管理与聚簇相关的信息的句柄(handle)的用法，可以延伸到与该聚簇相关的基于云的处理、与聚簇相关联的小玩意的进化，自始至终通过完整识别聚簇对象并基于此做出响应。(下面将介绍更详细命名的构造，例如包括会话 ID。)

[0223] 这些空间理解组件可以与其他系统软件组件并行工作，例如，维持公共 / 全局空间理解并设立代理和对象可以利用的空间框架。这样的操作可以包括把关于空间环境的当前信息发布到可分享的数据结构(例如，黑板)，识别代理可以查阅该可分享的数据结构以帮助理解它们正在看什么，并且图形系统可以在决定如何在当前景象上描绘小玩意的过程中参考该可分享的数据结构。不同的对象和代理可以设立与三个水平相关联的空间水平字段和属性条目。

[0224] 通过相继地产生几代这些系统，空间理解组件预期会成为装置的几乎反射性的生搬硬套的能力。

[0225] 直觉计算平台(ICP)状态机——服务的组成；面向服务的计算；识别代理

[0226] 如前所述，ICP 状态机可以实质上包括实时操作系统。它可以照管常规任务(如调度、多重任务处理、错误恢复、资源管理、信息传递和安全性)，以及对当前应用场合更特定的一些其他任务。这些附加的任务可以包括提供检查跟踪功能性、保证会话管理、以及确定服务的组成。

[0227] 检查跟踪功能性向商业实体提供保证，保证这些商业实体付款赞助的小玩意事实上确实被呈现给用户。

[0228] 保证会话管理涉及建立和维持与云服务和对窃听等有鲁棒性(例如通过加密)的其他装置的连接。

[0229] 服务的组成是指选择用于执行某些功能的操作(和这些组分操作的相关配合 / 编排)。在状态机操作的这些方面中会涉及到分派处理，例如使资源与各应用相协调。

[0230] 某些高级功能可能会使用来自各种低级操作的不同组合的数据来实现。对利用哪些功能以及在什么时候利用这些功能的选择可以基于许多因素。一个因素是有哪些其他操作已经在进行中或已经完成——其结果也可能服务于本需要。

[0231] 举例来说，条形码定位通常可以依赖于：计算所定位的水平对比度，并计算所定位的垂直对比度，并比较这些对比度数据。然而，如果跨越一图像的 16x16 像素块的 2D FFT 数据已经可从另一处理得到，那么作为替代可以将该信息用于定位候选的条形码区域。

[0232] 类似地，某一功能可能会需要关于图像中的长边缘的位置的信息，并且可以启动专用于产生长边缘数据的操作。然而，另一处理可能已经识别出该帧中的各种长度的边缘，并且可以简单地对这些现有结果进行过滤来识别长边缘，并使其得到再利用。

[0233] 另一实例是基于霍夫变换的特征识别。OpenCV 视觉库指示：该功能期望使用边缘

细化的图像数据作为输入数据。该功能还推荐通过将 Canny(坎尼)操作应用于边缘数据来生成边缘细化的图像数据。而该边缘数据共同地通过将 Sobel (索贝尔)滤波器应用于图像数据而生成。因此,霍夫程序的“常规”实现方案将会从 Sobel 滤波器开始,随后是 Canny 操作,然后调用霍夫法。

[0234] 但是边缘也可以通过除 Sobel 滤波器以外的方法来确定。并且细化的边缘可以通过除 Canny 以外的方法确定。如果系统已经具有边缘数据,即使该边缘数据是由除 Sobel 滤波器以外的方法生成的,那么仍可以使用该边缘数据。类似地,如果另一处理已经产生改良的边缘数据,即使该改良的边缘数据不是通过 Canny 操作生成的,仍可以使用该改良的边缘数据。

[0235] 在一个特定实现方案中,系统(例如,分派处理)可以查阅一数据结构,该数据结构具有建立不同类型的关键字向量之间的大致功能对应度的信息。通过 Canny 产生的关键字向量边缘数据会指示出与通过无限对称指数滤波器技术产生的边缘数据具有较高的功能对应度,并且与通过 Marr-Hildreth (马尔 - 希尔德雷斯)程序辨别出的边缘数据有略小的功能对应度。通过 Harris 算子检测的拐角可以与通过 Shi 和 Tomasi 方法检测的拐角互换。等等。

[0236] 该数据结构可以包括一个大表格,或者该数据结构可以分解为几个表格——每个表格专用于特定类型的操作。例如,图 5 示意性地示出指示出对应度(按比例缩放到 100)的与边缘寻找相关联的表格的一部分。

[0237] 特定的高级功能(例如,条形码解码)可能会需要由特定处理(如 Canny 边缘滤波器)生成的数据。Canny 滤波功能可以从系统可利用的软件处理算法库中获得,但是在调用该操作之前,系统可以参考图 5 的数据结构,以查看是否已经有适合的代用数据可用或者正在处理中(假定还没有优选的 Canny 数据可用)。

[0238] 该检查开始于寻找在最左侧一栏中具有名义上期望的功能的行。然后程序在该行中扫描以寻找最高值。在 Canny 的情况下,最高值是无限对称指数滤波器所对应的 95。系统可以检查共享的数据结构(例如黑板),以便确定对于主题图像帧而言这样的数据(或适合的替代者)是否可用。如果找到这样的数据,那么可以将它代替名义上指定的 Canny 数据使用,并且条形码解码操作可以在该基础上继续进行。如果没有找到这样的数据,那么状态机继续进行处理——寻找次最高的值(例如,Marr-Hildreth 所对应的 90)。再一次,系统检查是否有任何该类型的数据可用。处理继续进行,直到表格中的所有备选者用尽为止。

[0239] 在本优选实施例中,该检查是由分派处理进行的。在这样的实施例中,大多数识别处理是作为多个操作的级联序列执行的——每个操作具有指定的输入。分派处理的使用允许对服务的参与组成所做的决定被集中做出。这也允许操作软件组件被聚焦于图像处理,而不是还要涉及例如检查表格以查找适合的输入资源以及维持对其他处理的操作的注意,这些负担会使这些组件更复杂且更难以维持。

[0240] 在一些方案中,通过条形码解码功能指定一阈值,或者由系统全局地指定一阈值,指示对于数据替代方案而言可接受的最小对应值(例如 75)。在这种情况下,刚刚描述的处理将不会考虑来自 Sobel 和 Kirch (克希霍夫)滤波器的数据,因为它们与 Canny 滤波器的对应度只是 70。

[0241] 尽管其他实现方案可能是不同的,但应注意的是,图 5 的表格是不对称的。例如,

如果期望的是Canny，那么Sobel具有的指示出的对应度只有70。但是如果期望的是Sobel，那么Canny具有的指示出的对应度为90。因此，Canny可以代替Sobel，但是如果设定的阈值为75，Sobel并不能代替Canny。

[0242] 图5的表格是通用的。然而，对于一些特定应用场合，图5的表格可能是不适合的。例如，某一功能可能会需要用Canny（优选）或Kirch或Laplacian（拉普拉斯算子）来寻找边缘。由于该功能的特性，其他边缘寻找器可能不是令人满意的。

[0243] 系统可以允许特定功能提供它们自己的关于一个或更多操作的对应表——优先于通用表格使用。对应于某一功能的专用对应表的存在可以用与该功能相关联的标记位或以其他方式来指示。在刚刚给出的实例中，标记位可以指示：应该使用图5A的表格作为替代。该表格仅包括单一一行——用于名义上指定的在该功能中使用的Canny操作。并且该表格仅具有两栏——对应于无限对称指数滤波器和Laplacian。（没有适合的其他数据。）对应值（即，95、80）可以省略，使得该表格可以包括备选处理的简单列表。

[0244] 为了便于在共享数据结构中找到可替代的数据，可以使用指示特定关键字向量包含什么信息的命名规则。这种命名规则可以指示功能的类别（例如，边缘寻找）、功能的特定种类（例如，Canny）、数据所基于的图像帧、以及数据所特有的任何其他参数（例如，用于Canny滤波器的核的尺寸）。该信息可以以各种方式表示，如按字面意义表示、用缩写表示、用可以通过另一数据结构解析从而获得完整细节的一个或更多指标值表示、等等。例如，包含用5x5的模糊核产生的关于帧1357的Canny边缘数据的关键字向量可以命名为“KV_Edge_Canny_1357_5x5”。

[0245] 为了向其他处理提醒正在处理中的数据，当一功能被初始化时可以将空条目写入共享的数据结构——根据该功能的最终结果来命名所述空条目。因此，如果系统开始用5x5的模糊核对帧1357执行Canny操作，那么空文件可以用上面提到的名称写入共享的数据结构。（这可以由该功能执行、或者由状态机（例如分派处理）执行。）如果另一处理需要该信息、并且用空条目找到适当命名的文件，那么它会知道这样的处理已经被启动。于是它可以监视或复查该共享的数据结构，并在所需信息变得可用时获得所需信息。

[0246] 更特别地，需要该信息的处理阶段在其输入参数当中会包括期望的边缘图像的规格——包括描述其所需的质量的描述符。系统（例如，分派处理）会检查当前位于存储器中（例如，位于黑板上）的数据的类型、以及上面提到的表格，以便确定目前是否有适当的数据可用或正在处理中。可能的动作于是可以包括：采用可接收的可用数据开始该处理阶段；当预期数据在将来可用时，将开始时刻延迟到将来的时刻；延迟开始时刻，并安排生成所需数据的处理（例如，Canny）得以开始；或者由于缺少所需数据和生成该所需数据所需的资源，而延迟或终止该处理阶段。

[0247] 在考虑备选数据是否适合于供特定操作使用时，可以对来自其他帧的数据加以考虑。如果摄像机处于自由运行模式，那么该摄像机可以每秒钟拍摄许多帧（例如，30帧）。尽管（在上面给出的实例中）分析处理可能会特别考虑帧1357，但是分析处理也能够利用从帧1356或者甚至从帧1200或1500取得的信息。

[0248] 在这点上，识别出包括在内容上相似的图像的帧所构成的群组是有帮助的。两个图像帧是否相似自然地将取决于特定的情况，例如图像内容和所执行的操作。

[0249] 在一个示例性方案中，如果（1）相关的感兴趣区域出现在帧A和帧B这两个帧中

(例如,相同的面部主题或条形码主题),并且(2)帧 A 和帧 B 之间的每个帧也包括该同一感兴趣区域,那么帧 A 可以被认为与帧 B 相似(这提供了对如下情况的某种保护措施:主题在摄像机最初观察该主题的状态和摄像机返回到该主题的状态之间变化)。

[0250] 在另一方案中,如果两个帧的颜色直方图在指定阈值内相似(例如,它们具有大于 0.95 或 0.98 的相关度),那么这两个帧被认为是相似的。

[0251] 在又一方案中,可以将类似 MPEG 的技术应用于图像流,以确定两个帧之间的差异信息。如果该差异超过阈值,那么这两个帧被认为是非相似的。

[0252] 除了上面提到的那些标准之外还可以利用的另外的测试是,该帧中的感兴趣特征或感兴趣区域的位置是相对固定的("相对"使得容许的移动可以有一阈值,例如 10 个像素、帧宽度的 10%、等等)。

[0253] 大量种类的其他技术可以备选地使用;这些技术仅是例证性的。

[0254] 在一个特定实施例中,移动装置维持一数据结构,该数据结构标识出相似的图像帧。这可以与标识出每个群组的开始帧和结束帧的表格一样简单,例如:

[0255]

开始帧	结束帧
…	…
1200	1500
1501	1535
1536	1664
…	…

[0256] 在一些方案中,可以提供第三字段——指示出由于某种原因(例如散焦)而不相似的处于指示的范围内的帧。

[0257] 返回到前面提到的实例,如果某一功能期望获得输入数据"KV_Edge_Canny_1357_5x5"并且没有找到这样的数据,那么该功能可以将搜索扩展到基于上述表格指示的相似性(大致等效性)来寻找"KV_Edge_Canny_1200_5x5"至"KV_Edge_Canny_1500_5x5"。如上所示,该功能也能够利用通过其他方法产生的边缘数据(同样,从帧 1200-1500 中的任何一个帧产生)。

[0258] 因此,例如,可以通过寻找帧 1250 中具有高水平对比度的区域和帧 1300 中具有低垂直对比度的区域,来定位条形码。在定位之后,可以通过参考在帧 1350 中找到的边界线结构(边缘)以及在帧 1360、1362 和 1364 中找到的符号图案的相关性来解码该条形码。因为全部这些帧都处于共同的群组内,所以该装置把从这些帧中的每一个帧取得的数据视为可与从这些帧中的其它每个帧取得的数据一起使用。

[0259] 在更复杂的实施例中,可以辨别出各帧之间的特征径迹(流),并将其用于识别各帧之间的运动。因此,例如,该装置可以理解:帧 A 中开始于像素(100,100)的线对应于帧 B 中开始于像素(101,107)的同一条线。(再一次,可以使用 MPEG 技术以便进行例如帧到帧

的对象跟踪。）可以做出适当的调整以便再对准该数据，或者该调整可以以其他方式引入。

[0260] 在较简单的实施例中，各图像帧之间的等效性仅简单地基于时间接近度。主题帧的给定时间跨度（或帧跨度）内的帧被认为是相似的。因此在寻找关于帧 1357 的 Canny 边缘信息时，系统可以接受来自帧 1352-1362（即，加减五个帧）中的任何一个帧的边缘信息是等效的。尽管该方法有时会导致失败，但是其简单性使得它在某些情况下是合乎需要的。

[0261] 有时，使用替代的输入数据的操作会失败（例如，该操作未能找到条形码或未能识别出面部），因为来自该代用处理的输入数据不具有该操作的名义上期望的输入数据的精确特征。例如，尽管很少发生，但是基于霍夫变换的特征识别仍可能会由于这样的原因而失败：输入数据不是通过 Canny 算子产生的而是通过代用处理产生的。在操作失败的情况下，可以再尝试进行该操作——这次采用不同的输入数据源。例如，可以利用 Canny 算子来代替代用者。然而，由于重复该操作需要成本并且通常对第二次尝试能够成功的期待较低，所以这样的再尝试通常并不例行公事地进行。可以试图进行再尝试的一种情况是，操作以自顶向下的方式被启动（诸如响应于用户的动作）。

[0262] 在一些方案中，对服务的初始组成的决定多少会取决于操作是自顶向下启动的、还是自底向上启动的（这些概念将在下面讨论）。例如，在自底向上的情况下，可以给予比自顶向下的情况更多的自由度来替换不同的输入数据源（例如，指示出的与名义上的数据源的对应度较小的数据源）。

[0263] 在决定服务的组成时可考虑的其他因素可以包括：功率和计算限制、进行某些基于云的操作的财务成本、拍卖结果、用户满意度排序、等等。

[0264] 再一次，可以参考提供每个代用操作的相对信息的表格，来帮助决定服务的组成。一个实例在图 6 中示出。

[0265] 图 6 的表格给出执行不同的边缘寻找功能所需的 CPU 和内存的度量。这些度量可以是某种实际的值（例如，对给定尺寸（例如 1024x1024）的图像执行规定的操作所需的 CPU 循环，执行这样的操作所需的 RAM 的 KB 数），或者可以被任意地按比例缩放（例如缩放到 0-100 的范围）。

[0266] 如果某一功能需要边缘数据（优选地来自 Canny 操作）并且还没有适合的数据可用，那么状态机必须决定是调用所要求的 Canny 操作还是调用另一操作。如果系统内存的供给不足，那么图 6 的表格（结合图 5 的表格）建议：可以使用无限对称指数滤波器作为替代：它在 CPU 负担方面只是略微大一些，但是占用少 25% 的内存。（图 5 指示出无限对称指数滤波器与 Canny 具有 95 的对应度，因此在功能上它应该能替代 Canny。）Sobel 和 Kirch 需要更少的内存占用，但是图 5 指示出这些操作可能是不适合的（70 分）。

[0267] 对于每个备选边缘寻找操作，实时状态机可以考虑各种参数（如图 5 和 6 中的分数，加上对应于成本、用户满意度、当前系统限制（例如，CPU 和内存利用率）、和其他标准的分数）。可以将这些参数输入给一处理，该处理根据多项式方程对各参数的不同组合赋予权重并求和。该处理的输出对可能被调用的不同的操作中的每一个得到一分数。具有最高分数的操作（或最低分数，这取决于方程式）被认为是当前情况下的最佳选择，并且随后由系统启动。

[0268] 尽管图 5 和 6 的表格仅考虑了这些功能的本地装置执行，但是也可以考虑基于云的执行。在这种情况下，该功能的处理器和内存成本基本上是零，但是可能会引起其他成

本，例如接收结果的时间会增长、要消耗网络带宽、以及可能会发生财务微支付。对于备选服务提供商和功能而言，这些成本中的每一个都可能是不同的。为了评估这些因素，可以对例如每个服务提供商和备选功能计算附加的分数。这些分数可以包括以下信息作为输入：取回结果的紧急度的指示，和根据基于云的功能预期的周转时间的增长；网络带宽的当前利用情况，和把该功能委托给基于云的服务所要消耗的额外带宽；预期到的功能（例如无限对称指数滤波器）相对于名义上期望的功能（例如 Canny）的可替代性；以及用户对价格的敏感度的指示，和对该功能的远程执行会收取什么样的价格（如果有的话）。也可以涉及各种其他因素，包括用户偏好、拍卖结果、等等。通过这样的计算而产生的分数可以用于在不同的远程提供商 / 所考虑的功能当中识别出优选的选项。然后，系统可以把该练习所产生的获胜分数与和本地装置对功能的执行相关联的练习所产生的获胜分数进行比较。（合乎期望的是，分数按比例被缩放到的范围是相当的。）然后，可以基于这样的评估来采取行动。

[0269] 对服务的选择也可以基于其他因素。根据背景环境、用户意图的指示等，可以识别出与当前情况相关的一组识别代理。从这些识别代理中，系统可以识别出由这些识别代理期望的输入构成的集合。这些输入可能会涉及具有其他不同的输入的其他处理。在识别出全部相关输入之后，系统可以限定出一个解答树，其包括指示的输入以及备选者。然后，系统识别穿过该解答树的不同路径，并选择（例如，基于相关限制而）被认为是最佳的一个路径。再一次，可以考虑本地处理和基于云的处理这两者。

[0270] 一种最佳性度量是通过对解决方案被发现的概率和所涉及的资源赋予参数而计算出的成本度量。于是该度量是以下等式表示的商：

[0271] 成本 =（所消耗的资源）/（解决方案被发现的概率）

[0272] 状态机可以通过优化（最小化）该函数来管理 RA 服务的组成。在这样做的过程中，状态机可以与云系统合作以管理资源并计算各解答树遍历的成本。

[0273] 为了便于此，RA 可以由多个阶段构建而成，每个阶段都向解决方案前进一步。合乎期望的是，RA 应该在其进入点处呈颗粒状、并且在其输出方面是详细冗长的（例如，暴露记录信息及其他信息、关于收敛性的置信度的指示、状态、等等）。通常，被设计成使用流式数据模型的 RA 是优选的。

[0274] 在这些方面中，本技术可以在例如“智能环境”方面从人工智能（AI）领域已知的“规划模型”中吸取经验。

[0275] （下面对规划模型的讨论部分地取自 Marquardt 的“Evaluating AIPlanning for Service Composition in Smart Environments”（ACM Conf.onMobile and Ubiquitous Media 2008,pp.48-55。）

[0276] 施乐帕克研究中心（Xerox PARC）的 Mark Weiser 所设想的智能环境是这样的环境：“混合在该环境中的传感器、致动器、显示器和计算单元很丰富并且让人看不到，该环境无缝地嵌入我们生活的日常对象中，并且通过连续网络连接”。这样的环境的特征在于，以不显眼的方式向用户提供个性化服务（例如，照明、加热、冷却、加湿、图像投射、报警、图像记录、等等）的各装置的动态集成。

[0277] 图 7 是例证性的。用户的意图通过例如观察并且通过参考背景环境而得以识别。从该信息中，系统推断出用户的假定目标。策略合成步骤试图找到满足这些目标的动作序列。最终，这些动作通过使用环境中可用的装置而得以执行。

[0278] 因为环境是可变化的，所以决定服务的组成的策略合成步骤必须是可自适应的(例如随着目标和可用装置的变化而自适应)。服务任务的组成被认为是人工智能“规划”问题。

[0279] 人工智能规划涉及识别自主的代理必须执行以便实现特定目标的动作序列的问题。代理可以执行的每个功能(服务)被表示为算子。(前置条件和后置条件可以与这些算子相关联。前置条件描述要执行该算子(功能)就必须存在的先决条件。后置条件描述由该算子的执行所触发的环境中的变化——智能环境可能会需要对其做出响应的变化。)在规划术语中，图 7 的“策略合成”对应于计划产生，并且“动作”对应于计划执行。计划产生涉及关于智能环境的服务组成。

[0280] 大量规划器可从 AI 领域获知。参看例如 Howe 的“A CriticalAssessment of Benchmark Comparison in Planning”(Journal of ArtificialIntelligence Research,17:1-33,2002)。的确，存在着专门论述人工智能规划器之间的竞争的年会(参看 ipc<dot>icaps-conference<dot>org)。Amigoni 的“What Planner for Ambient Intelligence Applications?”(IEEESystems,Man and Cybernetics,35(1):7-21,2005)已经评估了用于在智能环境中组成服务的一些规划器。前面提到的 Marquardt 的论文特别考虑了用于智能环境中的服务组成的其他一些规划器，包括 UCPOP、SGP、和黑板(Blackbox)。这些规划器全都通常使用 PDDL (规划领域定义语言，Planning Domain Definition Language)的变型，其中 PDDL 是一种流行的用于对领域和问题进行规划的描述语言。

[0281] Marquardt 在简单的智能环境仿真中评估了不同的规划器，所述简单的智能环境的一部分由图 8 表示，采用 5-20 个装置，每个装置具有两个随机选择的服务和随机选择的目标。数据在模型组件之间以消息的形式沿着指示的线路交换。该仿真中的服务各自具有高达 12 个前置条件(例如，“light_on (灯 _ 开启)”、“have_document_A (具有 _ 文件 _A”、等等)。每个服务也具有各种后置条件。

[0282] 该研究得出的结论是：三个规划器全都令人满意，但是黑板表现最好(Kautz,“Blackbox:A New Approach to the Application of TheoremProving to Problem Solving”,AIPS 1998)。Marquardt 注意到：在目标无法解决的情况下，规划器通常会花费过量的时间来无用地尝试作出满足该目标的规划。该作者得出的结论是：如果该处理在一秒钟内没有产生解决方案，更好的做法是终止规划处理(或者启动不同的规划器)，以便避免浪费资源。

[0283] 尽管来自不同的研究领域，但是本申请人相信当在视觉查询领域中尝试安排服务的组成以达到特定目标时，上述的后一种洞察应该同样适用：如果不能快速地想出令人满意的穿过解答树(或其他规划程序)的路径，那么状态机或许应该将该功能视为无法用可用数据解决，而不是花费更多资源来设法找到解决方案。可以在软件中建立阈值间隔(例如，0.1 秒钟、0.5 秒钟等)，并且可以将计时器与该阈值进行比较，并且如果在达到阈值之前没有找到适合的策略，那么中断尝试找出解决方案。

[0284] 本技术的实施例也可以从万维网服务领域中的工作中吸取经验，所述万维网服务正日益作为复杂的网站的功能组件而被包含进来。例如，旅行网站可以使用一个万维网服务来进行航线预订，使用另一个万维网服务来选择飞机上的座位，并使用另一个万维网服务来从用户的信用卡收费。旅行网站不需要编写这些功能组件；它可以使用由其他方编写

并提供的网络服务的网络。这种利用由其他方在先前完成的工作的模块化方法能加快系统设计和交付。

[0285] 这种系统设计的特定方式有各种名字，包含面向服务架构(SOA)和面向服务计算。尽管这种设计风格为开发者节省了编写用于执行各个单独的组分操作的软件所需的努力，但是仍然存在着这样的任务：决定要使用哪些网络服务、以及使得向这些服务提交数据和从这些服务收集结果协调配合起来。解决这些问题的各种方法是已知的。参看例如 Papazoglou 的“Service-Oriented Computing Research Roadmap”(Dagstuhl Seminar Proceedings 05462,2006)和 Bichler 的“ServiceOriented Computing”(IEEE Computer,39:3,2006 年 3 月，pp.88-90)。

[0286] 服务提供商自然具有有限的提供服务的能力，并且有时候必须处理对超出其能力的请求进行鉴别分类的问题。该领域中的工作包括用于在相互竞争的请求当中进行选择、并根据需求使对服务的收费得到适应的算法。参看例如 Esmaeilsabzali 等人的“Online Pricing for WebService Providers”(ACM Proc.of the 2006Int'l Workshop on EconomicsDriven Software Engineering Research)。

[0287] 本技术的状态机可以采用面向服务计算方案来通过将处理负担的一部分部署到远程服务器和代理，而扩展移动装置的功能性(进行视觉搜索等)。相关的万维网服务可以向一个或更多基于云的中间人处理登记，例如以标准化的(例如 XML)形式说明其服务、输入、和输出。状态机可以在识别实现系统的需要所要进行的服务的过程中向这些中间人咨询。(状态机可以向多个中间人中的一个中间人咨询，以识别处理特定类型服务的中间人。例如，与第一类服务(例如面部识别)相关联的基于云的服务提供商可能会由第一中间人编入目录，而与另一类服务(例如 OCR)相关联的基于云的服务提供商可能会由第二中间人编入目录。)

[0288] 统一描述发现和集成(UDDI)规范定义了一种万维网服务可用来发布有关万维网服务的信息、以及状态机可用来发现有关万维网服务的信息的方式。其他适合的标准包括：使用可扩展标记语言(ebXML)的电子商务，和基于 ISO/IEC 11179 元数据注册(MDR)的那些标准。基于语义的标准(如 WSDL-S 和 OWL-S (下面将提到))允许状态机使用来自语义模型的术语描述期望的服务。随后可以使用推理技术(如描述逻辑推理)来找出状态机所提供的描述与不同的万维网服务的服务能力之间的语义相似性，从而允许状态机自动地选择适合的万维网服务。(如在其它地方提到的那样，可以使用反向拍卖模型来例如从几个适合的万维网服务当中进行选择。)

[0289] 直觉计算平台(ICP)状态机——并行处理

[0290] 为了将系统维持在响应状态，ICP 状态机可以监视图 9 中概念性示出的各种级别的并行处理(类似于认知)。四个这样的级别以及它们各自范围的大致摘要是：

[0291] • 反射——没有用户或云交互

[0292] • 有条件的——基于意图；最少的用户交互；使云卷入其中

[0293] • 直觉的或“浅的解决方案”——基于在装置上得出的解决方案，由用户交互辅助并且被告知以意图和历史的解释

[0294] • “深层解决方案”——通过与用户和云的会话得出的完整解决方案。

[0295] 图 10 进一步详述了这四个处理级别，其与执行视觉查询相关联，通过系统的不同

方面被组织起来，并且识别相互关联的元素。

[0296] 反射处理通常只需花费一秒钟的若干分之一来执行。一些反射处理可能会很少刷新(例如，摄像机分辨率是什么)。一些反射处理(如评估摄像机焦点)可能会一秒钟重复发生几次(例如，一次或两次，多达几十次——诸如对于每帧拍摄而言)。通信组件可以简单地检查网络连接是否存在。原型小玩意(模拟小玩意)可以基于对图像分割的总体评估(例如，存在亮点吗？)来放置。可以注意基本图像分割的时间方面，诸如流动——从一个帧到下一个帧，红色斑点向右移动了 3 个像素。所拍摄的 2D 图像被呈现在屏幕上。在该级别上通常不涉及用户，除了例如对用户输入(类似被轻拍的小玩意)进行确认。

[0297] 有条件的处理需花费更长的时间来执行(尽管通常少于一秒钟)，并且可能会例如每半秒钟左右就刷新一次。许多这些处理涉及背景环境数据和对用户输入采取的行动。这些处理包括：检索用户上一次在相似的背景环境情况下采取了什么动作(例如，用户经常在走着去上班的途中去星巴克)，对用户关于期望的冗长度的指令做出响应，基于当前装置状态(例如，飞机模式、节能模式)配置操作，执行基本的定向操作，确定地理位置，等等。激活或者准备激活与当前图像和其他背景环境相关地出现的识别代理(例如，图像看起来有点像文本，因此准备用于进行可能的 OCR 识别的处理)。识别代理可以注意也在运行的其他代理，并把结果发布到黑板以供它们使用。表示来自某些操作的输出的小玩意出现在屏幕上。执行与基于云的资源的信号交换，以使数据通道准备好供使用，并且检查通道的质量。对于涉及基于云的拍卖的处理，可以把这样的拍卖连同(例如，关于用户的)相关背景信息一起公告，使得不同的基于云的代理可以决定是否参与、并做出任何需要的准备。

[0298] 直觉处理仍需花费更长的时间执行，虽然主要在装置本身上。这些处理通常涉及在识别代理的工作过程中支持识别代理——组成所需的关键字向量，呈现相关联的用户界面，调用相关功能，响应并平衡相互竞争的对资源的请求，等等。系统辨别什么样的语义信息是用户期望或者可能期望的。(如果用户在星巴克通常会对纽约时报的头版进行成像，那么可以启动与 OCR 相关联的操作，而无需用户请求。同样，如果对类似文本的图像的呈现在历史上曾促使用户请求 OCR 和到西班牙文的翻译，那么可以启动这些操作——包括准备好基于云的翻译引擎。)可以识别并采用相关的本体论。由识别代理发布的所输出的小玩意可以根据装置对所拍摄景象的理解而被几何地重新映射，并且可以应用 3D 理解的其他方面。规则引擎可以监视外部数据通道上的通信量，并相应地做出响应。快速的基于云的响应可以被返回并呈现给用户——常常用菜单、窗口、和其他交互图形控制。在该级别上也可能会涉及到第三方功能库。

[0299] 最终的深层解决方案在时间安排方面是无限制的——它们可以从数秒钟延伸到数分钟或者更长，并且通常会涉及云和 / 或用户。然而，直觉处理通常会涉及各个单独的识别代理，深层解决方案可以基于来自几个这样的代理的输出，从而通过例如关联性进行交互。社交网络输入也可以牵扯到该处理中，例如使用关于同年龄群体、用户尊敬的时尚带头人、他们的历史等的信息。在外部的云中，精巧的处理可能正在开展(例如，在远程代理相互竞争提供服务给装置时)。一些早先提交给云的数据可能会引起对更多或更好的数据的请求。早先苦于缺少资源的识别代理现在可以被分配它们想要的全部资源，因为其他情况已经清楚表明对它们的输出的需求。与自由女神像相邻的渴望得到的 10x20 像素块被授给幸运的小玩意提供商，该小玩意提供商已经给轻拍那里的用户安排了愉快的交互体验。可以

建立去往云的定期的数据流，以便提供正在进行的基于云的对用户期望的满足。在该操作阶段可以由于视觉搜索而启动另外一些处理(许多是交互性的)，例如建立Skype会话，察看YouTube演示视频，把经OCR的法文菜单翻译成英文，等等。

[0300] 在装置启动(或者在该装置的其它操作阶段)时，该装置可以显示与该装置已经获得并且准备好应用的识别代理的一些或全部相对应的小玩意。这类似于在初次启动时汽车的仪表板上的全部警示灯都发光，表明警示灯在需要的情况下能够工作的能力(或者类似于在多玩家在线游戏中显示一玩家收集到的财宝和武器——用户在与龙战斗的过程中可从中领取的工具和资源，等等)。

[0301] 应认识到的是，该方案只是例证性的。在其他实现方案中，自然可以使用其他方案。

[0302] 自顶向下和自底向上；延迟激活结构

[0303] 应用程序可以以各种方式启动。一种方式是通过用户指令(“自顶向下”)。

[0304] 大多数应用程序需要某一组输入数据(例如，关键字向量)，并产生一组输出数据(例如，关键字向量)。如果用户指示系统启动一应用程序(例如，通过轻拍小玩意，与菜单交互，做手势，等等)，那么系统可以通过识别需要什么样的输入(诸如通过建立“需要的关键字向量”列表或树)而开始。如果全部所需的关键字向量都存在(例如，在黑板上，或者在“所存在的关键字向量”列表或树中)，那么可以执行该应用程序(或许呈现光亮的小玩意)并生成相应的输出数据。

[0305] 如果并不是全部的所需关键字向量都存在，那么可以显示与该应用程序相对应的小玩意，但是仅模糊地显示该小玩意。可以参考关键字向量输出的反向目录以识别其他应用程序，所述其他应用程序可以被运行以便提供所需的关键字向量作为用户启动的应用程序的输入。所述其他应用程序所需的全部关键字向量可以添加到“所需的关键字向量”中。处理继续进行，直到所述其他应用程序所需的全部关键字向量都处于“所存在的关键字向量”列表中。随后运行这些其他应用程序。所有它们产生的输出关键字向量被输入“所存在的关键字向量”列表中。每一次处于顶部级别的应用程序所需的另一关键字向量变得可用时，该应用程序的小玩意可以变亮。最终，全部必需的输入数据变得可用，并且由用户启动的应用程序得以运行(并且光亮的小玩意可以宣告该事实)。

[0306] 使应用程序得以运行的另一种方式是“自底向上”——由其输入数据的可用性触发。不是用户调用应用程序、然后等待必需的数据，而是使该处理反向。数据的可用性驱动应用程序的激活(并且常常会驱动对该应用程序的随后选择)。相关的工作可从绰号为“延迟评估”或“延迟激活”的技术获知。

[0307] 延迟激活结构的一种特定的实现方案可从人工智能领域吸取经验，即产生式系统架构。产生过程通常具有两个部分——条件(如果)和动作(则)。这些产生过程可以采取存储规则的形式(例如，如果存在椭圆形，则检查椭圆形内的大部分像素是否具有肤色颜色)。条件可以具有以逻辑组合形式组合起来的若干个元素(例如，如果存在椭圆形、并且如果该椭圆形的高度是至少50像素，则…)；然而，这样的规则通常可以分解为一系列较简单的规则，这在有时是优选的(例如，如果检测到椭圆形，则检查该椭圆形的高度是否是至少50像素；如果该椭圆形的高度是至少50像素，则…)。

[0308] 对照工作存储器来评估所述规则，其中所述工作存储器是表示求解过程的当前状

态的存储库(例如,黑板数据结构)。

[0309] 当陈述一条件的规则得到满足(匹配)时,通常会执行动作,有时会进行盘算。例如,如果若干个条件得到满足,那么系统必须进一步盘算以决定按照什么样的顺序来执行这些动作。(在一些情况下,执行一个动作可能会改变其他匹配条件,使得不同的结果取决于盘算的结论如何被决定而产生。进行盘算的方法包括:例如,基于所述规则在规则数据库中被列出的顺序来执行匹配的规则,或者参考赋予不同规则的不同优先级来执行匹配的规则。)

[0310] 这些方案有时被称为匹配/盘算(或评估)/执行方案(参看 Craig 的"Formal Specifications of Advanced AI Architectures"(Ellis Horward,Ltd.,1991))。在一些情况下,"匹配"步骤可以通过用户按压按钮来满足,或者由处于自底向上的形态的系统满足,或者可以由并未明确联系到被感测内容的某个其它条件满足。

[0311] 如上所述,条件规则启动该处理——必须被评估的标准。在本情况下,条件规则可能会涉及某一输入数据的可用性。例如,可以通过把当前"所存在的关键字向量"树与系统上安装的处于顶部级别的应用程序的完整列表进行比较,来定期地激活"自底向上"处理。如果某一应用程序的输入要求中的任何一个都已存在,那么可以将该应用程序投入执行。

[0312] 如果某一应用程序的输入要求中的一些(但不是全部都)已经存在,那么可以在适当的显示区域中以表明该应用程序的全部输入距离完全得到满足的接近度的亮度来显示相应的小玩意。一旦该应用程序的全部输入都得到满足,该应用程序可以启动而无需用户输入。然而,许多应用程序可以具有"用户激活"输入。如果用户轻拍了小玩意(或者如果另一用户界面装置接收到用户动作),那么该应用程序被切换到自顶向下的启动模式,从而启动其他应用程序(如上所述)以便搜集剩余的宣称输入数据,使得处于顶部级别的应用程序随后可以运行。

[0313] 以相似的方式,已经有一些输入(并非全部输入)可用的应用程序可以由具体情况(如背景环境)转变成自顶向下的激活方式。例如,用户在某些条件下激活某一特征的历史模式可以充当推断出的用户意图,从而预示着当这些条件再发生时应该激活该特征。即使有必不可少的输入不可用,但是如果推断出的用户意图足够有说服力,那么这样的激活仍可以发生。

[0314] (在一些实现方案中,由于要评估大量规则,常规的产生式系统技术可能是繁重的。可以采用优化,例如用于确定哪些规则的条件得到满足的通用 trie 模式匹配方法。参看例如 Forgy 的"Rete:A FastAlgorithm for the Many Pattern/Many Object Pattern Match Problem"(Artificial Intelligence,Vol.19,pp 17-37,1982)。)

[0315] 在类似上述方案的方案中,仅把资源应用于准备好运行或者几乎准备好运行的功能。当以适当的输入数据的可用性评价各功能时,各功能被机会主义地投入执行。

[0316] 有规律地执行的图像处理

[0317] 一些用户期望的操作将总是过于复杂以致于无法由便携式系统单独执行;必须要涉及到云资源。相反地,存在着一些图像相关操作,便携式系统应该能够执行这些图像相关操作而无需使用任何云资源。

[0318] 为了使后一种操作能够得到执行并且使前一种操作更容易得到执行,系统设计者可以规定一组无需由功能或用户请求就例行公事地对所拍摄图像执行的基线图像处理操

作。这些有规律地执行的背景功能可以提供其他应用程序可能会用作输入的素材(以关键字向量表示的输出数据)。这些背景功能中的一些也可以服务于另一目的:使图像相关信息得到标准化/精炼,以便能高效地传递给其他装置和云资源并由其他装置和云资源利用。

[0319] 第一类这种有规律地执行的操作通常取一个或更多图像帧(或其一部分)作为输入,并且产生图像帧(或部分帧)关键字向量作为输出。示例性操作包括:

[0320] • 遍及图像的(或遍及感兴趣区域的)采样或插值:输出图像可以不具有与源图像相同的尺寸,像素浓度也不必相同。

[0321] • 像素重新映射:输出图像具有与源图像相同的尺寸,尽管像素浓度不需要相同。每个源像素被独立地映射

[0322] ○实例:阈值处理,"假色(false color)",用范例值替换像素值

[0323] • 本地操作:输出图像具有与源图像相同的尺寸,或者以标准方式被增大(例如,增加黑色图像边界)。每个目的地像素由相应的源像素周围的固定尺寸的本地邻居限定

[0324] ○实例:6x6 的 Sobel 垂直边缘,5x5 的线边缘量值,3x3 的本地最大值、等等

[0325] • 空间重新映射:例如,校正透视图或曲率"变形"

[0326] • FFT 或其他映射到新空间中的"图像"

[0327] • 图像算术:输出图像是输入图像的总和、最大值等

[0328] ○序列平均:每个输出图像对 k 个连续的输入图像求平均

[0329] ○序列 (op)ing:每个输出图像是 k 个连续的输入图像的函数

[0330] 第二类这样的背景操作处理一个或更多输入图像(或其一部分),以便产生由一系列 1D 或 2D 区域或结构构成的输出关键字向量。该第二类中的示例性操作包括:

[0331] • 长线提取:返回一系列提取出的直线段(例如,以斜截式表示,带有端点和长度)

[0332] • 一系列点,长线在这些点处相交(例如,以行/列格式表示)

[0333] • 椭圆形寻找器:返回一系列提取出的椭圆形(在这种和其他情况下,被注意到的特征的位置和参数包含在该列表中)

[0334] • 圆柱寻找器:返回一系列可能的 3D 圆柱(使用长线)

[0335] • 基于直方图的斑点提取:返回一系列图像区域,这些图像区域通过它们的局部直方图而被区别开

[0336] • 基于边界的斑点提取:返回一系列图像区域,这些图像区域通过它们的边界特征而被区别开

[0337] • 斑点"树",在该斑点"树"中每个组分斑点(包括完整图像)具有完全包含在该组分斑点中的分离的子斑点。可以携带有用的缩放不变(或至少是抗缩放)信息

[0338] ○实例:以多个阈值对一图像进行阈值处理的结果

[0339] • 精确的边界,例如,经阈值处理的斑点区域的那些精确边界

[0340] • 不明显的边界,例如,一系列提供密集度适当的区域边界的边缘或点,但是可以具有小缺口或不一致性,不同于经阈值处理的斑点的边界

[0341] 第三类这种例行的正在进行的处理产生表格或直方图作为输出关键字向量数据。该第三类中的示例性操作包括:

[0342] • 色调、强度、颜色、亮度、边缘值、纹理等的直方图

[0343] • 表示例如 1D 值的特征同现的 2D 直方图或表格:(色调,强度),(x 强度,y 强度),

或某种其它配对

[0344] 第四类这种默认图像处理操作由对共同的非图像对象执行的操作构成。该第四类中的示例性操作包括：

[0345] • 分割 / 融合：输入斑点列表产生新的不同的斑点列表

[0346] • 边界修复：输入斑点列表产生一系列具有更平滑边界的斑点

[0347] • 边界跟踪：输入斑点列表的序列产生一系列斑点序列

[0348] • 归一化：图像直方图和基于直方图的斑点的列表返回用于对图像进行重新映射的表格(或许重新映射到“区域类型”值和“背景”值)

[0349] 自然，上述操作只是示例性的。存在着许多许多其他低级别操作可以例行公事地执行。然而，上述的相当大的一组类型通常是有用的，需要相当小的功能库，并且可以在通常可利用的 CPU/GPU 要求内实现。

[0350] 通过背景环境触发的图像处理：条形码解码

[0351] 前面的讨论提到了各种操作，系统可以例行公事地执行这些各种操作以提供能够充当各种更专门的功能的输入的关键字向量数据。那些更专门的功能可以以自顶向下的方式(例如通过用户指令)或者以自底向上的方式(例如通过全部宣称数据的可用性)启动。

[0352] 除了刚刚详述的操作之外，系统还可以启动处理来基于背景环境生成其他关键字向量。

[0353] 举例来说，考虑位置。通过参考地理位置数据，装置可以确定用户处于食品杂货店。在这种情况下，系统可以自动地开始执行附加的图像处理操作，这些附加的图像处理操作生成可能对通常与食品杂货店相关的应用程序有用的关键字向量数据。(这些自动触发的应用程序继而可以调用为所触发的应用程序提供输入所需的其他应用程序。)

[0354] 例如，在食品杂货店中用户预期会遇到条形码。条形码解码包括两个不同的方面。第一个方面是在视场内寻找条形码区域。第二个方面是对识别出的区域中的线符号进行解码。与前一方面相关联的操作可以在用户被确定为处于食品杂货店(或其他零售机构)时被例行公事地采取。即，前面详述的例行公事地执行的一组图像处理操作，通过追加由食品杂货店中的用户所处位置触发的、另外一组通过背景环境触发的操作而被暂时扩大。

[0355] 可以通过分析图像的灰度级版本以识别出在水平方向上具有高图像对比度并且在垂直方向上具有低图像对比度的区域，来进行寻找条形码的操作。因此，当处于食品杂货店中时，系统可以扩大例行公事地执行的图像处理操作的目录，以便还包括对所定位的水平灰度级图像对比度的度量的计算(例如主题像素的任意一侧的 2-8 个像素)。(一种这样的度量是对相邻像素的值的差分的绝对值求和。)该对比度信息帧(或向下采样的帧)可以包含关键字向量——被标注以关于其内容的信息，并且被发布给其他处理以供查看和使用。类似地，系统可以计算所定位的垂直灰度级图像对比度，并发布这些结果作为另一关键字向量。

[0356] 系统可以通过对图像中的每个点，从计算出的局部水平图像对比度的度量中减去计算出的局部垂直图像对比度的度量，来进一步处理这两个关键字向量。通常，该操作在取强烈正值的点和取强烈负值的点处产生混乱的一帧数据。然而，在条形码区域中，混乱程度要小得多，在条形码区域范围内具有强烈正值。该数据也可以发布给其他处理以供查看，作为在用户处于食品杂货店中时例行公事地产生的又一个(第三个)关键字向量。

[0357] 第四个关键字向量可以通过应用阈值处理操作(只识别具有高于目标值的值的那些点),从第三个关键字向量中产生。该操作因此识别出图像中看上去似乎在特征上可能类似于条形码的点,即水平对比度强并且垂直对比度弱的点。

[0358] 第五个关键字向量可以通过应用连通分量分析(限定看上去似乎在特征上可能类似于条形码的点所构成的区域(斑点)),从第四个关键字向量中产生。

[0359] 第六个关键字向量可以通过第五个关键字向量产生——由三个值构成:最大斑点中的点的数目;以及该斑点的左上角和右下角的位置(用距离图像帧的最左上角处的像素的行偏移和列偏移来定义)。

[0360] 这六个关键字向量是预期会产生的,而不需要用户明确地请求它们,这仅仅是因为用户处于与食品杂货店相关联的位置。在其他背景环境中,这些关键字向量通常将不会被产生。

[0361] 这六个操作可以包括单个识别代理(即,条形码定位代理)。或者这六个操作可以是较大的识别代理(例如,条形码定位/读取代理)的一部分,或者这六个操作可以是一些子功能,这些子功能单独地或者组合起来可以构成这六个操作自身的识别代理。

[0362] (条形码读取处理中的更少的操作或另外的操作可以类似地得到执行,但是这六个操作举例说明了要点。)

[0363] 条形码读取器应用程序可以处于装置上加载的那些应用程序中。当处于食品杂货店中时,条形码读取器应用程序可以在非常低的操作级别上活跃起来——仅仅检查上面提到的第六个关键字向量中的第一参数以查看其值是否超过例如15,000。如果该测试得到满足,那么条形码读取器可以指示系统呈现模糊的条形码——在由该第六个关键字向量的第二和第三参数标识的斑点拐角点位置之间的中途帧中的位置处指示出小玩意。该小玩意告诉用户:装置已经感测到可能是条形码的某个东西以及它出现在帧中的位置。

[0364] 如果用户轻拍该模糊的小玩意,那么这会(自顶向下地)启动对条形码进行解码所需的其它操作。例如,提取第六个关键字向量中标识的两个拐角点之间的图像区域,从而形成第七个关键字向量。

[0365] 然后接着发生一系列进一步的操作。这些操作可以包括用低频边缘检测器对提取出的区域进行滤波,并使用霍夫变换来搜索接近垂直的线。

[0366] 然后,对于经滤波的图像中的每一行,通过与用作引导标记的估算出的条形码的左右边缘的关联性来识别开始、中间和末端条形码图案的位置。然后,对于每个条形码数字,确定数字在该行中的位置,并且使该行中的该位置处的像素与可能的数字代码相关以便确定最佳匹配。对每个条形码数字重复该过程,从而产生候选条形码有效载荷。然后在来自该行的结果上执行奇偶校验数字测试,并且使该有效载荷的出现计数值加1。然后对经滤波的图像中的另外的几个行重复这些操作。然后,出现计数值最高的有效载荷被认为是正确的条形码有效载荷。

[0367] 在该点上,系统可以明亮地照亮条形码的小玩意——表明数据已经令人满意地得到提取。如果用户轻拍明亮的小玩意,那么装置可以呈现动作菜单,或者可以启动与解码出的条形码相关联的默认动作。

[0368] 尽管在刚刚描述的方案中,系统在产生第六个关键字向量之后停止其例行操作,但是系统也可以做进一步的处理。然而,由于资源限制,在每次机会来临时(例如当第六个

关键字向量中的第一参数超过 15,000 时）都进一步进行处理可能是不实际的。

[0369] 在一个备选方案中，系统可以例如每三秒钟一次地进一步进行处理。在每个三秒钟间隔期间，系统监视第六个关键字向量的第一参数，寻找：（1）超过 15,000 的值；和（2）超过该三秒钟间隔中的全部先前值的值。当满足这些条件时，系统可以对帧进行缓冲，或许对任何先前缓冲的帧进行盖写。在三秒钟间隔结束时，如果缓冲了一帧，那么该帧具有该三秒钟间隔中的任意第六个关键字向量的第一参数的最大值。随后系统可以从该帧中提取感兴趣区域、应用低频边缘检测器、使用霍夫程序寻找线条、等等——自始至终在成功解码了有效的条形码有效载荷的情况下明亮地照亮小玩意。

[0370] 作为机械地每三秒钟就尝试完成条形码读取操作这一方案的替代，系统可以机会主义地进行处理——当中间结果特别有前途时才进行处理。

[0371] 例如，尽管条形码读取处理可以在每当感兴趣区域中的点的数目超过 15,000 时就继续进行处理，但是该值是条形码读取尝试可能会富有成效的最小阈值。成功地读取条形码的概率随着所述点区域变得越大而增大。因此，作为每三秒钟就进一步进行解码处理一次这一方案的替代，进一步的处理可以由第六个关键字向量的第一参数中的值超过 50,000（或 100,000 或 500,000 等）的事件触发。

[0372] 这样大的值表明，明显的条形码占据了摄像机观察到的帧的相当大一部分。这暗示了用户有意做出的动作——拍摄条形码的高质量的视图。在这种情况下，可以启动条形码读取操作的剩余部分。这向装置的行为提供了一种直觉感：用户明显意图对条形码成像，并且系统在没有任何其它指令的情况下就启动完成条形码读取操作所需的进一步操作。

[0373] 以同样的方式，系统可以根据特别适合于某一类型操作的图像信息的可用性来推断：用户意图或者将会受益于所述某一类型的操作。系统于是可以进行该操作所需的处理，从而产生直觉响应。（类似文本的图像可以触发与 OCR 处理相关联的操作；类似面部的特征可以触发与面部识别相关联的操作，等等。）

[0374] 这可以在不考虑背景环境的情况下完成。例如，装置可以周期性地检查关于当前环境的某些线索，例如在可能看见条形码的情况下，偶而检查图像帧中的水平灰度级对比度与垂直灰度级对比度的比率。尽管这些操作可能不是例行加载的或者可能不是由于背景环境而加载的，但是无论如何这些操作可以例如每五秒钟左右一次地进行，因为计算成本小并且对视觉上有用的信息的发现可以由用户评价。

[0375] 返回到背景环境，就像系统由于用户的位置在食品杂货店而自动地采取一组不同的背景图像处理操作那样，系统也可以类似地基于其他情况或背景环境而使其例行发生的处理操作的集合得到适应。

[0376] 一个因素是历史（即，用户的历史或用户的同等社会阶层的历史）。通常，我们在自己家里不会使用条形码读取器。然而，图书收藏者可能会通过读取新图书的 ISBN 条形码来对家庭图书馆中的新图书编目录。第一次用户在家里把该装置用于这种功能时，生成上面提到的第一至第六个关键字向量的操作可能会需要以自顶向下的方式启动——因为用户通过装置的用户界面表示出对读取条形码的兴趣而启动。第二次也同样如此。然而，合乎期望的是，系统注意到（1）处于特定位置（即家里）的用户、以及（2）条形码读取功能的激活的反复的同现。在这样的历史模式已经建立之后，每当用户处于家位置时，系统就可以例行地启动上面提到的第一至第六个关键字向量的产生。

[0377] 系统可以进一步辨别出：用户只在晚上在家里激活条形码读取功能。因此，时间也可以是触发某些图像处理操作的自动启动的另一背景环境因素，即这些关键字向量是当用户在晚上在家里时产生的。

[0378] 社会信息也可以提供对数据的触发。用户可能仅把给图书编目录作为孤独时的追求。当配偶在家时，用户可能并不给图书编目录。配偶是否在家里可以以各种方式感测到。一种方式是通过来自配偶的手机的蓝牙无线电信号广播。这样，当(1)用户在家里、(2)在晚上、(3)用户的配偶不在附近时，可以自动地产生条形码定位关键字向量。如果配偶在家、或者如果是白天、或者如果用户远离家(和食品杂货店)，那么系统可以不例行地产生与条形码定位相关联的关键字向量。

[0379] 可以编辑并利用用户行为的贝叶斯或其它统计模型以检测反复出现的情况的这种同现，并且随后可以使用所述贝叶斯或其它统计模型以基于此来触发动作。

[0380] (在这一点上，微处理器设计中的分支预测方面的科学可以是有教益的。当代的处理器包括：可以包含许多阶段的管道——需要处理逻辑取回提前 15 或 20 步使用的指令。错误的猜测可能会需要注满管道——引起显著的性能损失。因此，微处理器包括分支预测寄存器，其跟踪条件分支在例如最后 255 次是如何得到解决的。基于这种历史信息，极大地提高了处理器的性能。以类似的方式，跟踪用户和代理这两者(例如，用户的同等社会阶层，或同等人口统计群体)对装置进行利用的历史模式、并基于这种信息来定制系统行为，可以提供重要的性能改善。)

[0381] (下面进一步讨论的)音频线索也可能会涉及某些图像处理操作的自动触发。如果听觉线索暗示用户在户外，那么可以启动一组额外的背景处理操作；如果该线索暗示用户在开车，那么可以启动一组不同的操作。如果音频具有电视音轨的特点、或者如果音频暗示用户在办公室环境中，也同样如此。在系统中加载并运行的软件组件因此可以在预期到在该特定环境中可能会遇到的刺激或者用户可能会请求的操作的情况下自动地自适应。(类似地，在应用不同的音频处理操作以生成不同的音频功能所需的关键字向量的听觉装置中，从视觉环境中感测到的信息可能会指示出这样的背景环境，该背景环境规定通常不会运行的某些音频处理操作的启动。)

[0382] 环境线索也可以引起某些功能被选择、启动、或定制。如果装置感测到的环境温度是负十摄氏度，那么用户推测起来可能在冬季的户外。如果指示面部识别(例如，通过用户指令，或者通过其他线索)，那么图像中描绘的任何面部可能会被包裹在帽子和/或围巾中。考虑到面部的某些部分会被遮盖，而不是例如背景环境是炎热的夏天、人的头发和耳朵预期会露出来的情况，因此可以采用一组不同的面部识别操作。

[0383] 其他与系统的用户交互可能会被注意到，并且会导致通常不运行的某些图像处理操作的启动，即使注意到的用户交互不涉及这样的操作。考虑用户通过装置上的万维网浏览器进行查询(例如通过文本或语音输入)以识别附近的餐馆的情况。该查询不涉及摄像机或图像。然而，从这样的交互中，系统可以推断出用户不久将会(1)改变位置，和(2)处于餐馆环境中。因此，系统可以启动在例如(1)导航到新位置、和(2)处理餐馆菜单的过程中可能会有帮助的图像处理操作。

[0384] 可以通过把来自摄像机的图像与沿着用户预期的路线的路边图像(例如，来自 Google Streetview 或其他图像储存库，使用 SIFT)进行图案匹配来对导航进行辅助。除了

从 Google 获得相关图像之外，装置还可以启动与尺度不变特征变换操作相关联的图像处理操作。

[0385] 例如，装置可以对摄像机以不同的缩放状态拍摄的图像帧进行重新采样，从而对每个图像帧产生一关键字向量。对于这些图像帧中的每一个，可以应用高斯差分函数，从而产生进一步的关键字向量。如果处理限制允许的话，那么这些关键字向量可以与模糊滤波器卷积，从而产生更进一步的关键字向量、等等——全都在预期到 SIFT 图案匹配的可能使用的情况下。

[0386] 在预期到要察看餐馆菜单的情况下，可以启动对 OCR 功能性而言难免的操作。

[0387] 例如，尽管默认的一组背景图像处理操作包括用于长边缘的检测器，但是 OCR 需要识别短边缘。因此，可以启动识别短边缘的算法；该输出可以用关键字向量表示。

[0388] 定义闭合轮廓的边缘可以用于识别字符候选斑点。字符的线条可以从这些斑点的位置取得，并且可以应用倾斜校正。从字符斑点的倾斜校正后的线条中，可以辨别出候选单词区域。随后可以应用图案匹配以便识别这些单词区域的候选文本。等等。

[0389] 如前所述，并不是所有这些操作都会在每个经处理的图像帧上执行。可以例行公事地执行某些早期操作，并且可以基于(1)定时触发、(2)迄今为止已处理的数据的有前途的属性、(3)用户指示、或者(4)其他标准来采取进一步的操作。

[0390] 回到食品杂货店的实例，不仅背景环境可以影响所采取的图像处理操作的类型，而且归因于不同类型的信息(图像信息以及其他信息(例如地理位置))的含义也可以影响所采取的图像处理操作的类型。

[0391] 考虑用户的手机在食品杂货店中拍摄了一帧图像的情况。手机可以立即做出响应——提出用户正面对着汤罐。手机可以通过参考地理位置数据和磁力计(罗盘)数据以及所存储的关于该特定店铺的布局的信息(表明摄像机正面对着放汤的货架)来实现该操作。处于初始阶段的小玩意可以例如通过表示食品杂货条目的图标、或者通过文本、或者通过链接的信息，来把该最初的猜测传达给用户。

[0392] 一会之后，在对所拍摄的帧中的像素进行初始处理期间，装置可以辨别出紧接着白色像素斑点的红色像素斑点。参考与食品杂货店背景环境相关联的参考数据源(再一次，或许也依靠地理位置和罗盘数据)，装置可以快速地猜测出(例如，在少于一秒钟的时间内)该条目(最可能)是一罐金宝(Campbell)汤、或者(可能性小一点地)是一瓶调味蕃茄酱。矩形可以叠加到屏幕显示上，从而勾画出装置所认为的对象的轮廓。

[0393] 一秒钟之后，装置可能已经对白色背景上的大字符完成了 OCR 操作，声称是番茄汤(TOMATO SOUP)—进一步证实了 Campbell 汤的假设。在进一步的短暂间隔之后，手机可能已经设法在图像的红色区域中识别出风格化的笔迹“Campbell's”，从而确认对象不是模仿 Campbell 的色彩设计的零售商品牌汤。在进一步的一秒中，手机可能已经解码出附近的罐上可见的条形码，其详述与 Campbell 番茄汤相关的尺寸、批号、生产日期、和 / 或其他信息。在每个阶段，小玩意或链接的信息根据装置对摄像机所指向的对象的精炼的理解而进化。(在任何时间点，用户可以指示装置停止其识别工作(可能通过快速摇动)，从而为其他任务保存电池电力和其他资源。)

[0394] 相反，如果用户在户外(例如通过 GPS 和 / 或明亮的阳光而感测到)，那么手机对紧接着白色像素斑点的红色像素斑点的最初猜测将可能不是 Campbell 汤罐头。而是，手机更

可能会猜测它是美国国旗、或者花、或者一件衣服、或者方格色桌布——再一次通过参考与户外背景环境相对应的信息的数据存储库。

[0395] 直觉计算平台(ICP)背景环境引擎,标识符

[0396] 引用 Arthur C.Clarke 说过的话:“任何足够先进的技术都很难与巫术区别开”。“先进”可以有许多含义,但是为了向移动装置灌输类似于巫术的东西,本说明书将该术语解释为“直觉的”或“智能的”。

[0397] 直觉行为的重要一部分是感测用户可能的意图、然后对用户可能的意图做出响应的能力。如图 11 所示,意图不仅随用户而变,而且随用户过去的历史而变。另外,意图也可以被认为随用户的同等群体的活动及其过去的历史而变。

[0398] 在确定意图的过程中,背景环境是关键。即,在知道例如用户在哪里、用户及其他人上一次在该位置从事了什么活动、等等对辨别用户在当前时刻可能的活动、需要和期望很有价值的意义上,背景环境能够向意图的推断提供信息。这种关于用户行为的自动推理是人工智能的核心目标,并且关于该主题已经有许多著述。(参看例如 Choudhury 等人的“Towards Activity Databases:Using Sensors and Statistical Modelsto Summarize People's Lives”(IEEE Data Eng.Bull,29(1):49-58,2006 年 3 月。)

[0399] 传感器数据(如图像、音频、运动信息、位置、和蓝牙信号)在推断用户可能的活动(或者排除不可能的活动)的过程中很有用。如 Choudhury 提到的那样,这些数据可以提供给一软件模块,该软件模块将传感器信息处理成可以帮助区别各活动的特征。这些特征可以包括:高级别信息(如对环境中的对象的识别,或者附近的人数,等等),或者低级别信息(如音频内容或幅度、图像形状、相关系数等)。从这些特征中,计算模型可以推断出可能的活动(例如,走路、谈话、取咖啡、等等)。

[0400] 合乎期望的是,来自手机的传感器数据被例行地记录,使得历史活动的模式能够得以辨别。继而,用户进行的活动可以被注意到并且与引起这些活动发生的背景环境(并发的背景环境和紧接着这些活动发生之前的背景环境)相关。继而,活动成为借以推断出用户兴趣的素材。所有这样的数据被存储并且充当参考信息,所述参考信息允许手机推断出用户在给定背景环境中可能会参与的可能行为,并且允许手机辨别出在这些环境中哪些用户兴趣可能是相关的。

[0401] 这样的智能可以以基于模板、模型或规则的形式(例如,详述背景环境数据和与所述背景环境数据明显相关的用户行为 / 兴趣(可能带有相关的置信度因子)的复发模式)被编码。给定实时传感器数据,这样的模板可以向便携式装置提供关于预期意图的建议,使得装置能够相应地做出响应。

[0402] 随着更多的体验被记录以及差别更细微的模式能够被辨别出来,这些模板可以不断地得到改进,从而与背景环境的额外方面(例如季节、天气、附近朋友、等等)相关。从专家系统获悉的技术可以被用来实现本技术的这些方面。

[0403] 除了由移动装置传感器提供的大量数据之外,在理解背景环境(并因此理解意图)的过程中有用的其他特征可以从附近对象取得。树会暗示户外背景环境;电视会暗示室内背景环境。一些对象具有相关联的元数据——极大地推进了背景环境理解。例如,用户环境内的一些对象可能会具有 RFID 等。RFID 传递唯一的对象 ID。通常在远程数据存储库中与这些唯一的对象 ID 相关联的是,关于附加有 RFID 的对象的固定的元数据(例如颜色、重

量、所有权、原产地、等等)。因此,胜于尝试仅从像素中推断相关信息,移动装置中的传感器或者移动装置所链接到的环境中的传感器可以感测这些信息载体,获得相关元数据,并使用该信息来理解当前背景环境。

[0404] (RFID 只是示例性的;也可以采用其他方案,例如数字水印法、条形码、指纹法、等等。)

[0405] 因为用户活动是复杂的并且对象数据和传感器数据都不适用于得出明确结论,所以用于推断用户可能的活动和意图的计算模型通常是概率性的。可以使用产生式技术(例如,贝叶斯、隐藏式马可夫、等等)。也可以采用关于组界(class boundary)的辨别技术(例如,后验概率)。因此也可以采用关系概率模型(relational probabilistic model)和马可夫网络模型。在这些方法中,概率也可以取决于用户的社交群体中的其他人的特性。

[0406] 在一个特定方案中,基于与可能存储在云中的(根据用户的历史、或者根据社交朋友或其他群组的历史等等得出的)模板匹配的、本地装置的与背景环境相关的观察来确定意图。

[0407] 通过辨别意图,本技术减小了对刺激的可能响应的搜索空间,并且可以使用本技术来对输入数据进行分割以辨别活动、对象并产生标识符。标识符可以用明显的推导出的元数据来构造。

[0408] 为了有一点后备,期望每个内容对象都得以识别。理想地,对象的标识符在全世界范围内是唯一且持久的。然而,在移动装置视觉查询中,这种理想情况常常是难以达到的(除了在例如对象承载有诸如数字水印之类的机器可读标志的情况下)。尽管如此,在视觉查询会话内,仍期望每个辨别出的对象具有在该会话内唯一的标识符。

[0409] 唯一标识符(UID)的一种可能的构造包括两个或三个(或更多)组分。一个是业务 ID,其可以是会话 ID。(一种适合的会话 ID 是例如由以装置标识符(如 MAC 标识符)作为种子的 PRN 发生器产生的伪随机数。在其他方案中,会话 ID 可以传递语义信息,诸如传感器最近一次从关闭或睡眠状态被激活的 UNIX 时间。)这种业务 ID 用于减小其他标识组分所需的范围,并且帮助使得标识符唯一。这种业务 ID 还将对象标识置于特定会话或动作的背景环境内。

[0410] 标识符的另一组分可以是明确的对象 ID,其可以是前面提到的聚簇 ID。这通常是分配的标识符。(如果聚簇被确定为包含几个明显可识别的特征或对象,那么可以将另外的比特附加到聚簇 ID 上以区分这些特征或对象。)

[0411] 另一组分可以以某种方式从对象或具体情况中取得。一个简单的实例是“指纹”——从对象本身的特征取得的统计上唯一的标识信息(例如,SIFT、图像签名、等等)。另外地或者可替换地,该组分可以由与背景环境、意图、推断出的特征(基本上是可由后续处理使用以帮助确定身份的任何东西)相关的信息构成。该第三组分可以被认为是推导出的元数据、或者与对象相关联的“先兆”。

[0412] 对象标识符可以是这些组分的级联或其他组合。

[0413] 扇形区、等等

[0414] 由系统调用的不同的识别处理可以并行地、或者以循环连续方式工作。在后一种情况下,时钟信号等可以提供借以激活不同的扇形区的步调。

[0415] 图 12 示出由扇形区构成的圆周这样的循环处理方案。每个扇形区表示识别代理

处理或另一处理。箭头指示从一个扇形区进行到另一扇形区。如位于右边的扩大的扇形区所示，每个扇形区可以包括几个不同的阶段或状态。

[0416] 本技术面临的问题是资源限制。如果不存在限制，那么视听装置可以对输入数据的每个帧和序列不断地应用无数资源密集型识别算法，从而在每个帧中检查用户潜在感兴趣的每个条目。

[0417] 在真实世界中，处理是有成本的。该问题可以表述为以下两者之一：动态地识别应该应用于输入数据的处理，以及动态地决定分配给每个处理的资源的类型和数量。

[0418] 在图 12 中，扇形区（识别代理处理）的不同阶段对应于资源消耗的进一步水平。最里面的（尖角的）阶段通常使用最少的资源。累积的资源负担随着扇形区的相继阶段的处理的执行而增大。（尽管每个阶段常常会比在它之前的阶段更加资源密集，但是这不是必需的。）

[0419] 能够实现这种类型的行为的一种方式是通过把识别和其他操作实现为“操作的级联序列”，而不是实现为整体式操作。这样的级联序列常常会涉及具有相对低的开销的初始操作，这些初始操作在被成功完成时可以通过可能会需要更多资源、但是现在只会在初步表明有可能成功后才被启动的操作而得到继续。本技术也可以促进机会主义地用已经可利用的关键字向量替代由一操作通常会使用的相关特征，从而再一次如前面提到的那样减少资源开销。

[0420] 出于讨论的目的，考虑面部识别代理。为了识别面部，应用一系列测试。如果有任何一个测试失败，那么不太可能存在面部。

[0421] （许多处理共用的）初始测试是检查由摄像机产生的图像是否具有任何种类的特征（与例如当处于黑暗的钱包或口袋中时的摄像机输出进行对比）。这可以通过对跨越整个图像范围的像素位置的稀疏采样的灰度级像素值进行简单的直方图分析来完成。如果该直方图分析指出全部的采样像素都具有基本上相同的灰度级输出，那么可以跳过进一步的处理。

[0422] 如果该直方图显示出在像素灰度级值中存在着一定差异性，那么接下来可以检查图像以查找边缘。不具有可辨别的边缘的图像很可能是不可用的图像，例如高度模糊或散焦的图像。如上所示，各种边缘检测滤波器是本领域技术人员所熟悉的。

[0423] 如果找到了边缘，那么面部检测程序可以接着检查是否有任何边缘是弯曲的并且限定出一闭合区域。（在某些实现方案中作为例程背景操作运行的椭圆形寻找器可以允许处理在该步骤开始。）

[0424] 如果有边缘是弯曲的并且限定出一闭合区域，那么可以执行颜色直方图以确定闭合区域内是否有相当大百分比的像素在色调上彼此相似（皮肤构成面部的一大部分）。“相当大”可以表示大于 30%、50%、70%、等等。“相似”可以表示 CIELAB 意义上的距离阈值或角度旋转。可以任选地应用预定义肤色范围内的颜色测试。

[0425] 接着，可以应用阈值处理操作以识别闭合区域内的最黑的 5% 的像素。可以对这些像素进行分析以确定它们是否形成与两个眼睛相一致的群组。

[0426] 这些步骤以类似的方式通过为各候选面部生成本征向量而继续进行。（面部本征向量是根据面部的高维向量空间表示形式的概率分布的协方差矩阵计算的。）如果是这样的话，可以在（本地或远程的）参考数据结构中搜索本征向量的匹配者。

[0427] 如果任何操作产生了否定结果，那么系统可以推断出不存在可辨别的面部，并终止对该帧进行的进一步的面部寻找努力。

[0428] 全部这些步骤可以形成单个扇形区处理中的各阶段。可替换地，一个或更多步骤可以被认为对几个不同的处理都是基本的且有用的。在这种情况下，这些步骤可以不形成专用扇形区处理的一部分，而是可以是分离的。这些步骤可以在一个或更多扇形区处理中实现——循环地由其他代理处理执行并将其结果发布到黑板上(而不管其他代理是否能找到这些结果)。或者这些步骤可以以其他方式实现。

[0429] 在把系统的有限资源应用于不同的正在进行的处理时，检测状态可能会是有用的概念。在每个时刻，每个代理所寻求的目标(例如，识别面部)看起来似乎或多或少可能会达到。即，每个代理可能会具有连续的瞬时检测状态，从非常有前途经过中性状态降至非常令人沮丧。如果检测状态很有前途，那么可以对该努力分配更多的资源。如果其检测状态趋向于很令人沮丧，那么可以分配更少的资源。(在某一点处，可能会达到沮丧的阈值，使得系统终止该代理的努力。)检测状态可以由(分离的、或者包含在代理处理中的)软件例程周期性地量化，所述软件例程根据代理处理所关心的特定参数而被裁制。

[0430] 当调用了代理处理的相继几个阶段时，倾向于会发生增加分配的资源这样的情况(例如，可能在第七阶段发生的 FFT 操作固有地会比可能在第四阶段发生的直方图操作更复杂)。但是除了基础操作复杂性以外，系统也可以对资源分配进行计量。例如，给定的图像处理操作可能在系统的 CPU 或 GPU 上执行。FFT 可能要用 1MB 或 10MB 的便笺式存储器来执行进行计算。一个处理在一些情况下可能被允许使用(更快速响应的)缓存数据存储器，而在另外一些情况下可能仅被允许使用(响应较慢的)系统内存。一个阶段在一种情况下可能被准予访问 4G 网络连接，而在另一种情况下可能仅被准予访问较慢的 3G 或 WiFi 网络连接。一处理可以公布详述这些不同的选项的信息(例如，我能够用该数量的资源做 X ；我能够用该更大数量的资源做 Y ；我能够用该更少数量的资源做 Z ；等等)，所述不同的选项可以被调用以提高处理的有效性或减少处理的资源消耗。可以明确地提供局部执行方案。状态机可以基于各种资源分配因素从这些选项中进行选择。在系统资源的消耗方面，产生最有前途的结果的处理或者提供最有前途的结果的可能性的处理可以被授予特权状态。

[0431] 在另外的方案中，资源分配不仅取决于在实现代理的目标的过程中该代理的状态，而且取决于该代理去往该目标的速度或加速度。例如，如果响应于初始资源努力水平而快速出现很有前途的结果，那么不仅可以应用额外的资源，而且可以应用比很有前途的结果出现的速度慢一些的情况更多的额外资源。因此，资源的分配可以不仅取决于检测状态(或者性能或结果的其他度量)，而且取决于这种度量的一阶或更高阶的导数。

[0432] 相关地，由检测代理处理的一个阶段产生的数据可能会如此有前途，以致于该处理可以向前跳过一个或更多阶段——跳过介于其间的阶段。这可能是这样的情况，例如跳过的阶段并不产生对该处理而言必需的结果，而是只是为了获得关于更进一步阶段的处理的值得程度的更高置信度而进行的。例如，识别代理可以执行阶段 1、2 和 3，然后基于来自阶段 3 的输出的置信度度量来跳过阶段 4 并执行阶段 5 (或者跳过阶段 4 和 5 并执行阶段 6，等等)。再一次，状态机可以基于一处理公布的关于该处理的不同进入阶段的信息来执行这种决策控制。

[0433] 本领域技术人员将认识到的是，这样的方案不同于熟悉的现有技术。先前，不同的

平台提供基本上不同的计算量子(quanta ofcomputing),例如大型机、PC、手机等。类似地,软件被认为是整体式功能块,具有固定的资源需求。(例如,可以取决于内存的可用性而加载或者不加载特定的 DLL。)设计者因此用具有既定尺寸的各功能块拼凑出计算环境。一些功能块适合,而另一些不适合。相异的是,本概念按照不同的进入点和不同的成本来描述任务,使得系统能够就应该在功能能力范围内进入多深做出智能决策。先前,范例是"如果你能够运行该功能,那么你可以运行该功能"。(成本可在事实之后确定。)本模型将范例转变为更类似于"我将买 31 美分的该功能。基于事情的发展情况,可能我以后会买更多。"在本方案中,因此对于执行某些任务而言,提供了多维的选择范围,由此系统可以在考虑到其他任务、当前资源限制及其他因素的情况下做出智能决策。

[0434] 这里描述的方案还允许操作系统预知资源消耗将会如何随时间变化。操作系统可以注意到:例如,有前途的结果很快将在特定识别代理中出现,这不久将导致对该代理分配的资源的增加。操作系统可以认识到:该代理的任务明显即将会令人满意地完成从而满足某些规则的条件——触发其他识别代理、等等。考虑到资源消耗的即将来临的峰值,操作系统可以主动抢先地采取其他步骤,例如将无线网络从 4G 减速到 3G、更积极地提早结束不会产生鼓励性结果的处理、等等。这样的预见度和响应性比与典型分支预测方法(例如,基于对特定分支判定的最后 255 个结果进行的生搬硬套式检查)相关联的预见度和响应性要高得多。

[0435] 正如资源分配和阶段跳跃可以由检测状态促使那样,资源分配和阶段跳跃也可以由用户输入促使。如果用户提供了对特定处理的鼓励,那么可以对该处理分配额外的资源,和/或该处理可以继续进行从而超过这样的点,在该点处该处理的操作可能因为缺乏很有前途的结果而已经自动地被提早结束。(例如,如果前面提到的检测状态连续过程在从 0 分<完全令人沮丧>到 100 分<完全令人鼓舞>的分数范围内变化,并且在处理的分数降到阈值 35 以下的情况下该处理通常会终止操作,那么如果用户提供了对该处理的鼓励,那么该阈值可以降至 25 或 15。阈值的变化量可以与接收到的鼓励量相关。)

[0436] 用户鼓励可以是明确的或暗示的。明确鼓励的实例是用户提供输入信号(例如,屏幕轻拍等),指示应执行特定操作(例如,指示系统处理图像以识别所描绘的人的用户界面命令)。

[0437] 在一些实施例中,摄像机连续拍摄图像——在无需特定用户指令的情况下监视视觉环境。在这种情况下,如果用户启动快门按钮等,那么该动作可以被解释为处理在该时刻成帧的图像的明确用户鼓励的证据。

[0438] 暗示鼓励的一个实例是用户在图像中描绘的人上轻拍。这可能具有作为了解关于该人的更多信息的信号的意图,或者这可能是随机动作。无论如何,这足够使系统对与该图像的该部分相关的处理(例如面部识别)增加资源分配。(其他处理也可以被列入优先,例如识别该人所穿戴的手提包或鞋,以及在通过面部识别进行鉴别之后调查关于该人的事实——诸如通过使用社交网络(例如 LinkedIn 或 Facebook);通过使用 Google、pipl<dot>com、或其他资源。)

[0439] 在决定应该把多少资源增加量应用于不同任务(例如,鼓励量)的过程中可以利用轻拍的位置。如果该人轻拍图像中的面部,那么与用户轻拍图像中该人的鞋的情况相比,可以将更多的额外资源应用于面部识别处理。在后一种情况下,可以对鞋识别处理分配比面

部识别处理更大的资源增加量。(对鞋进行的轻拍也可以启动鞋识别处理,如果该鞋识别处理尚未在进行中的话。)

[0440] 暗示用户鼓励的另一实例是用户将摄像机放置成使得特定主题位于图像帧的中心点。在系统注意到在多个帧构成的时间序列中摄像机被重新定位(将特定主题移动到中心点)的情况下,这是尤其令人鼓舞的。

[0441] 如前所述,该主题可能由几个部分构成(鞋、手提包、面部、等等)。每个这样的部分和帧的中心之间的距离可以被视为与鼓励量成反比。即,处于帧的中心的部分暗示着最大的鼓励量,而其他部分的鼓励量随着距离而相继减小。(数学函数可以使距离与鼓励相关。例如,作为该帧的中心的部分可以在 0 到 100 的范围内具有 100 的鼓励值。处于图像帧的遥远周边的任何部分可以具有 0 的鼓励值。中间位置可以通过线性关系、幂次关系、三角函数或其他方式与鼓励值相对应。)

[0442] 如果摄像机配备有变焦透镜(或数字缩放功能)、并且摄像机注意到在由多个帧构成的时间序列中摄像机镜头被推向特定主题(或者部分),那么这种动作可以被视为对该特定主题 / 部分的暗示用户鼓励。即使没有由多个帧构成的时间序列,表示变焦程度的数据也可以被视为用户对该成帧的主题的兴趣的度量,并且可以数学地转换成鼓励度量。

[0443] 例如,如果摄像机的变焦范围是 1X 至 5X,那么 5X 的变焦可以对应于 100 的鼓励因子,并且 1X 的变焦可以对应于 1 的鼓励因子。中间变焦值可以通过线性关系、幂次关系、三角函数等与鼓励因子相对应。

[0444] 对意图的推断也可以基于图像帧内的特征的取向。用户被认为通常会按照使意图的主题竖直成帧的取向来握持成像装置。通过参考加速计或陀螺仪数据等,装置可以辨别出用户是以拍摄“风景画”模式图像的方式还是以拍摄“肖像画”模式图像的方式握持成像装置,由此可以确定“竖直”定位。图像帧内具有竖直取向的主轴(例如大致的对称轴)的对象比从竖直方向倾斜的对象更可能是用户意图的主题。

[0445] (用于推断图像帧中用户意图的主题的其他线索在专利 6,947,571 中有讨论。)

[0446] 尽管前面的讨论考虑的是非否定的鼓励值,但是在其他实施例中,可以利用否定值,例如结合明确或暗示的用户对特定刺激的冷淡、图像特征与帧中心的遥远性、等等来利用否定值。

[0447] 肯定和否定这两种鼓励都可以由其他处理提供。如果条形码检测器开始感测到处于帧中心的对象是条形码,那么使其检测状态度量增大。然而,这种结论倾向于驳斥处于帧中心的主题是面部的可能性。因此,第一识别代理的检测状态度量的增大可以充当可能与第一代理相互排斥的其他识别代理的否定鼓励。

[0448] 用于多个识别代理的鼓励和检测状态度量可以通过各种数学算法组合起来,以产生混合控制度量。一种混合控制度量是它们的总和,从而在两个代理(且没有否定鼓励值)的情况下产生从 0 到 200 变化的输出。另一种混合控制度量是它们的乘积,从而产生从 0 到 10,000 变化的输出。在不同的识别代理各自的混合控制度量变化时,资源可以重新分配给这些不同的识别代理。

[0449] 识别代理可以取决于应用场合而具有不同的粒度和功能。例如,刚刚讨论的面部识别处理可以是由许多阶段构成的单个扇形区。或者它可以被实现为几个、或几十个相关的更简单的处理——每个更简单的处理都有自己的扇形区。

[0450] 应认识到的是，图 12 中的扇形区识别代理类似于 DLL——被选择性地加载 / 调用以提供期望种类的服务的代码。(确实，在一些实现方案中，可以使用与 DLL 相关联的软件构造，例如在操作系统中可以使用与 DLL 相关联的软件构造来管理代理代码的加载 / 卸载，以将这种功能性的可用性公布给其他软件，等等。基于 DLL 的服务也可以结合识别代理一起使用。)然而，优选的识别代理具有不同于 DLL 的行为。在一个方面中，这种不同的行为可以被描述为调速或者状态跳跃。即，它们的执行和支持它们的资源会基于一个或更多因素(例如，检测状态、鼓励、等等)而变化。

[0451] 图 13 示出图 12 方案的另一视图。该视图清楚地表明不同的处理可能会消耗不同量的处理器时间和 / 或其他资源。(当然，这些处理可以在单处理器系统上实现，或者在多处理器系统上实现。将来，不同的处理器或者多处理器系统的多个“核”可以被分配来执行不同的任务。)

[0452] 有时，识别代理因为缺乏令人满意的资源(或者是处理资源、或者是输入数据、等等)而未能实现其目标。在拥有额外的或更好的资源的情况下，可能会实现该目标。

[0453] 例如，面部识别代理可能会因为摄像机在拍摄图像时倾斜了 45 度而未能识别出图像中所描绘的人的面部。在该角度，鼻子不在嘴的上方，而鼻子在嘴的上方是代理在辨别是否存在面部的过程中可能已经应用的标准。在拥有更多处理资源的情况下，该标准可能被放宽或者去除。可替换地，在来自另一代理(例如定向代理)的结果已经可用的情况下(例如识别图像中的真实地平线的倾斜)，面部可能已经被检测到。获悉地平线的倾斜可以允许面部识别代理以不同的方式理解“上方”——这种不同的方式允许该面部识别代理识别出面部。(类似地，如果先前拍摄的帧或后来拍摄的帧得到了分析，那么可能已经辨别出面部。)

[0454] 在一些方案中，当其他资源变得可用时，系统对输入的刺激(例如，图像)进行进一步的分析。举个简单的情况来说，当用户将手机放入钱包中、并且摄像机传感器变黑或者无望地散焦时(或者当用户将手机放在桌上使得该手机注视固定的景象(或许是桌子或天花板)时)，软件可以重新激活先前未能实现其目标的代理处理，并对该数据进行重新考虑。在不用分心处理一连串输入的移动图像和相关联的资源负担的情况下，这些代理现在可能能够实现它们的初始目标，例如识别先前错过的面部。在完成该任务期间，系统可以再调用来自其他代理处理的输出数据——既包括在主题代理最初运行时可用的那些输出数据，又包括直到主题代理终止之后才变得可用的那些结果。这些其他数据可以帮助先前未成功的处理实现其目标。(可以对在手机先前进行的操作期间收集的“垃圾”进行复查以找出在该代理运行的初始处理环境中被忽略的或尚未变得可用的线索和有帮助的信息。)为了减少这种“事后思索”操作期间的电池电力耗尽，手机可以切换到省电状态，例如禁用某些处理电路、降低处理器时钟速度、等等。

[0455] 在相关方案中，未能实现其目标就在手机上结束的处理中的一些或全部可以在云中继续进行。手机可以把未成功的代理处理的状态数据发送给云，从而允许云处理器在手机停止的地方继续进行分析(例如，算法步骤和数据)。手机也可以向云提供来自其他代理处理的结果——包括当未成功的代理处理被结束时尚不可用的那些结果。再一次，也可以将数据“垃圾”作为可能的资源提供给云，以防先前丢弃的信息在云的处理中有了新的相关性。云可以对全部这些数据执行搜集操作——设法找出手机系统可能已经忽略的有用信息

或含义。这些结果在被返回给手机时可以继而使手机重新评估它过去或者现在正处理的信息，从而有可能允许它辨别出已经错过的有用信息。(例如，在云的数据搜集处理中，云可能会发现地平线似乎倾斜了 45 度，从而允许手机的面部识别代理识别出已经错过的面部。)

[0456] 尽管上面的讨论集中在识别代理方面，但是相同的技术也可以应用于其他处理，例如对识别而言起辅助作用的处理(诸如建立取向或背景环境等)。

[0457] 关于限制的更多说明

[0458] 图 14 是描绘可在某些实施例中采用的本技术的某些方面的概念视图。图的顶部示出充满识别代理(RA)服务的漏斗，这些识别代理服务大多数都可与用作该服务的输入的一个或更多关键字向量相关联地得到运行。然而，系统限制不允许全部这些服务都得到执行。因此，漏斗的底部被图解地显示为通过限制因素来控制，从而允许或多或少的服务取决于电池状态、CPU 上的其他需求、等等而被启动。

[0459] 允许运行的那些服务被显示在漏斗的下方。在这些服务执行时，它们可以将临时结果或最终结果发布到黑板上。(在一些实施例中，这些服务可以将输出提供给其他处理或数据结构，诸如提供给 UI 管理器、提供给另一识别代理、提供给检查跟踪(audit trail)或其他数据存储库、作为信号提供给操作系统——以便例如使状态机前进、等等。)

[0460] 一些服务运行到完成并且终止(在图中通过单个删除线示出)，从而释放出允许其他服务得到运行的资源。其他服务在完成之前被停止(通过双删除线示出)。这可能会由于各种原因而发生。例如，来自该服务的临时结果可能不是很有前途(例如，现在椭圆形似乎更可能是汽车轮胎而不是面部)。或者系统限制可能会变化——例如，因为缺乏资源而需要终止某些服务。或者其他更有前途的服务可能变得准备好运行，从而需要重新分配资源。尽管在图 14 的图解中未绘出，但是来自被停止的处理的临时结果可以发布到黑板上——或者在这些处理的操作期间，或者在它们被停止的时间点。(例如，尽管在椭圆形看起来更像车辆轮胎而不是面部的情况下面部识别应用程序可以终止，但是车辆识别代理可以使用这样的信息。)

[0461] 发布到黑板上的数据以各种方式被使用。一种方式是触发小玩意的屏幕显示，或者是为其他用户接口需求服务。

[0462] 还可以使得来自黑板的数据可用作对识别代理服务的输入，例如用作输入关键字向量。另外，黑板数据可以表示让新服务运行的原因。例如，黑板上所报告的椭圆形的检测可以表示 :面部识别服务应该运行。黑板数据还可以增大已经在(概念上的)漏斗中等待的服务的相关度分数，从而使得更可能让该服务运行。(例如，椭圆形实际上是车辆轮胎的指示可以将车辆识别处理的相关度分数增大到使该代理处理得到运行的分数。)

[0463] 相关度分数概念在图 15 中示出。一数据结构维持待运行的可能服务的列表(类似于图 14 中的漏斗)。对每个可能的服务示出了相关度分数。这是执行该服务的重要性的相对指示(例如在 1-100 的范围内)。该分数可以是多个变量的函数——取决于特定服务和应用，所述多个变量包括在黑板上找到的数据、背景环境、表达出的用户意图、用户历史等。相关度分数通常会随时间而变化，在更多数据变得可用时变化、在背景环境变化时变化、等等。正在进行的处理可以基于当前状况来更新相关度分数。

[0464] 一些服务可能被评分为高度相关，然而这些服务需要的系统资源比所能提供的结果更多，因此并未运行。另外一些服务可能仅被评分为弱相关，然而这些服务在资源消耗方

面可能如此适中，以致于它们可以得到运行而不管它们的相关度分数有如何低。(在该种类中，可以包含前面详述的有规律地执行的图像处理操作。)

[0465] 用资源需求表示运行服务的成本的数据以所示出的(图 15 中位于标题“成本分数”下方的)数据结构来提供。该数据允许相关度与成本的对比分析得以执行。

[0466] 所示出的成本分数是多个数字构成的数组——每个数字与特定的资源需求相对应，例如内存利用率、CPU 利用率、GPU 利用率、带宽、其他成本(诸如用于与财务收费相关联的那些服务的成本)、等等。再一次，任意的 0-100 的分数在说明性方案中示出。尽管仅示出三个数字(内存利用率、CPU 利用率、和云带宽)，但是当然可以使用更多或更少的数字。

[0467] 相关度与成本的对比分析可以与系统授权一样简单或复杂。简单的分析是例如从相关度分数中减去组合的成本组分，从而为该数据结构中的第一个条目产生 -70 的结果。另一种简单的分析是例如把相关度除以合计的成本组分，从而为第一个条目产生 0.396 的结果。

[0468] 可以为队列中的全部服务执行类似的计算，以产生借以确定各服务的排序的净分数。基于上面的第一种分析方法在图 15 中提供净分数栏。

[0469] 在简单的实施例中，直到许可给直觉计算平台的资源预算被达到时才启动服务。可以向平台授予以下资源：例如，300MB 的 RAM 内存，256 千比特 / 秒的去往云的数据通道，50 毫瓦的功率消耗，以及类似限定的关于 CPU、GPU 和 / 或其他限制性资源的预算。(这些分配可以由装置操作系统设定，并且随着其他系统功能被调用或终止而变化。)当达到这些阈值中的任何一个时，直到具体情况发生变化为止，不会再有识别代理服务被启动。

[0470] 尽管简单，但是当达到所限定的资源预算中的第一个时，该方案会凌驾于全部服务之上。通常优选的是这样的方案：考虑相关限制因素中的几个或全部来设法对所调用的服务进行优化。因此，如果达到 256 千比特 / 秒的云带宽限制，那么系统仍然可以启动不需要云带宽的另外的服务。

[0471] 在更复杂的方案中，向每个候选服务赋予与该服务相关联的不同的成本组分中的每一者所对应的品质因数分数。这可以通过上面提到的用于计算净分数的减法或除法方法等来完成。通过使用减法方法，图 15 中第一个列出的服务的内存利用率所对应的成本分数 37 可以产生数值为 9 的内存品质因数(即，46-37)。该服务的关于 CPU 利用率和云带宽的品质因数分别是 -18 和 31。通过按照候选服务的不同的资源需求来对这些候选服务评分，可以使得对服务的选择能够更高效地利用系统资源。

[0472] 当新的识别代理被启动且其他识别代理终止，并且其它系统处理发生变化时，资源的自由空间(限制)会变化。这些动态限制被跟踪(图 16)，并且会影响启动(或终止)识别代理的处理。如果内存密集型 RA 完成其操作并释放出 40MB 的内存，那么平台可以启动一个或更多其他内存密集型应用程序以利用最近释放出的资源。

[0473] (本领域技术人员将会认识到，通过选择不同的服务来优化不同资源的消耗这样的任务是线性规划方面的练习，存在着许多众所周知的方法来进行线性规划。尽管这里详述的方案比实践中可能会采用的方案简单，但是有助于说明这些概念。)

[0474] 返回到图 15，所示出的数据结构还包括“条件”数据。一服务可能是高度相关的，并且资源可能足以运行该服务。然而，执行的先决条件可能尚未得到满足。例如，提供必需数据的另一登记代理服务可能尚未完成。或者用户(或代理软件)可能尚未批准该服务所需

的花费、或者尚未同意服务的点选包装协议、等等。

[0475] 一旦服务开始执行，就可以存在着允许该服务运行到完成的程式化的偏置，即使资源限制发生变化从而将合计的直觉计算平台置于其最大预算之上。不同的偏置可以与不同的服务相关联，并且对于给定服务可以与不同的资源相关联。图 15 示出对应于不同的限制（例如内存、CPU 和云带宽）的偏置。在一些情况下，偏置可以小于 100%，在这种情况下，如果该资源的可用性低于偏置数值，那么将不会启动该服务。

[0476] 例如，一个服务可以继续运行，直到合计的 ICP 带宽是其最大值的 110%，然而另一服务可以在达到 100% 阈值时立即终止。

[0477] 如果服务是特定资源的低消耗使用者，那么可以允许较高的偏置。或者如果服务具有较高的相关度分数，那么可以允许较高的偏置。（偏置可以数学地根据相关度分数推导出来，诸如偏置 =90+ 相关度分数，或者 =100，取这两个值中的较大值。）

[0478] 这种方案允许在资源命令要求时，取决于赋给不同服务和不同限制的偏置而以可编程的方式削减服务。

[0479] 在一些方案中，可以允许一些服务运行，但是要用削减后的资源来运行。例如，服务的带宽需求通常可能是 50 千比特 / 秒。然而，在特定情况下，该服务的执行可以被限制于使用 40 千比特 / 秒。再一次，这是优化方面的练习，其细节会随着应用场合而变化。

[0480] 本地软件

[0481] 在一个特定实施例中，移动装置上的本地软件可以被概念化为执行六类不同的功能（不包括安装软件并向操作系统注册该软件自身）。

[0482] 第一类功能涉及与用户通信。这允许用户提供输入，从而指定例如用户是谁、用户对什么感兴趣、什么样的识别操作与用户相关（树叶 ：是 ；车辆类型 ：否）、等等。（用户可以预订不同的识别引擎，这取决于兴趣。）用户界面功能性还提供对硬件用户界面装置的所需支持——感测触摸屏和键盘上的输入、在显示屏上输出信息、等等。

[0483] 为了有效地与用户通信，合乎期望的是软件具有对用户环境的一些 3D 理解，例如如何组织在屏幕上呈现的 2D 信息，由存在着正被表现的 3D 领域的认识告知 ；以及如何理解由摄像机拍摄的 2D 信息，在知道其表现 3D 世界的情况下。这可以包括正射位块传输图元库。这会进入第二类。

[0484] 第二类功能涉及总体定向、正射和对象场景解析。这些能力提供了可以帮助向对象识别操作提供信息的背景环境共同点（例如，天空在上方，该图像中的地平线向右倾斜 20 度，等等。）

[0485] 第三类进入实际的像素处理，并且可以被称为关键字向量处理和打包。这是已知像素处理操作（变换、模板匹配、等等）的领域。采集像素并对其进行处理。

[0486] 尽管 8x8 像素块在许多图像处理操作（例如，JPEG）中是为人们所熟悉的，但是该分组形式在本背景环境中所占优势较小（尽管它可以在某些情况下使用）。而是，五种类型的像素分组形式是占优势的。

[0487] 第一种分组形式根本不是分组，而是全局的。例如镜头盖是否在上面？焦点的通常状态是什么？这是没有太多解析（如果有的话）的类别。

[0488] 第二种分组形式是矩形区域。可以对任意数量的操作请求矩形像素块。

[0489] 第三种分组形式是非矩形邻接区。

[0490] 第四种分组形式是枚举的像素拼凑物。尽管仍然处于单个帧内，但是这是第二种分组和第三种分组的组合——常常有某种相干性(例如，某种度量或某种直观推断(heuristic)，其表示所包含的像素之间的关系，诸如与特定识别任务的相关度)。

[0491] 第五种分组形式是像素的帧间采集。这些包括像素数据(常常不是帧)的时间序列。如其余分组形式那样，特定形式将会取决于应用场合而发生很大变化。

[0492] 这种像素处理功能分类的另一方面承认资源是有限的，并且应该以逐渐增大的量分配给看上去似乎正朝着实现自己的目标(例如识别面部的目标)前进的处理，反之亦然。

[0493] 由本地软件执行的第四类功能是背景环境元数据处理。这包括聚集极大种类的信息，例如由用户输入的信息、由传感器提供的信息、或者从内存中再调用的信息。

[0494] “背景环境”的一个正式定义是“可以用于表征实体(被认为与用户和应用程序(包括用户和应用程序本身)之间的交互相关的人、地点或对象)的情况的任何信息。

[0495] 背景环境信息可以具有许多种类，包括计算背景环境(网络连接性、内存可用性、CPU 争用、等等)、用户背景环境(用户概况、位置、动作、偏好、附近朋友、社交网络和境遇、等等)、物理背景环境(例如，光照、噪声水平、交通、等等)、时间背景环境(日时、日、月、季节、等等)、上述背景环境的历史、等等。

[0496] 用于本地软件的第五类功能是云会话管理。软件需要登记不同的基于云的服务提供商作为用于执行特定任务、执行与云的例示性双向会话(建立 IP 连接、管理通信流)、对远程服务提供商进行分组因特网探测(例如，提醒不久就会请求其服务)、等等的资源。

[0497] 用于本地软件的第六类并且是最后一类功能是识别代理管理。这些包括供识别代理和服务提供商用来向手机公布它们的输入需求、在运行时必须加载(或卸载)的它们所依赖的公共库函数、它们对其他系统组件 / 处理的数据及其他依赖性、它们执行共同点处理(从而可能替代其他服务提供商)的能力、关于它们对系统资源的最大利用率的信息、关于它们各自的操作阶段(参看图 12 的讨论)和各自发布的资源需求的细节、关于它们在资源削减的情况下的性能 / 行为的数据、等等的方案。该第六类功能于是基于当前情况，在给定这些参数的情况下管理识别代理，例如取决于结果和当前系统参数来提高或降低各个服务的强度。即，识别代理管理软件充当借以根据系统资源限制来调停各代理的操作的手段。

[0498] 示例性视觉应用程序

[0499] 一种说明性应用程序用于观察处于表面上的多个硬币并计算它们的总价值。系统应用椭圆形寻找处理(例如，霍夫算法)来定位这些硬币。这些硬币可能会彼此叠盖并且一些硬币可能只是部分可见；该算法可以确定它检测的椭圆形的每个扇区的中心——每个都对应于不同的硬币。椭圆形的轴通常应该是平行的(假定斜视图，即在图像中并不是全部硬币都被描绘为圆形)——这可以充当对程序的核对。

[0500] 在定位椭圆形之后，评估各硬币的直径以识别它们各自的价值。(可以对所评估的直径进行直方图分析以确保它们在预期的直径或者在预期的直径比率处聚簇。)

[0501] 如果这几个硬币的种类有多种，那么可以仅通过直径比率来识别各硬币，而不参考颜色或标志。一角硬币的直径是 17.91mm，一分硬币的直径是 19.05mm；五分硬币的直径是 21.21mm；二角五分硬币的直径是 24.26mm。相对于一角硬币，一分硬币、五分硬币和二角五分硬币的直径比率是 1.06、1.18 和 1.35。相对于一分硬币，五分硬币和二角五分硬币的直径比率是 1.11 和 1.27。相对于五分硬币，二角五分硬币的直径比率是 1.14。

[0502] 这些比率全部是唯一的，并且分隔得足够开从而允许简便的辨别。如果两个硬币的直径比率是1.14，那么较小的一定是五分硬币，另一个一定是二角五分硬币。如果两个硬币的直径比率是1.06，那么最小的一定是一角硬币，并且另一个一定是一分硬币，等等。如果发现了其他比率，那么某些东西有差错。(应注意的是，即使硬币被描绘为椭圆形，也可以确定直径比率，因为从相同的视点观察到的椭圆形的尺寸是类似成比例的。)

[0503] 如果全部硬币都是同一类型，那么可以通过暴露的标志来识别它们。

[0504] 在一些实施例中，也可以使用颜色(例如，用于帮助区分一分硬币和一角硬币)。

[0505] 通过对识别出的二角五分硬币的价值、识别出的一角硬币的价值、识别出的五分硬币的价值、以及识别出的一分硬币的价值进行求总和，来确定处于表面上的硬币的总价值。可以通过适合的用户界面方案将该价值呈现或通告给用户。

[0506] 相关的应用程序观察一堆硬币并确定它们的原产国。每个国家的不同硬币具有唯一的一组硬币间尺寸比率。因此，如上所述的直径比率的确定可以指示该硬币集合是来自美国还是加拿大、等等。(例如，加拿大的一分硬币、五分硬币、一角硬币、二角五分硬币、和半元硬币的直径是19.05mm、21.2mm、18.03mm、23.88mm和27.13mm，因此存在着该堆硬币是否只包含五分硬币和一分硬币的某种模糊性，但是这可在包含其他硬币的情况下得到解决。)

[0507] 增强环境

[0508] 在许多图像处理应用中，视觉背景环境被很好地限定。例如，胶合板工厂中的工序控制摄像机可以在已知的光照下察看传送带上的薄木片，或者ATM摄像机可以抓取距离十八英寸远的取现金的人的安全用图像。

[0509] 手机环境更加困难——关于手机摄像机正在察看什么可能知道得很少或一点也不知道。在这种情况下，合乎期望的是，将已知的可见特征(向系统提供视觉立足点的东西)引入该环境中。

[0510] 在一个特定方案中，通过将一个或更多特征或对象放置在视场中，来帮助对一景象的机器视觉理解，其中对于所述一个或更多特征或对象，其参考信息是已知的(例如，尺寸、位置、角度、颜色)，并且系统可以借助所述一个或更多特征或对象来理解其他特征(通过相关性)。在一个特定方案中，目标图案被包含在景象中，根据所述目标图案可以辨别出例如与观察空间内的表面之间的距离和所述表面的取向。这样的目标因此充当向摄像机系统通知距离和取向信息的信标。一种这样的目标是例如在de Ipiña的“TRIP:a Low-Cost Vision-BasedLocation System for Ubiquitous Computing”(Personal and Ubiquitous Computing,Vol.6,No.3,2002年5月,pp.206-219)中详述的TRIPcode。

[0511] 如Ipiña的论文中详述的那样，(图17中所示的)目标对包括目标的半径的信息进行编码，从而允许配备有摄像机的系统确定从摄像机到目标的距离和目标的3D姿势。如果目标被放置在观察空间中的表面上(例如，在墙壁上)，那么Ipiña的方案允许配备有摄像机的系统理解与该墙壁之间的距离和该墙壁相对于摄像机的空间取向。

[0512] TRIPcode有过各种实现方案，相继地被称为SpotCode，然后是ShotCode(并且有时候是Bango)。现在它被理解为已由OP3B.V.商业化。

[0513] TRIPcode目标的美观性不适合于一些应用场合，但是非常适合于另一些应用场

合。例如，可以把地毯塑造成包含 TRIPcode 目标作为复现的设计特征，例如跨越地毯的宽度范围在规则或不规则的位置处放置的设计特征。观察包含站在这种地毯上的人的景象的摄像机可以参考该目标来确定与人之间的距离（并且还限定出包含该地板的平面）。以类似的方式，该目标可以结合到其他材料（如壁纸、家具用的布罩、衣服、等等）的设计中。

[0514] 在另外一些方案中，通过用人类视觉系统不可见、但是在例如红外光谱中可见的墨水印刷 TRIPcode 目标，使得该 TRIPcode 目标的显著度较低。移动电话中使用的许多图像传感器对红外光谱非常敏感。这样的目标因此可以从所拍摄的图像数据中辨别出来，即使该目标逃避开人类的关注。

[0515] 在又一些方案中，TRIPcode 的存在可以以仍然允许移动电话检测到它的方式隐蔽在其他景象特征中。

[0516] 一种隐蔽方法依靠摄像机传感器对图像景象的周期性采样。这种采样可以在摄像机拍摄的图像中引入当人类直接检查一条目时不明显的视觉假象（例如混叠（aliasing）、莫尔效应）。一对象可以印刷有设计成引入 TRIPcode 目标的图案，所述 TRIPcode 目标在由图像传感器的多个规则间隔开的光传感器单元成像时会通过这种假象效应而显现，但是对人类观察者而言并不明显。（相同的原理也可以有利地用于进行对基于照相复制的伪造有抵抗力的检查。诸如单词 VOID 之类的潜像被结合到原稿设计的图形元素中。该潜像对人类观察者而言并不明显。然而，当由影印机的成像系统采样时，周期性采样使得单词 VOID 在影印件中浮现出来。）各种这样的技术在 van Renesse 的“Hidden andScrambled Images-a Review”（Conference on Optical Security andCounterfeit Deterrence Techniques IV, SPIE Vol. 4677, pp. 333-348, 2002）中有详述。

[0517] 另一种隐蔽方法依靠这样的事实，彩色印刷通常用四种墨水执行：青色、品红色、黄色和黑色（CMYK）。通常，黑色材料用黑色墨水印刷。然而，黑色也可以通过叠印青色、品红色和黄色来模仿。对于人类而言，这两种技术基本上是无法区分的。然而，对于数字摄像机而言，可以很容易地辨别它们。这是因为黑色墨水通常会吸收相对大量的红外光，而青色、品红色和黄色通道则不这样。

[0518] 在将要显现黑色的区域中，印刷处理可以（例如，在白色基底上）应用重叠青色、品红色和黄色墨水的区域。该区域随后可以使用黑色墨水进一步叠印（或者预先印刷）有 TRIPcode。对于人类观察者而言，它全部显现出黑色。然而，摄像机可以根据红外线行为而辨别出差别。即，在 TRIPcode 的黑色墨水区域中的点处，存在着遮掩白色基底的黑色墨水，其吸收可能从白色基底反射的任何入射的红外线照射。在另一点处（例如在 TRIPcode 目标外部，或者在其周边内但是白色通常会显现的地方），红外线照射穿过青色、品红色和黄色墨水，并且从白色基底被反射回到传感器。

[0519] 摄像机中的红色传感器对红外线照射最敏感，因此 TRIPcode 目标在红色通道中被区别出来。摄像机可以提供红外线照射（例如，通过一个或更多 IR LED），或者环境照明可以提供足够的 IR 照射。（在未来的移动装置中，可以设置第二图像传感器，例如设置有尤其适合于红外检测的传感器。）

[0520] 刚刚描述的方案可以适合于供任何彩色印刷图像使用（不仅仅是黑色区域）。专利申请 20060008112 提供了用于这样做的细节。通过这样的方案，可以在视觉景象中可能会出现印刷的任何地方隐藏 TRIPcode 目标，从而允许通过参考这些目标来准确地测量该景

象内的某些特征和对象。

[0521] 尽管诸如 TRIPcode 之类的圆形目标因为易于计算(例如在识别处于不同的椭圆形姿态的圆形的过程中易于计算)而是合乎期望的,但是也可以使用其他形状的标记。适合于确定表面的 3D 位置的正方形标记是 Sony 的 CyberCode,并且被详细记述在例如 Rekimoto 的“CyberCode:Designing Augmented Reality Environments with VisualTags”(Proc.of Designing Augmented Reality Environments 2000,pp.1-10)中。取决于特定应用场合的要求,可以使用各种其他参考标记作为替代。在某些应用场合中有利的一种参考标记被详述在 Aller 的已公开专利申请 20100092079 中。

[0522] 在一些方案中,TRIPcode(或 CyberCode)可以进一步被处理以传递数字水印数据。这可以通过上面讨论的并且在所提到的专利申请中详述的 CMYK 方案来完成。用于生成具有隐写数字水印数据的这种机器可读数据载体的其他方案和用于这些方案的应用程序被详细记述在专利 7,152,786 和专利申请 20010037455 中。

[0523] 可以采用的具有类似效果的另一种技术是在 MIT 的媒体实验室中研发的 Bokode。Bokode 利用了摄像机镜头的散景效果(bokeh effect)——把从散焦场景点发出的光线映射成摄像机传感器上的圆盘状污迹(blur)。无需定制的摄像机能够从 10 英尺以上的距离捕获小到 2.5 微米的 Bokode 特征。可以采用二进制编码来估算与摄像机的相对距离和角度。该技术被进一步详述在 Mohan 的“Bokode:Imperceptible VisualTags for Camera Based Interaction from a Distance”(Proc.ofSIGGRAPH'09,28(3):1-8)中。

[0524] 多触摸输入、图像重新映射和其他图像处理

[0525] 如在别处提到的那样,用户可以轻拍原型小玩意以表达对系统正在处理的特征或信息的兴趣。用户的输入会提高该处理的优先级(例如通过指示系统应该将额外的资源应用于该努力)。这样的轻拍可以导致原型小玩意更快地成熟化为小玩意。

[0526] 对小玩意的轻拍也可以用于其他目的。例如,小玩意可以是以类似于 Apple iPhone 所普及的方式(即,多触摸 UI)的方式、出于用户界面目的的触摸目标。

[0527] 先前的图像多触摸界面把图像作为无差别的整体来处理。缩放等操作是在不考虑图像中描绘的特征的情况下完成的。

[0528] 根据本技术另外的方面,多触摸和其他触摸屏用户界面执行的操作部分地取决于关于所显示图像的一个或更多部分表示什么的某种认识。

[0529] 举个简单的实例,考虑跨越一张书桌的表面分散的几个条目的倾斜角度视图。一种条目可以是硬币,其在图像帧中被描绘为椭圆形。

[0530] 移动装置应用前面详述的各种对象识别步骤,包括识别与潜在不同的对象相对应的图像的边缘和区域。小玩意可以显现。通过轻拍图像中硬币的位置(或与硬币相关联的小玩意),用户可以向装置告知:图像将被重新映射使得硬币被表现为圆形——仿佛在俯视图中俯视书桌。(这有时被称为正射校正。)

[0531] 为此,合乎期望的是系统首先知道该形状是圆形。这些认识可以从几个备选源取得。例如,用户可以明确地指示该信息(例如,通过用户界面——诸如通过轻拍硬币,然后轻拍在图像的空白处呈现的圆形控制,从而指示所轻拍的对象的实际形状是圆形)。或者这种硬币可以由装置在本地识别——例如通过参考其颜色和标志(或者云处理可以提供这种识别)。或者装置可以假定从倾斜视点观察具有椭圆形形状的任何分段的图像特征实际上是

圆形。(一些对象可以包括即使倾斜也能够感测到的并且指示对象的原本形状的机器可读编码。例如，QR 条形码数据可以从矩形对象上辨别出，指示对象的真实形状是正方形。)等等。

[0532] 在图像中描绘的硬币(或相应的小玩意)上轻拍可以使图像被重新映射，而不引起其他动作。然而，在另外一些实施例中，这种指令需要来自用户的一个或更多进一步的指示。例如，用户的轻拍可以使装置呈现详述可被执行的几个备选操作的(例如，图形的或听觉的)菜单。一个备选操作可以是平面重新映射。

[0533] 响应于这种指令，系统沿着椭圆形的短轴的维度放大所拍摄图像的尺寸，使得该短轴的长度等于椭圆形的长轴的长度。(可替换地，图像可以沿着长轴缩短，具有相似的效果。)在这样做的过程中，系统已将所描绘的对象重新映射为更接近其平面图形状，使图像的剩余部分也被重新映射。

[0534] 在另一方案中，作为仅在一个方向上应用缩放因子这一方案的替代，图像可以沿着两个不同的方向被缩放。在一些实施例中，可以使用剪切或者差动缩放(例如，以便解决透视效应)。

[0535] 存储器可以存储借以根据倾斜视图确定关于对象的平面形状的推断的一组规则。例如，如果对象具有四个近似笔直的边，那么它可以被假定为是矩形——即使相对的两边在摄像机的视图中并不平行。如果对象在第三维度中没有明显的长度、基本上均匀地呈浅色(或许在浅色中带有一些频繁出现的深色标记，那么该对象可以被假定为是一张纸——如果 GPS 指示处于美国的位置，那么该张纸或许带有 8.5:11 的纵横比(或者如果 GPS 指示处于欧洲的位置，那么该张纸或许带有 1:SQRT(2) 的纵横比)。在缺乏其他认识的情况下，重新映射可以采用这样的信息来实现从所描绘的对象到近似平面图的某种东西的视图变换。

[0536] 在一些方案中，关于图像帧中的一个分段对象的认识可以用于告知或精炼关于同一帧中的另一对象的推断。考虑描绘有最大尺寸为 30 像素的圆形对象、以及最大尺寸为 150 像素的另一对象的图像帧。后一对象可以通过一些处理而被识别为是咖啡杯。参考信息的数据存储库表明咖啡杯通常的最长尺寸是 3-6″。于是前一对象可以被推断为具有大约一英寸的尺寸(而不是例如一英尺或一米左右，这可能是其他图像中描绘的圆形对象的情况)。

[0537] 不是只有尺寸分类才能以这种方式推断。例如，数据存储库可以包括把相关联的条目一起聚成一组的信息。轮胎和汽车。天空和树。键盘和鼠标。剃须膏和剃刀。食盐和胡椒摇瓶(有时是调味番茄酱和芥末分配瓶)。硬币和钥匙和手机和皮夹。等等。

[0538] 这样的关联性可以从许多来源中搜集。一个来源是来自诸如 Flickr 或 Google Images 之类的图像档案库的文本元数据(例如，识别在描述性元数据中具有剃刀的所有图像，从这些图像的元数据中收集所有其他项，并且按照出现率进行排序，例如保持前 25%)。另一个来源是通过自然语言处理，例如通过对一个或更多文本(例如词典和百科全书)进行正向链接分析，通过辨别反向语义关系而得到扩充，如专利 7,383,169 中详述的那样。

[0539] 尺寸的认识可以以相似的方式被推断出。例如，可以将参考数据的种子集合输入到数据存储库(例如，键盘在最长的维度上大约是 12-20″，手机大约是 8-12″，汽车大约是 200″，等等。)随后可以从包括已知条目以及其它条目的 Flickr 收集图像。例如，目前 Flickr 具有将近 200,000 个用项“键盘”做标签的图像。在这些图像中，超过 300 个图像也

用项“咖啡杯”做标签。对这些 300+ 个图像中相似的非键盘形状的分析揭示出 :添加的对象的最长尺寸大致是键盘的最长尺寸的三分之一。(通过类似的分析,机器学习处理可以推断出咖啡杯的形状通常是圆柱形,并且这样的信息也可以添加到由装置参考的本地或远程的知识库。)

[0540] 类似于上述讨论的推断通常不会做出最终的对象识别结果。然而,这些推断使得某些识别结果比其他识别结果更可能(或更不可能),并且因此是有用的(例如在概率分类器中)。

[0541] 有时,图像的重新映射可以不仅仅基于图像本身。例如,图像可以是例如来自视频的图像序列中的一个图像。其他图像可以来自其他透视图,从而允许该景象的 3D 模型得以生成。同样,如果装置具有立体成像器,那么可以形成 3D 模型。可以参考这样的 3D 模型来进行重新映射。

[0542] 类似地,通过参考地理位置数据,可以识别出来自同一大致位置的其他图像(例如,来自 Flickr 等),并且可以使用这些其他图像来生成 3D 模型或者向重新映射操作提供信息。(同样,如果 Photosynth 软件继续获得普及性和可获得性,那么该 Photosynth 软件可以提供使重新映射进行下去的丰富数据。)

[0543] 这种重新映射是在应用诸如 OCR 之类的识别算法之前可以应用于所拍摄图像的有帮助的步骤。考虑例如先前实例的书桌照片,该书桌照片还描绘出从书桌向上倾斜的电话,带有显示电话号码的 LCD 屏幕。由于电话的倾斜和观察角度,显示屏并不显现为矩形、而是显现为平行四边形。通过识别四边形形状,装置可以将它重新映射为矩形(例如,通过应用剪切变换)。随后可以对重新映射的图像进行 OCR——识别电话屏幕上显示的字符。

[0544] 返回到多触摸用户界面,可以通过触摸装置屏幕上显示的两个或更多特征来启动额外的操作。

[0545] 一些额外的操作实现其他的重新映射操作。考虑先前的书桌实例,其描绘有从书桌表面向上倾斜的电话 /LCD 显示屏,以及平放的名片。由于电话显示屏相对于书桌倾斜,所以这两个承载有文字的特征处于不同的平面。来自单个图像的两个 OCR 操作需要折衷。

[0546] 如果用户触摸两个分段特征(或者与这两者相对应的小玩意),那么装置评估所选择的特征的几何形状。然后,装置对电话计算垂直于 LCD 显示屏的表观平面延伸的矢量的方向,并且同样计算从名片的表面垂直延伸的矢量。随后可以对这两个矢量求平均,以产生中间矢量方向。随后可以将图像帧重新映射成使得计算出的中间矢量笔直向上延伸。在这种情况下,已经对图像进行了变换,从而得到处于这样的平面上的平面图 :该平面所呈的角度是 LCD 显示屏的平面和名片的平面之间的角度。这种重新映射的图像呈现被认为是来自位于不同平面的两个对象的 OCR 文本的最佳折衷(假定每个对象上的文本在重新映射的图像上具有相似的尺寸)。

[0547] 相似的图像变换可以基于使用多触摸界面从图像中选择的三个或更多特征。

[0548] 考虑用户处于历史古迹,周围有解释性标志。这些标记处于不同的平面中。用户的装置拍摄描绘三个标记的一帧图像,并且从这些标记的边缘和 / 或其他特征中识别出这些标记作为潜在感兴趣的离散对象。用户触摸显示器上的全部三个标记(或者相应的小玩意,一起地触摸或顺序地触摸)。通过使用类似刚刚描述的程序的程序,三个标记的平面得以确定,然后生成图像被重新映射到的折衷的观察透视图——从与平均的标志平面垂直的

方向观察该景象。

[0549] 作为从折衷的观察透视图呈现三个标记这一方案的替代，备选方法是分别对每个标记进行重新映射，使得其出现在平面图中。这可以通过将单个图像转换成三个不同的图像(每个图像利用不同的重新映射来生成)来完成。或者包含不同标记的像素可以在同一图像帧内被不同地重新映射(使附近的图像扭曲从而容纳重新整型的、可能被放大的标记图示)。

[0550] 在又一方案中，触摸三个标记(同时地或者顺序地)会启动这样的操作，其涉及从诸如 Flickr 或 Photosynth 之类的图像档案库获得指定对象的其他图像。(用户可以与装置上的用户界面进行交互，以使用户的意图变清晰，例如“用来自 Flickr 的其他像素数据进行扩充”。)这些其他图像可以通过与所拍摄图像的姿势相似性(例如，纬度/经度，加上方位)或通过其他方式(例如，其他元数据对应性，模式匹配、等等)而得到识别。可以从这些其他来源处理这些标记的更高分辨率或者聚焦更锐利的图像。这些摘选出的标记可以适当地被缩放和水平移位，然后被混合并粘贴到用户所拍摄的图像帧中——或许如上面详述的那样被处理过(例如，被重新映射到折衷的图像平面上，被分别重新映射——或许被分别重新映射到 3 个不同的图像中或者合成照片中、等等，所述合成照片被扭曲从而容纳重新整型的摘选出的标记)。

[0551] 在刚刚详述的方案中，对所拍摄的图像中可见的阴影的分析允许装置从单个帧中获得关于景象的某些 3D 认识(例如，对象的深度和姿势)。该认识可以帮助向上面详述的任何操作提供信息。

[0552] 正如对图像(或摘选)进行重新映射可以帮助进行 OCR 那样，它也可以帮助决定应该启动什么样的其他识别代理。

[0553] 在图像中轻拍两个特征(或小玩意)可以启动一处理以确定所描绘的对象之间的空间关系。在 NASCAR 比赛的摄像机视图中，小玩意可以叠盖在不同的赛车上并跟踪赛车的移动。通过轻拍用于与赛车相连的小玩意(或者轻拍所描绘的赛车本身)，装置可以获得每个赛车的位置数据。这可以用来自取景器的透视图的相对值来确定，例如通过从赛车在图像帧中的缩放比例和位置推断赛车的位置(在知道摄像机光学系统的细节和赛车的真实尺寸的情况下)。或者装置可以链接到一个或更多跟踪赛车的实时地理位置的万维网资源，例如根据所述实时地理位置，用户装置可以报告赛车之间的间隙是八英寸并且正在靠拢。

[0554] (如在前面的实例中那样，该特定操作可以在用户轻拍屏幕时从由几个可能的操作构成的菜单中选择。)

[0555] 作为简单地轻拍小玩意这一方案的替代，进一步的改进涉及在屏幕上拖动一个或更多小玩意。它们可以被拖动到彼此上，或者拖动到屏幕的一区域上，由此用户可以告知期望的动作或询问。

[0556] 在具有几个面部的图像中，用户可以将两个对应的小玩意拖动到第三个小玩意上。这可以指示分群操作，例如所指示的人具有某种社交关系。(关于该关系的更多细节可以由用户使用文本输入来输入，或者通过语音识别由口述的文本输入。)在网络图的意义上，在表示两个个体的数据对象之间建立链路。该关系可以影响其他装置处理操作如何处理所指示的个体。

[0557] 可替换地，全部三个小玩意可以拖动到图像帧中的新位置。该新位置可以表示与

分群相关联的操作或属性(或者是推断出的(例如,背景环境),或者由用户输入明示)。

[0558] 特征代理小玩意的另一种交互式应用是对图像进行编辑。考虑图像含有三个面部:两个朋友和一个陌生人。用户可能想要将该图像发布到在线储存库(Facebook),但是可能想先去除陌生人。为此可以操纵小玩意。

[0559] Adobe Photoshop CS4 引入了被称为智能缩放的特征,它在之前可从诸如rsizr<dot>com 之类的在线站点获知。(例如,用鼠标绘出的边界框)标示出将要保存的图像区域,然后缩小或删除(例如,具有多余特征的)其他区域。图像处理算法使保存的区域保持不变,并且将它们与先前具有多余特征的编辑过的区域混合。

[0560] 在本系统中,在处理一帧图像以生成与辨别出的特征相对应的小玩意之后,用户可以执行一系列手势,指示将要删除一个特征(例如,陌生人)、并且将要保存其他两个特征(例如,两个朋友)。例如,用户可以触摸不想要的小玩意,并将手指扫到显示屏的底部边缘以指示相应的视觉特征应该从图像中去除。(小玩意可以跟随手指,或者不跟随手指。)然后用户可以双击每个朋友小玩意以指示将要保存它们。另一手势会唤出一菜单,用户从该菜单中指示已经输入了全部编辑手势。然后处理器根据用户的指示来编辑图像。如果编辑的图像被证明不令人满意,那么“取消”手势(例如,用手指在屏幕上划出逆时针半圆轨迹)可以撤销该编辑,并且用户可以尝试另一编辑。(系统可以被设置在这样的模式中:通过屏幕上的手势(例如用手指划出字母“e”的轨迹)、或者通过从菜单选择、或者通过其它方式接收编辑小玩意手势。)

[0561] 由多次小玩意轻拍构成的序列的顺序可以把关于用户意图的信息传递给系统,并引出相应的处理。

[0562] 考虑一游客在新城镇中观看介绍各种兴趣点、并且带有每个旅游胜地(例如,埃菲尔铁塔、凯旋门、卢浮宫、等等)的照片的标志。用户的装置可以识别这些照片中的一些或全部,并呈现与每个描绘的旅游胜地相对应的小玩意。按特定顺序触摸这些小玩意可以指示装置获得按轻拍的顺序去往所轻拍的旅游胜地的行走方向。或者可以使得装置为每个旅游胜地取来 Wikipedia 条目并按指示的顺序呈现这些 Wikipedia 条目。

[0563] 由于特征代理小玩意与特定对象或图像特征相关联,所以这些特征代理小玩意在被轻拍或者被包含在手势中时可以具有取决于它们所对应的对象 / 特征的响应。即,对手势的响应可以随与所涉及的小玩意相关联的元数据而变。

[0564] 例如,在对应于一个人的小玩意上轻拍可以意味着与在对应于雕像或餐馆的小玩意上轻拍的情况不同的某事(或者会唤出不同的可用操作的菜单)。(例如,在前者上轻拍可以引出例如来自 Facebook 的该人的名字和社交概况的显示或通告;在第二者上轻拍可以唤出关于该雕像或其雕刻家的 Wikipedia 信息;在后者上轻拍可以得到该餐馆的菜单和关于任何当前促销的信息。)同样,涉及在两个或更多小玩意上轻拍的手势具有的含义也可以取决于所轻拍的小玩意表示什么、并且任选地取决于小玩意被轻拍的顺序。

[0565] 随着时间,跨越不同的小玩意而总体保持一致的手势词典可以得到标准化。例如,轻拍一次可以唤出与小玩意的类型相对应的特定类型的介绍信息(例如,如果轻拍与人相关联的小玩意,那么唤出名字和概况;如果轻拍与建筑物相关联的小玩意,那么唤出地址和办公室目录;如果轻拍用于历史古迹的小玩意,那么唤出 Wikipedia 页面;如果轻拍用于零售产品的小玩意,那么唤出产品信息,等等)。轻拍两次可以唤出例如四个最频繁调用的操

作的精华菜单，同样被裁制成适合于相应的对象 / 特征。对小玩意的触摸和手指在该位置的摆动可以启动另一响应——诸如具有滚动条的完整选项菜单的显示。再一次抖动可以使该菜单收起。

[0566] 关于架构的注释

[0567] 本说明书详述了许多特征。尽管各实现方案可以利用这些特征的子集来实现，但是这些实现方案距优选方案还有一点距离。实现更丰富而不是较稀少的一组特征的原因在下面的讨论中阐述。

[0568] 示例性的软件框架使用各种组件来支持在智能手机上运行的视觉应用程序：

[0569] 1. 屏幕是实时更改的摄像机图像，上面叠盖有动态图标（小玩意），所述动态图标可以附着在图像的一部分上并且同时充当用于立即发生的（可能的）多个动作的价值显示器和控制点。屏幕也是有价值的可货币化的广告空间（以类似于 Google 的搜索页面的方式）——正好处在用户关注的焦点上。

[0570] 2. 用于装置的许多应用程序处理摄像机图像的实时序列，而不仅仅是处理“快照”。在许多情况下，需要复杂的图像判断，尽管响应性保持一定优先级。

[0571] 3. 实际应用程序通常与所显示的小玩意和显示器所显示的当前可见“景象”相关联，从而允许用户交互成为这些应用程序的所有级别中的普通一部分。

[0572] 4. 基本的一组图像特征提取功能可以在背景中运行，从而允许可见景象的特征一直都可被各应用程序利用。

[0573] 5. 合乎期望的是，各个单独的应用程序不被允许“贪心攫取”系统资源，因为许多应用程序的有用性会随着可见景象的变化而盛衰，因此不止一个应用程序常常会立即处于活跃状态。（这通常需要多重任务处理，具有适合的分派能力以保持足够活跃从而有用的应用程序。）

[0574] 6. 应用程序可以设计成多层，使相对低负荷的功能监视景象数据或用户需求，使更资源密集的功能在适当的时候被调用。分派方案可以支持该代码结构。

[0575] 7. 许多应用程序可以包括基于云的部分来执行超出装置本身的实际能力的操作。再一次，分派方案可以支持该能力。

[0576] 8. 应用程序常常需要用于发布和访问相互有用的数据的方法（例如，黑板）。

[0577] 下面以宽松无序的方式描述一些相互关系，所述相互关系可以使上述的各方面部分成为一个整体（而不仅仅是合乎期望的个体）。

[0578] 1. 参考实时景象的应用程序通常会依靠从全部（或者至少许多）帧中高效地提取基本图像特征——因此使实时特征可被获得是一个重要考虑因素（即使对于某些应用程序，可能不需要它）。

[0579] 2. 为了允许高效的应用程序开发和测试以及为了在能力各不相同的多个装置上支持这些应用程序，将任何应用程序的重要部分任选地放置“在云中”的能力会变得几乎是强制性的。许多益处产生于这种能力。

[0580] 3. 许多应用程序会受益于超出未受协助的软件的当前能力的识别能力。这些应用程序会要求与用户的交互是有效的。此外，移动装置通常会请求用户交互，并且只有当 GUI 支持该需求时，一致的友好的交互才是可能的。

[0581] 4. 在具有有限的不可改变的资源的装置上支持复杂的应用程序需要来自软件架

构的充分支持。把 PC 风格的应用程序在不经过细致的重新设计的情况下硬塞进这些装置中通常是不令人满意的。分层式软件的多重任务处理会是在这种装置受限制的环境中提供邀请用户体验的重要组件。

[0582] 5. 以高效的方式向多个应用程序提供图像信息，最好是通过只产生一次信息、并允许该信息按照使信息访问和缓存无效最少化的方式由每个需要它的应用程序使用来完成。“黑板”数据结构是实现该高效率的一种方式。

[0583] 因此，尽管所详述的技术的一些方面各自单独是有用的，但是在组合中它们的最大效用才会实现。

[0584] 关于黑板的更多说明

[0585] 可以在黑板中采用垃圾收集技术以去除不再相关的数据。去除的数据可以转移到长期存储库（如磁盘文件）以便充当其他分析中的资源。（去除的数据也可以转移或复制到云，如其它地方所述。）

[0586] 在一个特定方案中，当多个备选标准中的第一个得到满足时，例如当新的发现会话开始时、或者用户位置的变化大于阈值（例如，100 英尺或 1000 英尺）时、或者自从基于图像和音频的关键字向量数据被生成以来已逝去失效时间（例如，3 或 30 或 300 秒）时，从黑板中去除所述关键字向量数据。在前两种情况下，可以保留旧数据，例如在新的发现会话开始后将旧数据再保留 N 个时间增量（例如再保留 20 秒），或者在用户位置的变化大于阈值后将旧数据再保留 M 个增量（例如，再保留 30 秒）。

[0587] 考虑到非图像 / 音频关键字向量数据（例如，加速计、陀螺仪、GPS、温度）对存储的要求很有限，所以非图像 / 音频关键字向量数据通常在黑板上保存的时间要长于图像 / 音频关键字向量数据。例如，这样的数据可以持续保存在黑板上，直到手机下一次处于睡眠（低电池消耗）运行状态的时间超过四小时为止、或者直到若干个这样的相继睡眠状态已经发生为止。

[0588] 如果新利用了任何老化的黑板数据（例如，被识别代理用作输入，或者被新发现为与其他数据相关），那么延长该数据在黑板上的准许驻留时间。在一个特定方案中，所延长的时间等于从数据最初被生成到该数据新近被利用为止的时间（例如，将该数据新近被利用的时间作为新的生成时间对待）。与共同的对象相关的关键字向量数据可以以新的关键字向量的形式聚集在一起，从而类似地延长其在黑板上的准许使用寿命。

[0589] 如果被去除的数据在与用户当前位置的地理接近度的阈值度量内被搜集到，那么该数据也可以在被去除（例如，从长期存储库中被去除）之后恢复到黑板上。例如，如果当用户处于购物中心中时黑板被填充有与图像相关的关键字向量数据、随后用户开车返回家中（从而对黑板进行冲刷），那么当用户下一次回到该购物中心时，与该位置相对应的最近被冲刷掉的关键字向量数据可以恢复到黑板上。（恢复的数据量取决于黑板尺寸和可用性。）

[0590] 在一些方面中，可以将黑板实现为一种聚焦于传感器融合的关于对象的自动维基系统（Wiki），或者另一种数据结构可以充当一种聚焦于传感器融合的关于对象的自动维基系统。每过几秒（或一秒的数分之一），多页数据就被产生，并且数据元素之间的链接就被断开（或者新的链接就被建立）。识别代理可以填充页面并设立链接。页面被频繁地编辑，通常将状态机用作编辑器。每个维基系统作者可以看到所有其他的页面并且可以对其做贡献。

[0591] 系统也可以例如与黑板相关联地调用信任程序。每次识别代理试图将新数据发

布到黑板上时，可以在信任系统数据库中对该代理进行调查以确定其可靠性。数据库也可以指示该代理是否是商用代理。在确定将要对该代理所发布的数据赋予的可靠性分数（或者是否应该完全准许与黑板的合作）的过程中可以考虑用户对该代理的评级。基于信任调查结果和存储的策略数据，可以准许或拒绝向代理授予某些特权（例如贡献链接、断开链接（自己的链接或第三方的链接）、删除数据（自己的数据或第三方的数据）、等等）。

[0592] 在一个特定方案中，装置可以与独立的信任权威（例如 Verisign 或 TRUSTe）协商以调查识别代理的可信赖度。可以采用已知的密码技术（例如数字签名技术）来认证提供代理业务的第三方是该第三方声称的那一方、并且认证任何代理软件未被篡改。只有当这样的认证成功时、和 / 或只有当独立的信任权威将该提供商评级为高于阈值（例如，“B”或者 100 中的 93，这可由用户设定）的级别时，才向该识别代理授予与装置的黑板结构交互（例如，通过读取和 / 或写入信息）的特权。

[0593] 装置可以类似地（例如，通过 TRUSTe）调查服务提供商的隐私措施、并且只有当超过某些阈值或满足参数时才允许交互。

[0594] 关于处理、使用模型、罗盘和会话的更多说明

[0595] 如上所述，一些实现方案在自由运行的状态下拍摄图像。如果有限的电池电力是一个限制因素（如目前的通常情况那样），那么系统可以在某些实施例中以高选择性模式处理该连续的图像流——很少将装置的计算能力的显著部分（例如，10% 或 50%）应用于该数据的分析。而是，系统在低耗电状态下工作，例如执行没有显著功率成本的操作，和 / 或每秒钟或每分钟只检查（例如每秒钟拍摄的 15、24 或 30 个帧中的）几个帧。只有在（A）初始的低级处理指示出图像中描绘的对象可以得到准确识别的高概率，并且（B）背景环境指示出这种对象的识别会与用户相关的高概率的情况下，系统才会提速为使功率消耗增大的第二模式。在该第二模式中，功率消耗可以大于第一模式中的功率消耗的两倍、或 10 倍、100 倍、1000 倍或更多倍。（上面提到的概率可以基于计算出的取决于特定实现方案的数值分数。只有当成功的对象识别以及与用户的相关度所对应的这些分数超过各自的阈值（或者按照公式组合起来从而超过单个阈值）时，系统才会切换为第二模式。）当然，如果用户明确地或暗示地告知兴趣或鼓励，或者如果背景环境有指示，那么系统也可以从第一模式切换为第二模式。

[0596] 用于某些增强现实（AR）应用程序的新兴的使用模型（例如，用户被预期在拿出智能电话并全神贯注于它不断变化的显示（例如，以便导航到期望的咖啡店或地铁车站）的同时沿着城市的街道行走）是考虑不周的。许多备选者似乎是优选的。

[0597] 一个备选者是通过耳机或扬声器以可听的方式提供引导。胜于提供语音引导，可以利用更精巧的听觉线索，从而允许用户更加注意其他听觉输入（如汽车喇叭声或同伴的谈话）。一种听觉线索可以是重复率或频率发生变化从而告知用户是否在正确的方向上行走并靠近期望的目的地的偶尔的音调或滴答声。如果用户在十字路口尝试做出错误的转弯、或者远离目的地移动而不是朝向目的地移动，那么样式可以以独特的方式变化。一个特定方案采用类似盖革式计数器的声音效果，使滴答声的稀疏样式随着用户朝向期望的目的地前进而变得更频繁，并且在用户从正确的方向转开的情况下降低频繁度。（在一个特定实施例中，听觉反馈的音量随着用户的运动而变化。如果用户暂停（例如在交通信号灯处），那么可以增大音量，从而允许用户面向不同的方向并通过音频反馈识别前进的方向。一旦用

户恢复行走，那么音频音量可以变小，直到用户再一次暂停。音量或其他用户反馈强度水平因此可以在用户按照导航方向前进时减小，并且在用户暂停或从预期路径转向时增大。）

[0598] 运动可以以各种方式检测，诸如通过加速计或陀螺仪输出、通过变化的 GPS 坐标、通过摄像机感测到的变化的景象、等等。

[0599] 作为听觉反馈的替代，上述方案可以采用振动反馈。

[0600] 移动装置中的磁力计可以在这些实现方案中用来感测方向。然而，移动装置可以以任意方式相对于用户和用户向前行进的方向而被定向。如果移动装置被夹在面向北的用户的腰带中，那么磁力计可以指示出装置指向北方、或南方、或任何其他方向——取决于装置如何被定位在腰带上。

[0601] 为了解决该问题，该装置可以辨别应用于磁力计输出的校正因子，以便正确地指示出用户正面向的方向。例如，该装置可以通过参考偶尔的 GPS 测量值来感测用户沿着其移动的方向矢量。如果在十秒钟内用户的 GPS 坐标的纬度增大、但是经度保持恒定，那么用户已经向北移动（推测起来可能是面向北方方向向北移动）。该装置可以在该时期期间注意磁力计输出。如果装置被定位成使得其磁力计一直指示“东”，而用户明显面向北方，那么可以辨别出 90 度的校正因子。此后，该装置知道从磁力计指示的方向中减去九十度以确定用户正面向的方向——直到这种分析指示出应该应用不同的校正。（这种技术可广泛地应用，并且不限于这里详述的特定方案。）

[0602] 当然，这种方法不仅对行走适用，而且对自行车和其他交通方式也适用。

[0603] 尽管详述的方案假定图像是在它被拍摄到时进行分析的、并且假定拍摄是由用户装置执行的，但是这两者都不是必需的。相同的处理可以对先前拍摄的和 / 或在其它地方拍摄的图像（或音频）执行。例如，用户的装置可以处理一小时或一星期以前由例如城市停车场中的公共摄像机拍摄的图像。其他图像源包括 Flickr 和其他这样的公共图像储存库、YouTube 和其他视频站点、通过在公共万维网上爬行而收集的图像、等等。

[0604] （优选的是，将处理软件设计成使得其能够可交替地处理实时图像数据和已存图像数据（例如实时的静止图像或图像流，以及先前记录的数据文件）。这允许看上去不同的用户应用程序采用相同的内核。对于软件设计者而言，这也是有用的，因为其允许实时图像应用程序利用已知的图像或序列被重复测试。）

[0605] 许多人更喜欢以转录的文本形式回顾语音邮件——略读相关内容，而不是收听散漫的谈话者的每段发言。以类似的方式，基于视觉图像序列的结果可以比拍摄该序列所花费的时间更快速地由许多用户回顾和理解。

[0606] 考虑下一代移动装置包含在头饰上安装的摄像机，由沿着城市街区散步的用户佩戴。在跨越该街区期间，摄像机系统可能会收集到 20、60 或更多秒钟的视频。作为（在行走的同时）分散注意力地察看给出基于图像的结果的叠盖的 AR 呈现这一方案的替代，用户可以将注意力集中于避开步行者和障碍物的即时任务。同时，系统可以分析所拍摄的图像并存储结果信息以供之后回顾。（或者，作为在行走的同时拍摄图像这一方案的替代，用户可以暂停、摆动配备有摄像机的智能电话以拍摄全景图像、然后将智能电话放回到口袋或钱包中。）

[0607] （结果信息可以具有任何形式，例如图像中的对象的识别结果、与这些对象相关地获得的音频 / 视频 / 文本信息、关于响应于视觉刺激而采取的其它动作的数据、等等。）

[0608] 在方便的时候，用户可以看一下智能电话屏幕（或者激活眼镜上的平视显示器）来回顾基于所拍摄的帧序列产生的结果。这种回顾可以仅涉及响应信息的呈现，和 / 或可以包括各个响应所基于的拍摄影像。（在响应基于对象的情况下，对象可能出现在该序列的几个帧中。然而，该响应只需要针对这些帧中的一个帧呈现。）对结果的回顾可以由装置以标准化的呈现方式指引，或者可以由用户指引。在后一种情况下，用户可以采用用户界面控制来航行通过结果数据（可以与图像数据相关联地呈现，或者不这样）。一种用户界面是由 Apple iPhone 家族普及的为人们所熟悉的触摸界面。例如，用户可以扫过一景象序列（例如以 1 或 5 秒钟、或者 1 或 5 分钟为间隔拍摄的多个帧），每个景象都叠盖有可被轻拍从而呈现附加信息的小玩意。另一种导航控制是图形的或物理的往复式控制（从诸如 Adobe Premier 之类的视频编辑产品而为人们熟悉），允许用户对图像和 / 或响应的序列进行向前加速、暂停、或倒退。一些或全部结果信息可以以听觉形式呈现，而不是以视觉形式呈现。用户界面可以是语音响应式的，而不是例如触摸响应式的。

[0609] 尽管已经以视频方式收集到视觉信息，但是用户可能会发现以静态景象方式回顾该信息能够提供最多信息。这些静态帧通常由用户选择，但是可以由装置选择或预先过滤，例如略去低品质的帧（例如，模糊的帧、或者被前景中的障碍物遮蔽、或者不具有很多信息内容的帧）。

[0610] 装置获得的响应的导航不需要横跨整个序列（例如，显示每个图像帧或每个响应）。一些模式可以向前跳过一些信息，例如仅呈现与每两帧中的第二帧、或者每十帧中的第十帧、或者每个帧数量或帧时间的某个其它间隔的帧中的最后一帧相对应的响应（和 / 或图像）。或者回顾可以基于突出性或内容而向前跳跃。例如，不具有任何识别出的特征或相应的响应的部分序列可以被全部跳过。具有一个或几个识别出的特征（或其他响应数据）的图像可以呈现较短的时间。具有许多识别出的特征（或其他响应数据）的图像可以呈现较长的时间。用户界面可以呈现用户借以设定回顾的整体速度（例如，使得花费 30 秒钟拍摄的序列可以在十秒钟、或 20、或 30 或 60 秒钟等内得到回顾）的控制。

[0611] 应认识到的是，刚刚描述的回顾时间到拍摄时间的映射可以是非线性的，例如由于影像的突出性随时间变化（例如，一些影像摘选中的感兴趣对象较丰富；而另一些影像摘选则不这样）、等等。例如，如果在 15 秒中回顾的序列所花的拍摄时间是 60 秒，那么回顾过程的三分之一可能不对应于拍摄过程的三分之一，等等。因此，对象可以在回顾数据中的与该对象在拍摄数据中的时间位置不成比例的时间位置处出现。

[0612] 用户界面还可以提供用户借以暂停任何回顾从而允许进一步研究或交互、或者请求装置对特定的描绘特征进行进一步分析并报告的控制。响应信息可以按照与图像被拍摄的顺序相对应的顺序、或者相反的顺序（最新近者优先）回顾，或者可以基于估算出的与用户的相关度来排序、或以非时间先后顺序的某种其他方式来排序。

[0613] 这种交互和分析可以被认为是采用基于会话的构造。用户可以在图像序列的中间开始回顾，并向前或向后穿越（连续地，或者跳来跳去）。这种会话方案的一个优点是后来获取的图像可以帮助告知对先前获取的图像的理解。只举一个实例来说，人的脸部可能会在帧 10 中被揭示出来（并且使用面部识别技术得以识别），而帧 5 可能只显示出这个人的头部的背面。而通过分析作为集合的图像，这个人可以在帧 5 中得到正确地标注，并且对帧 5 的景象的其他理解可以基于这样的认识。相反，如果景象分析仅仅基于当前帧和前面的帧，那

么这个人在帧 5 中将是匿名的。

[0614] 可以通过这里详述的实施例来使用会话构造。一些会话具有自然的起点和 / 或终点。例如，所拍摄视频中的突然的场景转换可以用于开始或结束会话，如同用户从口袋中取出摄像机扫描场景、并在之后将其放回到口袋中的情况那样。（从 MPEG 借鉴的技术可以用于该目的，例如检测需要开始新的图像群（GOP）的场景变化，所述新的图像群以“I”帧为起点。）失去新颖性的场景可以用于结束会话，正如呈现出新兴趣的场景可以开始会话那样。（例如，如果摄像机已经从床头柜凝视空间了一整夜、并在随后被拿起来从而将运动新近引入到影像中，那么这可以触发会话的开始。相反，如果将摄像机以固定取向留在静态环境中，那么这种缺乏新视觉刺激的状态会很快引起会话结束。）

[0615] 可以可替换地或者附加地采用与基于图像的会话类似的基于音频的会话。

[0616] 手机中的其他传感器也可以用于触发会话的开始或结束，例如发出用户已经拿起手机或改变手机方向的通知信号的加速计或陀螺仪。

[0617] 用户动作也可以明确地发出会话的开始或结束的通知信号。例如，用户可以口头地指示装置“看 TONY”。这样的指示是充当新会话的逻辑起点的事件。（指示也可以通过除语音以外的方式来发出，例如通过与用户界面交互、通过摇动手机以发出应该将其计算资源集中 / 增加到当时存在于环境中的刺激上的通知信号、等等。）

[0618] 一些会话可以通过诸如“发现”或“开始”之类的词语来明确地调用。这些会话可以响应于来自软件定时器的信号（例如，在 10、30、120、600 秒之后，这取决于所存储的配置数据）而终止，除非在这之前被诸如“停止”或“退出”之类的指示停止。发出定时器正在接近会话的终点的警告的 UI 可以被发布给用户，并且可以提供对按钮的选择或其他控制方案，从而允许会话延长例如 10、30、120 或 600 秒或者从而允许无限期地延长会话（或者从而允许用户输入另一个值）。

[0619] 为了避免不必要的数据捕获和含糊的指示，用户可以发出诸如“只需要看”或“只需要听”之类的指示。在前一种情况下，不会对音频数据进行采样（或者，如果被采样，则并不存储它）。对于后一种情况而言，则相反。

[0620] 类似地，用户可以说出“听音乐”或“听讲话”。在每个情况下，所捕获的数据可以按照类别被分段和识别，并且可以将分析聚焦于指定的类型。（其他类型可以被丢弃。）

[0621] 同样，用户可以说出“听电视”。除了该指示可以调用的其他处理之外，该指示还提示处理器在电视音频中寻找由 Nielsen 公司编码的某种数字水印数据。（这样的水印被编码在特定频谱范围中，例如 2KHz-5KHz。在了解这样的信息的情况下，装置可以相应地对其采样、滤波和分析进行裁制。）

[0622] 有时，会捕获到与意图的发现活动无关的数据。例如，如果会话的长度由定时器设定或者由视觉无变化状态持续时间（例如，十秒）确定，那么该会话可能会捕获到对于意图的发现操作而言没有价值的信息（特别是在接近终点时）。系统可以采用识别哪些数据与意图的发现操作相关、并丢弃剩余数据的处理。（或者，类似地，系统可以识别哪些数据与意图的发现操作不相关并将其丢弃。）

[0623] 考虑用户在电子产品商店中拍摄潜在感兴趣的产品的影像——特别是产品的条形码。该会话也可能捕获到例如商店顾客的音频和其他影像。根据视频数据并且特别是根据去往用户驻足的相继出现的条形码的移动，系统可以推断出用户对产品信息感兴趣。在

这种情况下，系统可以丢弃音频数据和不含有条形码的视频。（同样，系统可以丢弃与条形码不相关的关键字向量数据。）在一些实现方案中，系统在采取这样的动作之前向用户核对，例如详述该系统对用户感兴趣的东西的假设、并请求确认。只有与影像的条型码区域相对应的关键字向量数据可以被保留。

[0624] 尽管会话通常表示一种时间构造（例如包含一系列逻辑上相关的事件或过程的时间间隔，但是也可以采用其他会话构造。例如，可以参考图像帧内的特定空间区域或者图像序列内的特定空间区域（在这种情况下，区域可以展现出运动）来定义逻辑会话。（MPEG-4 对象可以各自按照空间会话来看待。对于其他面向对象的数据表示形式而言，也同样如此。）

[0625] 应认识到的是，多个会话可以同时进行，从而完全或部分重叠、相互独立地开始和结束。或者多个会话可以共享共同的起点（或终点），同时这多个会话相互独立地结束（或开始）。例如，摇动手机（或在手机上轻拍）可以使手机更加注意进入的传感器数据。手机可以通过增加应用于麦克风和摄像机数据的处理资源来做出响应。然而，手机可能会快速地辨别出：不存在值得注意的麦克风数据，而视觉场景正在动态地变化。手机因此可以在几秒之后终止音频处理会话，从而减少应用于音频分析的资源，同时继续视频处理会话更长的时间，直到例如该活动消退、用户动作表明要停止、等等。

[0626] 如前所述，来自发现会话的数据被共同地存储并且可以在之后被召回。然而，在一些情况下，用户可能会希望丢弃会话的结果。UI 控制可以允许这样的选项。

[0627] 诸如“看 TONY”之类的口头指示可以极大地帮助装置执行其操作。在一些方案中，手机不需要一直处于提高的警惕状态——试图在永不枯竭的传感器数据洪流中辨别出有用的东西。而是，手机可以通常处于较低的活动状态（例如，执行由所存储的调速数据建立的背景级别的处理），并且仅根据指示来调配额外的处理资源。

[0628] 这样的指示也充当可以快速剪除其他处理的重要线索。通过参考所存储的数据（例如，本地或远程数据库中的存储数据），手机可以快速地识别出“Tony”是诸如人类、个人、男性、FaceBook 朋友和 / 或面部之类的一个或更多逻辑类别的成员。手机可以启动处理或者对处理进行裁制以辨别并分析与这类实体相关联的特征。用另一种方式表述，手机可以识别不需要关注的某些任务或对象类别。（“看 TONY”可以被认为是不要寻找钞票、不要解码条型码、不要执行歌曲识别、不要聚焦于汽车、等等的指示。这些处理在正在进行的情况下可以被终止，或者干脆在会话期间不被启动。）所述指示因此大大地缩小了装置必须应对的视觉搜索空间。

[0629] 手机在解释用户的指示的过程中查阅的存储数据可以具有各种形式。一种是简单的词汇表，其对每个单词或短语指出一个或更多相关联的描述符（例如，“个人”、“地方”或“东西”；或者一个或更多其他的类描述符）。另一种是用户的电话簿，其列出名字，并且任选地还提供联系人的图像。另一种是用户的社交网络数据，例如标识出朋友及感兴趣的对象。一些这样的资源可以处于云中——在多个用户群之间共享。在诸如电话簿之类的一些情况下，所存储的数据可以包括图像信息或线索，以便辅助手机进行图像处理 / 识别任务。

[0630] 在这样的实施例中有用的语音识别技术是本领域技术人员所熟悉的。可以通过限制识别算法必须对其进行匹配的候选单词的范围来提高识别的准确度。通过将词汇表限制到一千个（或一百个或更少的）单词，可以用有限的处理和有限的时间获得极高的识别准确度。（这样的精简词汇表可以包括：朋友的名字，诸如开始、停止、看、听、是、否、执行、退出、

结束、发现之类的通用指示单词，通用颜色、数位和其他数字，当前地域中的流行地理术语，等等。）如果速度（或本地数据存储）不是最需关注的因素，那么可以采用 Google 公司在其 GOOG411 产品中使用的语音识别技术。关于语音识别技术的相关信息被详述在本受让人的申请 20080086311 中。

[0631] 来自用户的指示不需要是具有既有定义的熟悉单词。这些指示可以是发声、喷鼻息、鼻音、咕哝声或者用户在某些背景环境中发出的其他声音。“UH-UNH”可以被认为是否定的——向手机指示其当前聚焦对象或结果不令人满意。“UM-HMM”可以被认为是肯定的——确认手机的处理与用户的意图一致。手机可以被训练成适当地响应这样的发声，如同其他未识别出的单词一样。

[0632] 指示不需要是听觉的。指示可以是其他形式，例如通过手势。再一次，手机可以通过训练经历而将各含义归因于相应的手势。

[0633] 在一些实施例中，视觉投影可以将手机指引到感兴趣的对象。例如，用户可以使用具有已知的光谱颜色或独特的时间或频谱调制的激光指示器来指向感兴趣的对象。显微投影仪可以类似地被用来把独特目标（例如图 17 的目标，或 3x3 的点阵列）投影到感兴趣的对象上——使用可见光或红外线。（如果使用可见光，则所述目标可以被不频繁地投影（例如，对于每秒而言，投影的时间占一秒的三十分之一），使得检测软件可以与该定时同步。如果使用红外线，则可以用红色激光指示器的光点来投影，以便向用户示出红外线图案被布置在哪里。在一些情况下，所述目标可以针对不同的用户而被个性化（例如被串行化），从而允许许多投影目标能同时存在，例如在公共场所中。）这样的投影目标不仅指示出感兴趣的对象，而且允许该对象的取向和与该对象间的距离（其姿势）得到确定，从而建立在其他分析中有用的“地面实况（ground truth）”。一旦在影像内发现投影的特征，系统就可以对图像进行分割 / 分析以便识别在其上发现所述目标的对象、或者采取其他响应性动作。

[0634] 在一些方案中，手机总是在寻找这样的投影指示。在另一些方案中，这样的动作由用户的口头指示“寻找激光”或“寻找目标”来触发。这是采用多种指示的组合的实例：讲话和视觉投影的组合。不同类型的指示的其他组合也是可能的。

[0635] 如果系统没有识别出特定的指示或者在完成相关联的任务的尝试中失败，那么系统可以通过向用户反馈（例如通过咂舌声、音频问题（例如，“是谁？”或“是什么？”）、通过视觉消息、等等）来表明该情况。

[0636] 例如，手机可以理解：“看 TONY”是处理影像以辨别出用户的朋友的指示（关于该朋友的参考影像可从存储库中获得）。然而，由于该手机摄像机的观察角度，手机可能无法在视场内识别出 Tony（例如，Tony 的背部可能对着摄像机），并且可能会向用户指示识别失败的情况。用户可以通过尝试诸如“帽子”、“绿色衬衣”、“附近”、“右边的人”等（借以识别出意图的对象或动作的其他线索）之类的其他指示来做出响应。

[0637] 购物中心中的用户可以拍摄显示出货架上的三个物品的影像。通过说出“中间的一个”，用户可以将手机的处理资源聚焦于了解中间的物体，从而把位于左侧和右侧的物体（及其他地方）排除在外。其他描述符可以同样地得到使用（例如“红色的那一个”或者“正方形的那一个”、等等）。

[0638] 根据这些实例，应认识到的是，音频线索（和 / 或其他线索）可以被用作对 ICP 装置的处理努力划出界线的一种手段。对象识别因此可通过语音识别（和 / 或其他线索）而得到

补充 / 辅助。

[0639] (相反,语音识别也可以通过对象识别而得到补充 / 辅助。例如,如果装置识别出用户的朋友 Helen 在摄像机的视场中、并且如果口头说出的单词模糊不清——可能是“英雄(hero)”或“Helen”或“喂(hello)”,那么在影像中识别出海伦这个人可以使所述模糊单词的解释倾向到“Helen”。类似地,如果视觉背景环境表示的是有鸭子的池塘,那么模糊不清的单词可以被解析为“家禽(fowl)”,而如果视觉背景环境表示的是棒球场,那么同样的单词可以被解析为“犯规(foul)”。)诸如来自 GPS 的数据之类的位置数据可以类似地用于解析讲话中的模糊性。(如果位置数据表明用户在星巴克咖啡店(例如通过使描述符与纬度 / 经度数据相关联的已知服务之一),那么模糊不清的发声可能被解析为“茶(tea)”,而在高尔夫球场上,同样的发声可能被解析为“球座(tee)”。)

[0640] 系统对讲话的响应可以取决于手机正在进行什么处理或者手机已经完成了什么处理而变化。例如,如果手机已经分析了街道场景并且叠盖了与不同的商店和餐馆相对应的视觉小玩意,那么用户讲出这些商店或餐馆之一的名称可以被认为是等同于对所显示的小玩意进行轻拍。如果被称为“鸭子”的酒吧在屏幕上具有小玩意,那么讲出名称“鸭子”可以使手机显示该酒吧的快乐时间菜单。相反,如果在步行中用户的手机已经识别出池塘中的野鸭、并且用户讲出“鸭子”,那么这可以唤出野鸭的 Wikipedia 页面的显示。此外,如果在十一月,手机识别出汽车窗户上的俄勒冈州大学的“O”标志并将相应的小玩意叠盖在用户的手机屏幕上,那么讲出单词“鸭子”可以唤出俄勒冈鸭子橄榄球队的名单或比赛时间表。(如果是在二月,那么同样的情况可以唤出俄勒冈鸭子篮球队的名单或比赛时间表。)因此,可以对同一个讲出的单词提供不同的响应,这取决于手机已经进行的处理(和 / 或随显示在手机屏幕上的标志而变化)。

[0641] 如刚刚提到的那样,响应也可以取决于位置、日时或其他因素而有所不同。在中午,讲出其小玩意已被显示的餐馆的名称可以唤出该餐馆的午餐菜单。在傍晚,可以改为显示正餐菜单。当希尔顿酒店在附近时讲出名称“希尔顿”可以使附近房地产的房价得到显示。(在底特律和纽约市说出同一个单词“希尔顿”会促使不同的房价得到显示。)

[0642] 对手机讲话允许指令的对话模式得以实现。响应于初始指令,手机可以进行初始的一组操作。看到响应于初始指令而采取的动作(或者由此得到的结果),用户可以发出进一步的指令。手机继而用进一步的操作做出响应。以迭代的方式,用户可以交互地引导手机产生用户期望的结果。在任何时刻,用户可以指示将会话保存起来,使得该迭代过程可以在之后继续进行。在“被保存”时,处理可以例如在云中继续进行,使得当用户在之后返回到交互过程时,可以获得额外的信息。

[0643] 可以基于用户偏好或应用场合以及隐私考虑而不同地实现“保存”。在一些情况下,只保存会话的摘要。摘要可以包括位置数据(例如,来自 GPS)、方向 / 方位数据(例如,来自磁力计)、和日期 / 时间。可以保留最初捕获的图像 / 音频,但是常常不这样做。而是,可以保存派生物。一种类型的派生物是内容指纹——一种从人类可理解的内容派生出的数据,但是根据该数据无法重建出所述人类可理解的内容。另一种类型的派生物是关键字向量数据,例如标识出形状的数据、单词、SIFT 数据、以及其他特征。另一种类型的派生数据是解码出的机器可读信息,例如水印或条型码有效载荷。也可以保存派生出的标识内容的数据,例如歌曲名称和电视节目名称。

[0644] 在一些情况下，可以保存最初捕获的图像 / 音频数据——假如从这样的数据所表示的个人接收到许可。如果派生数据与个人相关联(例如，面部识别矢量、声波纹信息)，那么该派生数据也可以要求许可进行保存。

[0645] 正如流行的摄像机在摄像机取景器中围绕察觉到的面部绘出矩形以指示出摄像机的自动聚焦和曝光所要基于的对象，ICP 装置也可以围绕装置屏幕上所呈现的视觉对象绘出矩形或者提供其他视觉标志，以便向用户通知影像中的什么东西将成为装置处理的焦点。

[0646] 在一些实施例中，不是通过讲出的线索或指令来引导装置的注意力(或者除此之外)，而是用户可以触摸屏幕上所显示的对象或者围绕该对象画圈，以便指示出装置应该将其处理努力集中于的对象。即使系统还没有显示(或者即使系统不显示)与所述对象相对应的小玩意，也可以启用该功能。

[0647] 传感器相关系统的声明配置方案

[0648] 本节进一步详述上面提到的一些概念。

[0649] 在现有技术中，智能电话已经为了诸如免提拨号和口头因特网查询(语义搜索)之类的目的而使用语音识别。根据本技术的某些实施例，把语音识别与对一个或更多基于传感器的系统的操作进行调谐这一手段相结合地使用，以便增强对用户期望的信息的提取。

[0650] 参考图 25，示例性智能电话 710 包括各种传感器，例如麦克风 712 和摄像机 714，它们各自具有相应的接口 716、718。智能电话的操作由处理器 720 控制，所述处理器 720 由存储在存储器 722 中的软件指令配置。

[0651] 智能电话 710 被显示为包括语音识别模块 724。该功能可以由智能电话的处理器 720 结合存储器 722 中的相关指令来实现。或者其可以是专用硬件处理器。在一些实施例中，该功能可以处于智能电话的外部——数据通过智能电话的 RF 蜂窝功能或数据收发器功能被传递给外部的语言识别服务器并从外部的语言识别服务器传递回来。或者语音识别功能可以分布在智能电话和远程处理器之间。

[0652] 在使用中，用户讲出一个或更多单词。麦克风 712 感测相关联的音频，并且接口电子装置 716 把由麦克风输出的模拟信号转换成数字数据。该音频数据被提供给语音识别模块 724，语音识别模块 724 返回识别出的讲话数据。

[0653] 用户可以讲出例如“听那个男人说话”。智能电话可以通过将男声滤波器应用于麦克风所感测到的音频来对该识别出的语音指令做出响应。(典型男性的讲话声音具有低至约 85 赫兹的基频，因此滤波器可以去除低于该值的频率。)如果用户说“听那个女人说话”，那么智能电话可以通过应用去除低于 165Hz (即，典型女性声音的范围下限)的频率的滤波功能来做出响应。在两种情况下，智能电话响应于这些指令而应用的滤波功能可以裁切出大约 2500 或 3000Hz (即，典型语音频带的上限)的音频频率。(音频滤波有时被称为“均衡化”，并且可以涉及对不同的音频频率进行增强以及衰减。)

[0654] 智能电话因此接收用户环境中的用户感兴趣的对象的口头指示(例如，“男人”)，并相应地配置其对所接收音频的信号处理。图 26 描绘出这样的方案。

[0655] 智能电话的配置可以通过建立与信号处理相关联地使用的参数(例如采样率、滤波器截止频率、水印密钥数据、要查阅的数据库的地址、等等)来完成。在其他方案中，所述配置可以通过执行与不同的信号处理操作相对应的不同的软件指令来完成。或者所述配置

可以通过激活不同的硬件处理电路或者将数据路由到外部处理器等来完成。

[0656] 如图 27 的表格摘选所示，在一个特定实现方案中，智能电话包括表格或其他数据结构，其使用户讲出的不同对象（例如，“男人”、“女人”、“无线电广播”、“电视”、“歌曲”、等等）与不同的信号处理操作相关联。把由语音识别引擎识别出的每个单词应用于该表格。如果任何识别出的单词匹配表格中标识出的“对象”之一，那么智能电话于是将指定的信号处理指令应用于此后接收到的音频（例如，当前会话中的音频）。在所描绘的实例中，如果智能电话识别出“男人”，那么智能电话将相应的男声滤波功能应用于音频，并将经滤波的音频传递给语音识别引擎。然后，将根据语音识别而输出的文本呈现在智能电话的显示屏上——按照该表格规定的指示。

[0657] 用户可以讲出“听无线电广播”。查阅图 27 的表格，智能电话通过试图借助对 Arbitron 数字水印进行检测来识别音频，而对识别出的讲话数据做出响应。首先以 6KHz 的采样频率对音频进行采样。然后对其进行滤波，并应用与 Arbitron 水印相对应的解码程序（例如，按照所存储的软件指令）。将解码出的水印有效载荷发送给 Arbitron 的远程水印数据库，并且与该无线电广播相关的元数据从该数据库被返回到该手持机。然后，智能电话在其屏幕上呈现该元数据。

[0658] 如果在音频中没有发现 Arbitron 水印，那么表格中的指令规定备选的一组操作。特别地，该“否则（Else）”条件指示智能电话应用与对象“歌曲”相关联的操作。

[0659] 与“歌曲”相关联的指令开始于以 4KHz 对音频进行低通滤波。（先前捕获的音频数据可以在存储器中缓冲，以便允许对先前捕获的刺激的这种再处理。）然后（使用独立存储的指令）计算 Shazam 歌曲识别指纹，并将所得的指纹数据发送给 Shazam 的歌曲识别数据库。在该数据库中查找相应的元数据并将其返回给智能电话进行显示。如果没有发现元数据，则显示内容指示未识别该音频。

[0660] （应理解的是，所详述的信号处理操作可以在智能电话上执行，或者由远程处理器（例如，在“云”中）执行，或者以分布式方式执行。应进一步理解的是，图 27 中所示的信号处理操作只是基于用户输入可以触发的信号处理操作和操作序列所构成的大集合的小子集。当表格中详述的指令中没有规定各参数时，可以使用默认值，例如对于采样率取 8KHz，对于低通滤波取 4KHz，等等。）

[0661] 一些智能电话包括两个或更多麦克风。在这样的情况下，由用户输入触发的信号处理指令可以涉及对麦克风阵列进行配置，例如通过控制各麦克风对组合而成的音频流的相位和幅度贡献。或者，所述指令可以涉及分别处理来自不同的麦克风的音频流。这对于例如声音定位或讲话者识别是有用的。可以应用额外的信号调整操作来改善对期望的音频信号的提取。通过传感器融合技术，讲话者的位置可以尤其基于摄像机和姿态估算技术而得到估算。一旦识别出声源，在存在多个麦克风的情况下，可以利用波束形成技术来隔离出讲话者。在一系列样本的基础上，表示信道的音频环境可以被模拟并去除，以便进一步改善对讲话者声音的恢复。

[0662] 智能电话通常包括除麦克风以外的传感器。摄像机是普遍存在的。其他传感器也是常见的（例如，RFID 和近场通信传感器、加速计、陀螺仪、磁力计、等等）。可以类似地采用用户语音来配置这些其他传感器数据的处理。

[0663] 在一些实施例中，该功能可以通过由用户讲出独特的关键字或措词（例如

“DIGIMARC 看”或“DIGIMARC 听”)来触发,从而启动应用程序并向装置提示后面跟随的单词不仅仅是命令。(在另一些实施例中,可以提供不同的提示——口头的或者其他方式(例如手势)。在又一些实施例中,这样的提示可以被省略。)

[0664] 例如,“DIGIMARC 看电视”可以唤出一个专用的命令程序库以便触发由信号处理操作(例如设定帧捕获率,应用某些滤色器、等等)构成的序列。“DIGIMARC 看这个人”可以启动这样的过程,该过程包括 :进行颜色补偿以获得精确的肤色、提取面部信息、以及将面部信息应用于面部识别系统。

[0665] 再一次,表格或其他数据结构可以用于使相应的信号处理操作与不同的动作和感兴趣的对象相关联。在表格中为其指出指令的不同对象是“报纸”、“书”“杂志”、“海报”、“文本”、“印刷物”、“票券”、“盒子”、“包裹”、“纸箱”、“包装纸”、“产品”、“条形码”、“水印”、“照片”、“相片”、“人”、“男人”、“男孩”、“女人”、“女孩”、“他”、“她”、“他们”、“人们”、“显示器”、“屏幕”、“监视器”、“视频”、“电影”、“电视”、“无线电广播”、“iPhone”、“iPad”、“Kindle”、等等。相关联的操作可以包括 :应用光学字符识别,数字水印解码,条型码读取,计算图像或视频指纹,以及辅助图像处理操作和参数(例如颜色补偿、帧速率、曝光时间、焦距、滤波、等等)。

[0666] 可以利用额外的冗词来帮助用对象描述符、颜色、形状或位置(前景、背景、等等)对视觉场景进行分割。跨越多个样本,可以利用时间描述符(例如眨眼、闪光),可以利用额外的运动描述符(例如快速或缓慢)。

[0667] 含有使得能够识别装置的运动的传感器的装置会增加另一层控制单词,这些控制单词陈述装置和期望的对象之间的关系。诸如“跟踪”之类的简单命令可以指示装置应该分割视觉或听觉场景以便仅包含轨迹与装置的运动近似的那些对象。

[0668] 在更精巧的方案中,智能电话包括几个这样的表格,例如用于音频刺激的表格 1、用于视觉刺激的表格 2、等等。智能电话可以基于识别出的用户讲话中的其他术语和 / 或语法来决定要使用哪个表格。

[0669] 例如,如果识别出的用户讲话包括诸如“看”、“注视”、“观察”、“注意”或“阅读”之类的动词,则这可以向智能电话告知用户所感兴趣的是视觉刺激。如果在用户的讲话中检测到这些单词中的一个,那么智能电话可以把来自用户的识别出的讲话的其他单词或语法应用于表格 2。相反,如果识别出的用户讲话包括诸如“收听”或“听”之类的动词,那么这会指示出用户对听觉刺激感兴趣并且应该查阅表格 1。

[0670] 通过这样的基于规则的方案,智能电话对两个讲出的短语“DIGIMARC 看这个男人”和“DIGIMARC 听这个男人说话”做出不同的响应。在前一种情况下,查阅表格 2 (与由摄像机拍摄的视觉刺激相对应)。在后一种情况下,查阅表格 1 (与由麦克风捕获的听觉刺激相对应)。图 28 和 29 示出这些系统的实例。

[0671] (本领域技术人员应理解的是,所描述的表格方案只是借以实现所详述的功能的许多方式之一。本领域技术人员会类似地认识到,除了上面详述的那些单词以外的各种各样的动词及其他单词都可以被解释为是关于用户对视觉刺激感兴趣还是对听觉刺激感兴趣的提示。)

[0672] 有时,讲出的名词也会揭示关于刺激类型的一些信息。在短语“DIGIMARC 看这本杂志”中,“DIGIMARC”会唤出专用程序库和操作,“看”暗示视觉刺激,并且“杂志”也告知了

关于视觉刺激的一些信息，即它包含静态的印刷图像和 / 或文本(这可以通过使用“阅读”而不是“看”来区别)。相反，在短语“DIGIMARC 看电视”中，术语“电视”指示该内容具有时间特征，使得拍摄多个帧用于分析是适当的。

[0673] 应认识到的是，通过使不同的参数和 / 或信号处理操作与不同的关键字词条相关联，智能电话基本上通过讲出的用户输入而被重新配置。在一个时刻，智能电话被配置为无线水印检测器。接着，智能电话被配置为面部识别系统。等等。对传感器相关系统进行动态调谐以便服务于用户的明显兴趣。此外，用户通常不明确地声明某一功能(例如，“读取条形码”)，而是确定某一对象(例如，“看这个包裹”)，并且智能电话推断期望的功能(或者推断由可能的功能构成的分级结构)，并相应地改变智能电话系统的操作。

[0674] 在涉及相同操作(例如，数字水印解码)的一些情况下，操作的细节可以取决于特定对象而变化。例如，杂志中的数字水印由于所使用的油墨、介质和印刷技术之间的差别，而通常使用与报纸中嵌入的数字水印不同的编码参数来编码。因此，“DIGIMARC，看这本杂志”和“DIGIMARC，看这张报纸”都可能会涉及数字水印解码操作，但是前者可以利用与后者不同的解码参数(例如，相关颜色空间、水印缩放比例、有效载荷、等等)。(在后面的实例中将略去前奏“DIGIMARC”，但是本领域技术人员应理解的是，仍然可以使用这样的提示。)

[0675] 不同的对象通常可以与不同的摄像机观察距离相关联。如果用户发出“看这本杂志”的指示，那么智能电话可以(例如根据表格中存储的其他信息而)了解到该对象将会处于大约 8 英寸远的地方，并且可以指示机械或电子系统将摄像机系统聚焦在该距离。如果用户发出“看电子布告板”的指示，那么作为对比，摄像机可以聚焦在 8 英尺的距离处。智能电话预期会辨别出的图像特征的缩放比例可以类似地得到建立。

[0676] 有时，用户的口头指令可以包括否定词，例如“不是”或“不”或“忽视”。

[0677] 考虑这样的情况：智能电话通常通过检查拍摄的图像数据以获得条型码，来响应用户讲出的“看那个包裹”。如果发现条形码，则对该条型码进行解码，在数据库中查找有效载荷，然后在屏幕上呈现所得的数据。如果没有发现条型码，那么智能电话求助于所存储数据中的“否则(Else)”指令(例如分析拍摄的图像数据以获得水印数据)，并将任何解码出的有效载荷数据提交给水印数据库以获得相关的元数据，然后将该元数据显示在屏幕上。(如果没有发现水印，那么另外的“否则”指令可以使智能电话检查影像以获得可能的文本，并将任何这样的摘选提交给 OCR 引擎。然后将来自 OCR 引擎的结果呈现在屏幕上。)

[0678] 如果用户说出“看那个包裹；忽视条形码”，那么这会改变通常的指令流。在这种情况下，智能电话不会试图解码来自所拍摄影像的条形码数据。而是，智能电话直接进行到第一个“否则”指令，即检查影像以获得水印数据。

[0679] 有时，用户可以不特别确定某一对象。有时，用户可以仅提供否定词，例如“没有水印”。在这样的情况下，智能电话可以把按优先顺序排列的一系列内容处理操作应用于刺激数据(例如，按照所存储的列表)，从而跳过根据用户的讲话而被指示(或推断)为不适用的操作的那些操作。

[0680] 当然，对感兴趣对象的口头指示可以被理解为对其他潜在感兴趣对象的否定，或者可以被理解为对可能应用于刺激数据的其他类型的处理的否定。(例如，“看那个男人”提示智能电话不需要检查影像来获得数字水印或条型码。)

[0681] 因此应理解的是，用户的声明能帮助智能电话的处理系统决定要采用什么样的识

别技术和其他参数来最好地满足用户的可能期望。

[0682] 适合于供本技术使用的语音识别软件可从 Nuance 通信公司获得(例如该公司的 SpeechMagic 和 NaturallySpeaking SDK)。免费语音识别软件(例如,可根据开源许可证利用)包括由 Carnegie Mellon 大学提供的 Sphinx 系列产品。这包括 Sphinx 4 (JAVA 实现方案)和 PocketSphinx (对 ARM 处理器上的使用进行优化的简化版本)。其他免费语音识别软件包括 Julius (由在交互语音技术联盟(Interactive SpeechTechnology Consortium)中合作的日本大学的联盟提供)、ISIP (来自密西西比州)和 VoxForge (可与 Sphinx、Julius 和 ISIP 一起使用的开源语音语料库和声学模型)。

[0683] 尽管在参考用户的口头言语来感测用户兴趣的背景环境下进行描述,但是也可以采用其他类型的用户输入。可以采用视线(眼睛)跟踪方案来识别用户正在察看的对象。通过手或激光指示器做出的指示运动可以同样被感测并用于识别感兴趣的对象。可以使用不涉及用户与智能电话进行触觉交互(例如通过键盘或者通过触摸手势)的各种这样的用户输入。图 30 大体描绘出这样的方案。

[0684] 在一些实施例中,由智能电话应用的信号处理也可以部分地基于背景环境信息。

[0685] 如其它地方讨论的那样,“背景环境”的一种定义是“可以用于表征实体(被认为与用户和应用程序(包括用户和应用程序本身)之间的交互相关的人、地点或对象)的情况的任何信息”。背景环境信息可以具有许多种类,包括计算背景环境(网络连接性、内存可用性、CPU 争用、等等)、用户背景环境(用户概况、位置、动作、偏好、附近朋友、社交网络和境遇、等等)、物理背景环境(例如,光照、噪声水平、交通、等等)、时间背景环境(日时、日、月、季节、等等)、内容背景环境(主题、行为者、风格、等等)、上述背景环境的历史、等等。

[0686] 关于视觉操作和相关概念的更多说明

[0687] 因为本技术的某些实施例能够在装置上的资源和“云”之间动态地分配期望的任务,所以本技术的某些实施例非常适合于在存储和计算资源有限的背景环境中优化应用程序响应。

[0688] 对于复杂的任务(如确认钞票的面额),人们可以把整个任务提交给具有最高时间效率或成本效率的提供商。如果用户想要识别美国钞票、并且找到了能够完成该任务的外部提供商(例如投标人),那么高级别的任务可以在云中执行。为了提高效率,云服务提供商可以使用由装置上执行的子任务(例如处理图像数据以使所需的外部带宽最小化,或者对图像数据进行过滤以便去除使个人可被识别出来的数据或无关数据)所提取的图像特征数据。(这种本地处理的数据也可以同时处于可供其他本地和远程任务利用的状态。)

[0689] 在一些方案中,本地装置不知道外部提供商进行的处理的细节,本地装置仅被通知所需的输入数据的类型和格式以及将被提供的输出数据的类型 / 格式。在另一些方案中,提供商公布关于在执行其处理的过程中所应用的特定算法 / 分析的信息,使得本地装置能够在备选提供商之间做选择时考虑该信息。

[0690] 在计算模型聚焦于始终能够在装置上执行的某些任务的意义上,这些基本操作会被裁制成适合于为各装置预见到的可能需要的云应用程序的类型。例如,如果应用程序需要钞票或其他文件的具有特定分辨率、对比度和覆盖率的图像,那么所提供的“图像获取”功能将会需要匹配能力。

[0691] 通常,自顶向下的思维提供了装置所要提供的一些非常具体的低级特征和能力。

在这一点上，设计者将会集思广益一点。这些低级特征和能力会建议什么样的更有用的特征或能力？一旦已经编辑出这种通常有用的能力的列表，就可以选择一套基本操作并预作安排以使内存和功率需求最小化。

[0692] 作为旁白，Unix 已经长期利用能够使中间存储最少化的“过滤器链”。为了执行由多个变换构成的序列，那么对每个步骤提供可级联的“过滤器”。例如，假定变换 A->B 实际上是序列：

[0693] A|op1|op2|op3>B

[0694] 如果每个步骤都要将一条目变为同样或相似尺寸的新条目，并且假定 A 在结束时仍然可获得，那么内存需求是尺寸 (A)+ 尺寸 (B)+2 个缓冲器，每个缓冲器通常比完整对象尺寸小得多、并且在操作完成时被释放。例如，复杂的本地变换可以通过以这种方式组合一些简单的本地操作来获得。存储量和所执行的操作次数都可以得到减少，从而节省时间、功率或者时间和功率这两者。

[0695] 自然，至少一些应用程序被构想为用短图像序列作为输入。系统设计可以通过提供短的、或许是长度固定(例如，三或四或 40 个帧)的图像序列缓冲器来支持这种想法，所述图像序列缓冲器是每个图像获取操作的目的地。变化的应用程序需求可以通过提供各种向缓冲器写入的方式(插入一个或更多新图像 FIFO ；一个或更多新图像经由过滤器(最小、最大、平均、…) 被组合起来，然后插入 FIFO ；一个或更多新图像与对应的当前缓冲器元素经由随后插入的过滤器被组合起来，等等) 来支持。

[0696] 如果图像序列由以特定方式填充的固定尺寸缓冲器表示，那么从一序列中提取图像的操作会由从缓冲器中提取图像的操作替代。每个这样的提取操作可以从缓冲器中选择一组图像，并经由过滤器将它们组合起来以形成所提取的图像。在提取之后，缓冲器可以保持不变，可以去除一个或更多图像，或者可以使它的一些图像通过基本图像操作而得到更新。

[0697] 存在着至少三种类型的图像子区域，它们通常在模式识别中被使用。最通常的只是一组提取出的点，这些点的几何关系原封未动，通常是作为一系列点或行片段。下一种是图像的连通区域，或许作为一系列连续的行片段。最后一种是矩形子图像，或许作为像素值的阵列和在图像内的偏移量。

[0698] 在已经决定这些特征类型中要支持的一个或更多特征类型的情况下，可以关于效率或通用性来选择表示形式——例如，位于图像上任何地方的“1 维”曲线只是一个像素序列，并且因此是一种斑点。因此，两者都可以使用相同的表示形式，因此所有相同的支持功能(内存管理等) 也可以使用相同的表示形式。

[0699] 一旦选定了表示形式，任何斑点“提取”都可能是单一的两步骤操作。第一步 ：定义斑点“主体”，第二步 ：从图像中复制像素值到它们对应的斑点位置。(这可以是“过滤器”操作，并且可以模仿产生图像而且适用于静态图像的任何过滤器操作序列。)

[0700] 即使对于图像，对处理的“拍卖”过程也可以涉及使得各操作可被用于内部格式与适当的外部格式之间的转换。对于斑点和其他特征，相当多种类的格式转换可能会得到支持。

[0701] 或许有用的是，从图像处理或计算机视觉包的“正常”讨论离题一点，从而返回到可在详述的方案中运行的应用程序、以及所涉及的(非典型) 限制和自由度的本质。

[0702] 例如，尽管一些任务将会由直接的用户动作“触发”，但是另外一些任务可以简单地在适当的时候被启动并且预期会触发其自身。即，用户可能将智能电话瞄准停车场并触发“找到我的车”应用程序，该应用程序会快速拍摄一张图像并设法分析它。更可能地，用户更喜欢触发应用程序，然后在停车场中徘徊，到处摇动摄像机镜头，直到装置告知汽车已经得到识别。然后显示器可以呈现从用户的当前位置拍摄的图像，使汽车加亮显示。

[0703] 尽管这种应用程序可能会变得普及或者可能不会变得普及，但是很可能许多应用程序会包含这样的处理循环：图像被获取、采样并检查可能存在的目标，检测到该目标会触发“真实的”应用程序，该应用程序会带来更多的计算能力以对候选图像施加作用。该处理继续进行，直到应用程序和用户同意该处理已经成功，或者明显未成功会使用户终止它。合乎期望的是，“试探性的检测”循环应该能够仅在摄像机上运行，并且任何外部资源仅当存在着希望这些外部资源可能有用的原因时才被调入。

[0704] 另一种应用程序用于跟踪对象。这里，已知类型的对象已经被定位（不管是如何被定位的），此后获取到一连串图像，并且该对象的新位置被确定并指示出来，直到该应用程序被终止或者对象消失。在这种情况下，应用程序可能使用外部资源来在最初定位该对象，并且非常可能会使用这些外部资源来使已知的检测模式专门用于已经检测到的特定实例，而随后的使用新模式实例的“跟踪”应用程序合乎期望地在手机上不受协助地运行。（或许这种应用程序会帮助在运动场上留意一个小孩。）

[0705] 对于一些应用程序，模式识别任务可能是相当粗略的——或许是跟踪帧序列中的一块蓝色（例如，毛线衫），而在另外一些应用程序中模式识别任务可能是高度复杂的（例如，鉴定钞票）。很可能的是，相当少数量的控制循环（类似上面提到的两种控制循环）对于大量的简单应用程序而言是足够的。它们的不同之处在于所提取的特征、所采用的模式匹配技术、以及所求助的外部资源（如果有的话）的特性。

[0706] 如上所示，至少一些模式识别应用程序可以在基本的移动装置上天然地运行。并不是所有的模式识别方法都适合于这些受限制的平台。可能性包括：简单模板匹配，尤其是与非常小的模板或使用非常小的元素的合成模板；霍夫式匹配，对于所检测的参数有适度的分辨率要求；以及神经网络检测。应注意的是，对该神经网络进行训练可能会需要外部资源，但是应用该训练可以在本地完成，尤其是在可以采用 DSP 或图形芯片的情况下。采用大型数据库查找或者过于计算密集的任何检测技术（例如 N 空间最近邻居）或许最好使用外部资源来完成。

[0707] 关于聚簇的更多说明

[0708] 如前所述，聚簇是指用于将像素群识别为彼此相关的处理。

[0709] 一种特定方法是利用“共同的命运”（例如共同拥有共同的运动）使各场景条目分群。另一种方法依赖多阈值或尺度空间树。数据结构（包括黑板）可以存储表示借以识别聚簇的方法的符号标签，或者聚簇可以与表示其类型的标签一起存储。（识别代理可以对标签字典做出贡献。）

[0710] 所述标签可以根据所使用的聚簇方法和所涉及的特征（例如，与亮度边缘相结合的颜色均一性）来得出。在最低级别上，可以使用“局部亮度边缘”或“最均一的颜色”。在较高级别上，可以采用诸如“类似的均一性水平，接近但是被局部亮度边缘分开”之类的标签。在更高级别上，可以基于来自识别代理的信息来对各聚簇赋予诸如“像树叶”或“像脸

部”之类的标签。其结果是填充有带标签的特征的 n 维空间，从而（可能在将各特征投影到特定平面上时）便利于更高阶的识别技术。

[0711] 共同运动方法考虑点 / 特征在图像之间的 2D 运动。运动可以是 :例如，几乎相同的位移，或者沿着图像方向的几乎线性的位移，或者绕着图像点的几乎共同的旋转。也可以使用其他方法，例如光流、点群、运动向量、等等。

[0712] 多阈值树法可以用于使图像内的呈树结构的嵌套斑点相互关联。图 20A 和 20B 是说明性的。简要地，对图像（或摘选）进行阈值处理——检查每个像素值以确定其满足阈值还是超过阈值。最初，阈值可以被设定为黑色。每个像素都符合该标准。然后升高阈值。图像的一些部分开始不满足阈值测试。在阈值测试得到满足的地方会出现一些区域（斑点）。最终，阈值达到明亮（高）级别。只剩下一些小区域仍然能通过该测试。

[0713] 如图 20A 和 20B 所示，整个图像都通过黑色阈值。在暗阈值下，单个斑点（矩形）满足该测试。随着阈值的增大，两个椭圆形斑点区域区分开来。继续将阈值升高到明亮值会使第一区域分成两个明亮的椭圆形，并且第二区域转变成单个小明亮区域。

[0714] 对照这种变化的阈值对像素值进行测试可提供一种快速的检查方法来识别图像帧内相关的像素聚簇。

[0715] 在实际的实现方案中，图像首先可以利用高斯或其他模糊来处理以防止轻微的噪声伪像不适当地影响结果。

[0716] （该方法的变型可以充当边缘检测器。例如，如果尽管阈值升高了几个值，多个斑点之一的轮廓仍保持大体固定，那么该轮廓被辨别为边缘。边缘的强度由使轮廓基本上保持固定的阈值范围表示。）

[0717] 尽管详述了对照亮度值进行阈值处理，但是也可以类似地对照其他阈值度量（例如颜色、纹理度、等等）来进行比较。

[0718] 通过这种方法识别出的聚簇可以充当用于其他数据（如图像特征和关键字向量）的组织构造。例如，一种用于识别从图像数据中提取出的特征 / 关键字向量相互关联的方法是识别包含这些特征 / 关键字向量的最小阈值斑点。该斑点越小，这些特征可能就更相关。类似地，如果第一和第二特征已知是相关的，那么其他相关的特征可以通过寻找包含前两个特征的最小阈值斑点来估计。该斑点内的任何其他特征也可能与第一和第二特征相关。

[0719] 自由度和限制因素

[0720] 一些模式识别方法的实用性取决于平台的在应用程序请求时执行浮点操作或调用 DSP 的矢量操作的能力。

[0721] 更一般地，在直觉计算平台上存在着许多具体的自由度和限制因素。自由度包括任务利用装置外的资源（不管是处于附近的通信辅助装置上，还是处于云中）、从而允许“不可能”在装置上运行的应用程序看上去似乎能这样做的能力。限制因素包括 :有限的 CPU 功率，有限的可用内存，以及应用程序在资源不断变化的情况下需要继续进行的需求。例如，可用的内存可能不仅会受到限制，而且可能会突然被减少（例如在开始通电话时）、然后在更高优先级的应用程序终止时再次变得可用。

[0722] 速度也是限制因素——通常与内存的关系紧张。对迅速响应的期望可能会甚至把寻常的应用程序向上推到靠近内存上限。

[0723] 在特征表示方面，内存限制会鼓励维持有序的元素列表(内存需求与条目数成比例)而不是数值的明确阵列(内存需求与可能的参数的数量成比例)。操作序列可能会使用最少的缓冲器(如上所述)而不是完整的中间图像。长的图像序列可能会通过短的实际序列以及一个或更多平均结果来“伪造”。

[0724] 一些“标准”图像特征(如 Canny 边缘算子)对于通常的用途而言可能过于资源密集。然而，在以前关于 FFT 处理也有过同样的评价，但是智能电话应用程序正越来越多地采用该操作。

[0725] 适合于考虑的装置内处理

[0726] 在上述限制因素的背景环境中，下面的概要详述了可包含在本地装置的指令系统(repertoire)中的广泛有用的操作的种类：

[0727] I. 任务相关操作

[0728] A. 与图像相关

[0729] i. 图像序列操作

[0730] a) 从序列中提取图像

[0731] b) 从序列范围中生成图像

[0732] c) 贯穿该序列跟踪特征或 ROI

[0733] ii. 图像变换

[0734] a) 逐点重新映射

[0735] b) 仿射变换

[0736] c) 本地操作：例如边缘、本地平均、…

[0737] d)FFT 或相关操作

[0738] iii. 从图像中提取视觉特征

[0739] a)2D 特征

[0740] b)1D 特征

[0741] c) 近乎 3D 的特征

[0742] d) 完整图像 ->ROI 列表

[0743] e) 非本地特征(颜色直方图、…)

[0744] f) 缩放、旋转不变强度特征

[0745] iv. 特征操纵

[0746] a) 来自 2D 特征的 2D 特征

[0747] b)1D 到 1D、等等

[0748] c) 来自 2D 特征的 1D 特征

[0749] v. 用户界面图像反馈(例如，在图像上叠盖与标签相关的符号)

[0750] B. 模式识别

[0751] i. 从一组特征集合中提取图案

[0752] ii. 使序列、图像或特征集合与标签相关联

[0753] iii. 从特征集合中“识别”标签或标签集合

[0754] iv. 从较简单的一组“识别出的”标签中“识别”合成的或复杂的标签

[0755] C. 与应用程序相关的通信

[0756] i. 从系统状态中提取必要功能的列表

[0757] ii. 广播对投标的请求——收集响应

[0758] iii. 发送精炼出的数据，接收外包结果

[0759] II. 与动作相关的操作(许多操作将准备好存在于基本系统动作中)

[0760] i. 激活 / 停用系统功能

[0761] ii. 产生 / 消耗系统消息

[0762] iii. 检测该系统状态

[0763] iv. 使系统转变到新状态

[0764] v. 维持待决的、活跃的、和已完成的动作的队列

[0765] 用户体验和用户界面

[0766] 本技术的一个特定实施例允许未经训练的用户通过使用移动装置发现关于他所处环境(和 / 或关于他所处环境中的物体)的信息，而无需决定使用哪些工具，同时该特定实施例提供了在期望的任何时间和地点继续进行被中断的发现体验的能力。

[0767] 读者将会认识到的是，现有的系统(如 iPhone)无法满足这种需求。例如，用户必须决定应该启动数千个不同的 iPhone 应用程序中的哪一个(哪些)来提供所期望的特定类型的信息。并且，如果在指引操作时用户被中断，那么就没有办法在后来的时间或地点继续该发现处理。即，用户必须在与物体或环境交互时的时间点经历该发现过程。无法“保存”该体验以便在之后探查或共享。

[0768] 图 19 示出具有说明性用户界面的智能电话 100，该说明性用户界面包含屏幕 102 和发现按钮 103。

[0769] 发现按钮 103 是被硬接线或者被编程为使智能电话起动其发现模式(分析输入的刺激以辨别含义和 / 或信息)。(在一些模态中，智能电话始终分析这种刺激，并且不需要按钮动作。)

[0770] 所绘出的屏幕 102 具有顶部长方格部分 104 和下部长方格部分 106。两个长方格的相对大小由滑条 108 控制，该滑条 108 将所绘出的两个长方格分开。滑条 108 可以使用为图形用户界面设计者所熟悉的构造，由用户拖动从而使顶部长方格更大或者使底部长方格更大。

[0771] 说明性的底部长方格 106 用来呈现空间信息(如地图、图像、GIS 层、等等)。这可以被称为地理位置长方格，尽管这不应该被解释为是对其功能性进行限制。

[0772] 说明性的顶部长方格 104 在下面的讨论中被称为传感器(尽管这同样不是限制性的)。在所示的模式中，该长方格呈现音频信息，即听觉场景可视化。然而，在 UI 上呈现借以把该顶部长方格切换成呈现视觉信息(在这种情况下在按钮上于是显示音频(AUDIO)，从而允许用户切换回来)的按钮 131。其他类型的传感器数据(如磁力计、加速计、陀螺仪、等等)也可以呈现在该长方格中。

[0773] 从顶部长方格开始，智能电话中的一个或更多音频传感器(麦克风)侦听音频环境。讲话者 / 语音识别软件分析捕获的音频，以便试图识别出讲话的人并辨别所讲的话。如果实现匹配(使用例如存储在本地或云中的讲话者表征数据)，那么沿着显示器的边缘呈现与识别出的讲话者相对应的图标 110。如果智能电话可以使用识别出的讲话者的所存储的图像 110a (例如，来自用户的电话簿或来自 Facebook)，那么该图像可以用作图标。如果不

能使用图像 110a，则可以采用默认图标 110b。（如果识别软件能够做出具有规定置信度阈值的性别确定，则可以对男性和女性讲话者采用不同的默认图标。）所绘出的 UI 示出已经检测到两个讲话者，尽管在其他情况下可能会存在更多或更少的讲话者。

[0774] 除了语音识别之外，诸如水印检测和指纹计算 / 查找之类的处理可以应用于音频流以识别讲话的人和所讲的话。通过这些或其他方法，软件可以检测出环境音频中的音乐，并呈现指示这种检测结果的图标 112。

[0775] 也可以检测并指示出其他不同的音频类型（例如，路面噪声、鸟叫声、电视、等等）。

[0776] 在每个图标（110、112、等等）的左边是波形显示部 120。在所绘出的实施例中，显示基于实际数据的波形，尽管根据需要可以使用千篇一律的绘图。（可以使用其他表示形式，如谱直方图。）所示出的模拟波形向左移动，使最新的数据留在右边（类似于我们在阅读一行文字时的体验）。在向左移动出视线之前，只呈现每个波形的最新的一段时间间隔（例如，3、10 或 60 秒）。

[0777] 将环境音频分割成不同的波形只是一种近似；精确的分离是困难的。在采用两个不同的麦克风的简单实施例中，确定两个音频流之间的差分信号，从而提供第三音频流。当感测到第一个讲话者在讲话时，呈现这三个信号中较强的一个（波形 120a）。当该讲话者不在讲话时，以大大衰减的刻度来呈现该波形（或另一波形）——表明他已变得沉默（尽管环境音频水平可能在级别上还没有减弱很多）。

[0778] 对于用图标 110b 表示的第二个讲话者也同样如此。当识别出该人的声音（或者辨别出人声，但不能识别出是谁的声音——但是已知不是由图标 110a 表示的讲话者）时，于是以波形形式 120b 显示三个音频信号中声音最大的一个。当该讲话者变得沉默时，呈现衰减了很多的波形。

[0779] 类似地呈现波形 120c 以表示感测到的背景音乐。可以呈现来自三个声源中与讲话者的音频关联度最小的那个声源的数据。此外，如果音乐被中断，那么波形可以由软件衰减以表明这种中断情况。

[0780] 如上所述，只有几秒的音频是用波形 120 表示的。同时，智能电话正在分析该音频，辨别含义。该含义可以包括：例如，该讲话者的语音识别文本，和音乐的歌曲识别结果。

[0781] 当辨别出有关音频流的信息时，该信息可以由小玩意（图标）122 表示。如果该小玩意与仍然在横穿屏幕的波形所表示的音频摘选相对应，则该小玩意可以被放置到与该波形相邻，诸如小玩意 122a（其可以指示例如讲话者最近所说的话的文本文件）。该小玩意 122a 与他所对应的波形一起向左移动，直到该波形在虚拟的停止门 123 处从视线中消失。在该点，小玩意被穿到短线 124 上。

[0782] 小玩意 122 在线 124 上排成队，就像一条绳上的珍珠那样。线 124 的长度仅足够保持有限数目的小玩意（例如，两个到五个）。在线被穿满后，每个追加的小玩意将最老的小玩意推出视线。（消失的小玩意仍然可从历史文件中获得。）如果没有新的小玩意到达，现有的小玩意可以被设定成在一段时间间隔之后“死去”，使得它们从屏幕消失。该时间间隔可以由用户配置；示例性的时间间隔可以是 10 或 60 秒、或者 10 或 60 分钟、等等。

[0783] （在一些实施例中，甚至在已经辨别出任何相关信息之前，就可以将原型小玩意与波形或其他特征相关联地呈现。在这种情况下，轻拍原型小玩意会使智能电话将其处理注意力集中于获得与相关特征有关的信息。）

[0784] 小玩意 122 可以包括可见标志以便用图形形式指示其内容。如果例如识别出一首歌曲，那么相应的小玩意可以包含相关联的 CD 封面插图、艺术家的脸、或者音乐发行商的标志(如小玩意 122b)。

[0785] 另一种音频场景可视化方法通过参考不同的音频流相对于智能电话的方向来识别并描绘这些音频流。例如，一个波形可能会被显示为从右上方进入；另一个波形可能会被显示为从左侧到来。处于中心位置的中心部充当这些波形的停止门，小玩意 122 累积在这些波形上(如同串在绳 124 上)。轻拍中心部会调出所存储的历史信息。这样的方案在图 19A 中示出。

[0786] 由智能电话做出的所有动作和发现的历史可以在本地和 / 或远程被编辑并存储。所存储的信息可以仅包括发现的信息(例如，歌曲名称、讲话文本、产品信息、电视节目名称)，或者所存储的信息可以包括更多信息(如音频流的录制版本，和由摄像机拍摄的图像数据)。如果用户通过适当的概况设定进行了选择，那么该历史可以包括在会话中由智能电话处理的全部数据(包括关键字向量、加速计和所有其他传感器数据、等等)。

[0787] 附加地或者可选地，用户界面可以包括“保存”按钮 130。用户启动该控制件会使系统的信息状态被存储。另一种用户控制件(未示出)允许所存储的信息被复原到系统中，使得装置分析和用户发现能够继续进行——甚至是在不同的地方和时间。例如，如果用户在书店翻阅图书并且传呼器(pager)召唤他去附近餐馆的空闲桌位，那么用户可以按下“保存”。之后，该会话可以被调出，并且用户可以继续该发现，例如让装置参考感兴趣的书的封面套纸图或条形码来查找这本书，以及让装置识别背景中播放的歌曲。

[0788] 尽管图 19 在传感器长方格 104 中示出关于音频环境的信息，但是可以采用类似的构造来呈现关于视觉环境的信息，例如使用本说明书中其他地方详述的方案。如上所述，轻拍摄像机按钮 131 会使程式从音频切换到视觉(以及从视觉切换到音频)。在视觉模式下，该传感器长方格 104 可以用于显示交互的增强现实模式。

[0789] 转向图 19 下方的地理位置长方格 106，其示出地图数据。该地图可以从在线服务(如 Google Maps、Bing 等等)下载。

[0790] 地图数据的分辨度 / 粒度最初取决于智能电话借以知道其当前位置的粒度。如果已知高度精确的位置信息，那么可以呈现示出精细细节的地图(例如，被放大)；如果只已知大体位置，那么呈现示出较少细节的地图。如同常规技术那样，用户可以通过比例控制件 140 来放大或缩小地图以获得更多或更少细节。用户的位置由较大的推针 142 或其他标志表示。

[0791] 每次用户例如通过轻拍所显示的小玩意来进入发现会话或进行发现操作时，更小的推针 146 留驻在地图上，从而记住遭遇地点。关于发现操作的信息(包括时间和地点)与推针相关联地存储。

[0792] 如果用户轻拍推针 146，那么从存储装置中调用关于先前进行的发现的信息，并将其呈现在新窗口中。例如，如果用户具有关于商场中的一双长靴的发现体验，那么可以显示长靴的图像(用户拍摄的或者库存的照片)以及在先前的遭遇期间呈现给用户的价格和其他信息。另一个发现可能会涉及识别夜总会中的歌曲或者识别教室中的面部。所有这些事件都通过所显示地图上的推针来记忆。

[0793] 地理位置长方格通过时间控制件 144 (例如，图形滑条)来使对先前进行的发现的

回顾变得容易。在一个极端，不示出任何先前进行的发现(或者仅示出过去一小时内的发现)。然而，通过使该控制件发生变化，可使地图填充有额外的推针146，每个额外的推针146指示先前的发现体验和其发生的位置。该控制件144可以被设定成示出例如过去一星期、一月或一年内的发现。可以激活“H”(历史)按钮148以使滑条144出现，从而允许访问过去的发现。

[0794] 在一些地理位置(例如，商场或学校)，用户的发现历史可能是如此丰富以致于必须对推针进行过滤以便不会乱七八糟地堆满地图。因此，一种模式允许发现的开始和结束日期由用户设定(例如，通过一对类似滑条144的控制件)。或者可以通过相应的UI控制来应用关键词过滤器，例如诺德斯特龙百货公司(Nordstrom)、长靴、音乐、面部、人名、等等。

[0795] 指南针箭头146呈现在显示器上，以帮助理解地图。在所绘出的模式中，地图上的“向上”方向是智能电话所指向的方向。如果轻拍箭头146，则箭头快速移动到垂直取向。然后地图被旋转使得地图上的“向上”方向对应于北。

[0796] 用户可以使得数量多少与自己的期望相应的、关于自己的动作的信息可供别人获得从而与其他人分享。在一种场景下，用户的概况设置允许分享她在本地商场的发现，但是仅允许与她在FaceBook社交网络账户上的选定的朋友分享，并且仅在用户特意地保存了该发现(与通常记录所有动作的系统历史存档相反)的情况下才分享。如果该用户在书店发现了有关特定的一本书的信息并保存了该发现，那么该信息被发布到数据存储云中。如果她一周后返回到该商场并且回顾来自先前造访事件的小玩意，那么基于该用户的存储的发现体验，她可能会发现有一个朋友当时正在书店看那本书。该朋友可能已经发布了关于该书的评论，并且可能已经推荐了关于同一主题的另一本书。因此，关于发现的云存档可以与其他人分享，从而发现这些其他人自己的内容并利用该内容得到扩充。

[0797] 类似地，用户可以同意使该用户的发现历史的一部分或全部可供商业实体使用，例如用于受众测量、交通拥挤分析等目的。

[0798] 说明性的操作序列

[0799] 应理解的是，图19的方案可以无用户交互地进行呈现。所显示的操作模式可以是装置的默认操作模式(如屏幕保护程序，在任何无活动时间段之后装置回复到该屏幕保护程序)。

[0800] 在一个特定方案中，当智能电话被拾起时软件被激活。该激活可以通过装置移动或其他传感器事件(例如，视觉刺激变化，或感测到屏幕上的轻拍)来触发。在操作的第一秒左右之后，如果摄像机和麦克风还没有被激活，则激活摄像机和麦克风。智能电话快速地估计当前位置(例如，通过识别本地WiFi节点，或者其他粗略检查)，并且可利用的位置信息被写到黑板上以供其他处理使用。一旦有某个位置信息可用，就在屏幕上呈现相应的地图数据(如果智能电话与地图中心所对应的位置之间的距离没有超过所存储的阈值诸如100码或一英里，那么缓存的一帧地图数据可能就足够了)。智能电话也建立去往云服务的连接并传送该智能电话的位置。用户的概况信息(任选地连同最近的历史数据一起)被调用。

[0801] 在激活后的一秒到三秒之间，装置开始处理关于环境的数据。启动图像和/或音频场景分割。所捕获的影像中记录的特征可以通过屏幕上显示的原型小玩意来表示(例如，这里是影像中的可能值得注意的明亮区域；在这里，这可能也是值得观看的……)。与感测到的数据相关的关键字向量可以开始以流形式传送到云处理。更细化的地理位置可以得到

确定，并且可以获得 / 呈现更新后的地图数据。与先前的发现体验相对应的推针可以绘在地图上。也可以呈现其他图形叠盖物诸如示出用户朋友的位置的图标。如果用户在市中心区或在商场，那么另一叠盖物可以示出正提供待售商品的商店或商店内的位置。(该叠盖物可以基于选择性加入的方式而提供给例如零售商的频繁购物者俱乐部的成员。RSS 型分发可以将这种预订信息馈送给智能电话以供叠盖呈现。) 另一个叠盖物可以示出附近道路上的当前交通状况等。

[0802] 在视觉场景内可能已经识别出感兴趣的显著特征(例如条型码) 并在摄像机视图中将其加亮显示或绘出轮廓。快速图像分割操作的结果(例如，那是脸部) 可以类似地被标出(例如通过绘出矩形轮廓)。装置侧识别操作的结果可以显现出来，例如显现为传感器长方格 104 上的小玩意。在小玩意 UI 可被轻拍并且将呈现相关信息的意义上，激活小玩意 UI。小玩意可以被类似地拖动从而跨过屏幕，以便指示期望的操作。

[0803] 仍然，用户没有对智能电话采取任何动作(除了例如将智能电话从口袋或钱包中举起)。

[0804] 如果智能电话处于视觉发现模式，那么对象识别数据可以开始显现在传感器长方格上(例如，在本地或来自云)。智能电话可能会识别出例如一盒 Tide 清洁剂并叠盖相应品牌的小玩意。

[0805] 用户可以将 Tide 小玩意拖动到屏幕的不同角落，以指示不同的动作。一个角落可以具有垃圾桶图标。另一个角落可以具有保存图标。把 Tide 小玩意拖到那里可将其添加到历史数据存储库，使得它可以在之后被调用和回顾以继续进行该发现。

[0806] 如果用户轻拍 Tide 小玩意，那么任何其他小玩意可以在屏幕上变灰。智能电话将资源分流给进一步分析由所选小玩意指示的对象的处理，从而将轻拍理解为用户对兴趣 / 意图的表达。

[0807] 轻拍小玩意也可以为该小玩意唤出一个背景环境菜单。这样的菜单可以来源于本地、或者从云提供。对于 Tide，菜单选项可以包括 :用法说明，用户借以向制造商提供反馈的博客，等等。

[0808] 菜单选项之一可以用来发出用户想要另外的菜单选项的通知。轻拍该选项会指引智能电话获得其他流行度较低的选项并将其呈现给用户。

[0809] 可选地或者附加地，菜单选项之一可以发出用户对对象识别结果不满意的通知。轻拍该选项会指引智能电话(和 / 或云)“搅拌更多原料”以试图作出另外的发现。

[0810] 例如，书店中的用户可以拍摄一本书的绘出 Albert Einstein 的封面套纸的图像。智能电话可以识别该书，并提供诸如书评和购买选项之类的链接。然而，用户的意图可能是获得关于 Einstein 的进一步的信息。告诉电话从原处理路线返回并做另外一些工作，可以导致电话识别 Einstein 的脸部并随后呈现与该人而不是该书相关的一组链接。

[0811] 在一些用户界面中，菜单选项可以取决于它们被轻拍一次还是两次而具有交替的两个含义。对特定菜单选项进行单次轻拍可以表明用户想要显示更多菜单选项。对同一菜单选项进行两次轻拍可以发出用户不满意原来的对象识别结果并且想要其他对象识别结果的通知。双重含义可以在所显示的菜单图例中用文本指示。

[0812] 可选地，在单次轻拍的含义被给定的情况下，用户借以推断两次轻拍的菜单含义的惯例可能会出现。例如，单次轻拍可以指示使用智能电话的本地资源执行所指示的任务

的指令，而两次轻拍会指引同一任务由云资源执行。或者，单次轻拍可以指示仅使用计算机资源执行该指示的任务的指令，而两次轻拍可以指示把任务提交给带有人类辅助的执行处理(诸如通过使用 Amazon 的土耳其机器人(MechanicalTurk)服务)的指令。

[0813] 作为轻拍小玩意这一方案的替代，用户可以通过环绕一个或更多小玩意画圈(使手指沿围绕屏幕上的图形的路线行进)来指示兴趣。该输入形式允许用户指示对一组小玩意的兴趣。

[0814] 这种手势(指示对两个或更多小玩意的兴趣)可以用于触发与简单地分别轻拍两个小玩意不同的动作。例如，把图 24 中的苹果和 NASA 小玩意圈在共同的圆圈内可以指引系统寻找与苹果和 NASA 两者都相关的信息。作为响应，该装置可以提供关于例如 NASA iPhone 应用软件的信息，所述 NASA iPhone 应用软件使得 NASA 图像可供 iPhone 的用户使用。通过分别轻拍苹果和 NASA 标志，这种发现是不会发生的。类似地，把 NASA 标志和滚石乐队标志圈在一起可以触发导致与以下事件有关的维基百科文章被发现的搜索：在旅行者号(Voyager)宇宙飞船上装载的镀金铜唱片上包含有滚石乐队的歌曲(a fiction——由电影 Starman 引入)。

[0815] 图 21A 示出与图 19 稍微不同的发现 UI。视觉发现占据屏幕的大部分，屏幕的底部地带显示感测到的音频信息。尽管在该黑白绘图中并不明显，但是跨过图 21A 屏幕的中心位置的是叠盖的红色小玩意 202，其由风格化的字母“O”构成(使用来自俄勒冈的报纸的头号标题的字体)。在这种情况下，智能电话感测来自俄勒冈的文章的数字水印信号——触发小玩意的显示。

[0816] 点击小玩意会使它以动画方式变换成图 21B 中所示的背景环境感测菜单。在中心处是表示在图 21A 中发现的对象的图形(例如，报纸中的文章)。在左上方是用户借以将该文章或链接发邮件给其他人的菜单项。在右上方是准许该文章被保存在用户存档中的菜单项。

[0817] 在左下方是去往用户可在其上编写与该文章有关的评论的博客的链接。在右下方是去往与该文章相关联的视频的链接。

[0818] 报纸的读者接下来可能会遇到娱乐场的广告。当被智能电话感测到时，小玩意再次显现。轻拍该小玩意会带来另外一组菜单选项(例如，购买表演者的即将来临的音乐会的票、进入比赛、以及对娱乐场大厅进行 360 度沉浸式游览。也提供“保存”选项。在屏幕的中心位置是具有该娱乐场的标志的矩形。

[0819] 观察带有数字水印的药瓶会带来图 22 中所示的又一个背景环境菜单。在中心位置是药丸看起来应该像什么样子的图像，从而允许在服药(例如，来自旅行者混合了几种不同药丸的瓶子中的药)时进行安全检查。药品也通过名称(“Fedratryl”)、浓度(“50mg”)和开药医生(“Leslie Katz”)来标识。一个菜单选项使智能电话呼叫用户的医生(或药剂师)。该选项在用户的电话簿中搜索开药医生的名字并拨打该号码。另一个选项将自动的药房补充请求提交给药房。另一个链接通向呈现关于药品的常见问题、并且包括 FDA 要求的公开信息的网站。另一个选项可以示出以用户的当前位置为中心的地图——用推针标出贮备有 Fedratryl 的药房。竖直地拿着智能电话而不是水平地拿着智能电话，会将视图切换为无标记的增强现实呈现，从而示出贮备有 Fedratryl 的药房的标志，该标志随着智能电话被移动成面向不同的方向而以叠盖在实际视野的影像上的方式出现或消失。(在说明性实施例

中，来自俄勒冈州波特兰的 SpotMetrix 的用于 iPhone 手机的 3DAR 增强现实 SDK 软件被用于增强现实呈现。）也提供“保存”选项。

[0820] 以类似方式，PDF 文档中的水印可以揭示文档特定的菜单选项；Gap 牛仔裤标签上的条形码可以导致保养说明和时尚提示；对图书封面套纸上的插图的识别可以触发包括书评和购买机会在内的菜单选项的显示；并且对脸部的识别可以带来诸如察看这个人的 FaceBook 页面、在 Flickr 上存储注释有名字的照片等之类的选项。类似地，带有水印的电台或电视音频 / 视频可以导致关于所采样的节目的信息的发现，等等。

[0821] 在一些方案中，（例如，零售商店中的）数字标志可以呈现用水印数据隐写地编码的视觉（或音频）内容。例如，商店可以显示为某些牛仔裤做广告的视频呈现。视频可以用多比特有效载荷（例如，传递索引数据，该索引数据可以被用于访问远程服务器上的相应数据库记录中的相关信息）来编码。该相关信息可以包括标识从其解码出视频水印的标志的位置的地理位置坐标数据。该信息可以被返回给用户装置，并且被用于向装置告知其位置。在一些情况下（例如，如果装置处于室内），其他位置数据（例如来自 GPS 卫星）可能无法获得。然而与解码出的水印信息相对应的从远程服务器返回的数据可以提供智能电话借以获得或提供其他基于位置的服务（甚至是与该商店、该水印等无关的那些服务）的信息。例如，在已知装置处于与例如特定购物中心相对应的地理位置的情况下，智能电话可以提供与附近店主相关的优待券或其他信息（例如，通过同一软件应用程序、通过另一软件应用程序、或以其他方式）。

[0822] 图 23 绘出与图像处理相关联的“雷达”用户界面提示。照亮的红色杆 202（图 24A 中所示）从虚拟的支点反复扫过图像。（在所绘出的情况下，该支点在屏幕外。）该扫掠提醒用户注意智能电话的图像处理活动。每次扫掠可以指示对所捕获数据的新的分析。

[0823] 数字水印通常具有在能够检测出水印有效载荷之前必须辨别出的取向。如果拍摄的图像与水印的取向大体对准地得到定向，那么会使检测变得容易。一些水印具有可被快速辨别出以识别出水印取向的取向信号。

[0824] 在图 23B 的屏幕快照中，雷达扫描线 202 使得瞬时的幻影图案苏醒从而显现出来。该图案示出与水印取向对准的网格。看到（诸如图 23B 中绘出的）倾斜网格可以促使用户将智能电话略微重新定向，使得该网格线与屏幕边缘平行，从而帮助水印解码。

[0825] 作为另一种视觉提示（这种提示是与时间有关的），小玩意可以失去其空间停留处，并在一段时间已经逝去之后漂移到屏幕的边缘。最终，它们可以滑出视线（但是在用户的历史文件中仍然可获得）。这样的方案在图 24 中示出。（在其他实施例中，小玩意在空间上与图像特征相关联地停留——仅当相关联的视觉特征移出视线时消失。对于音频（并且任选性地对于图像），小玩意可以随着时间的推移有选择地在适当的地方冒泡。）

[0826] 音频发现可以平行于上面详述的处理。原型小玩意可以立即与检测到的声音相关联，并且当有更多信息可用时被精炼成完整的小玩意。不同类型的音频水印解码和指纹 / 查找可以用于识别歌曲等。语音识别可能正在进行。一些音频可以在本地被快速处理，并在云中经历更彻底的处理。一旦云处理被完成并确认原始的结论，那么由本地处理产生的小玩意就可以呈现不同的外观（例如，粗体、或变得更亮、或采用彩色与单色的对比）。（当通过本地和云处理或者通过备选的识别机制（例如 SIFT 和条型码读取）确认第一识别结果时，对于视觉分析也同样如此。）

[0827] 如前所述，用户可以轻拍小玩意以揭示出相关联的信息和背景环境菜单。当轻拍一个小玩意时，对其他对象的处理被暂停或减少，使得处理可以聚焦于用户指示出兴趣的地方。如果用户轻拍所显示的菜单选项之一，那么装置 UI 会转变成支持所选操作的 UI。

[0828] 对于识别出的歌曲，背景环境菜单可以包括中心长方格，其呈现艺术家名字、音轨名称、发行商、CD 名称、CD 插图、等等。围绕着其周界的可以是各链接，从而例如允许用户在 iTunes 或 Amazon 购买该音乐作品、或者允许用户观看该歌曲的 YouTube 音乐视频。对于讲话音频，轻拍可以打开菜单，该菜单显示讲话者所说的话的转录文本，并且提供诸如发送给朋友、发布到 FaceBook、播放所存储的讲话者讲话的录制版本等之类的选项。

[0829] 由于音频的时间特性，用户界面合乎期望地包括这样的控制，其允许用户访问来自较早时间(对应于该较早时间的小玩意可能已经从屏幕中去除)的信息。一种方法是允许用户往回扫过期望的音频声轨(例如，将波形 120b 向右扫)。该动作会暂停正在进行的波形的显示(尽管全部信息被缓冲)，并且反而从存储的历史中顺序地调出音频和相关联的小玩意。当期望的小玩意以这种方式被恢复到屏幕上时，用户可以轻拍它来获得相应的发现体验。(也可以作为替代而提供用于在时间域中航行的其他装置，例如往复式控制。)

[0830] 为了便于进行这种时间航行，界面可以提供相对时间信息的显示，诸如沿着所调出的波形每 10 或 60 秒出现一次的 tic 代码，或者利用与调出的小玩意相关联的文本时间戳(例如，“2:45 以前”)。

[0831] 软件的用户界面可以包括“之后”按钮等，从而发出用户不打算实时回顾发现信息的通知。例如，音乐会上的用户可以激活该模式，从而确认她的注意力将会集中于其他地方。

[0832] 该控制会向智能电话指示其不需要用发现数据更新显示器，甚至不需要立即处理数据。而是，该装置可以简单地将全部数据转送给云进行处理(不仅包括捕获的音频和图像数据，而且包括 GPS 位置、加速计和陀螺仪信息等)。来自云的结果在被完成时可以存储在用户的历史中。在之后的更方便的时间，用户可以调用所存储的数据并探索注意到的发现(因为这些发现没有在直接的限制下进行处理，所以其细节可能更丰富)。

[0833] 另一用户界面特征可以是“停泊坞”，小玩意被拖动到该停泊坞并且停留在这里以便例如在之后访问(类似于苹果的 OS X 操作系统中的停泊坞)。当小玩意以这种方式被停泊时，保存与该小玩意相关联的全部关键字向量。(可选地，保存与当前会话相关联的全部关键字向量，从而为之后的操作提供更有用的背景环境。)装置偏好可以设定成使得如果小玩意被拖动到停泊坞，那么相关数据(小玩意特定数据或者整个会话)由云处理以辨别出与所指示对象相关的更详细信息。

[0834] 又一界面特征可以是“蛀洞”(或共享图标(SHARE icon))，小玩意可以被拖动到该蛀洞中。这会发布用于与用户的朋友共享的小玩意或相关信息(例如，小玩意相关关键字向量或整个会话数据)。沉积到蛀洞中的小玩意可以在用户朋友的装置上弹出，例如作为地图显示上的独特推钉。如果该朋友正伴随着该用户，那么小玩意可以显现在该朋友的装置的摄像机视图上，作为由该朋友的装置所观察的场景的相应部分上的叠盖物。当然也可以使用其它相关信息的显示。

[0835] MAUI 项目

[0836] 微软研究院在其 TechFest 2010event 中公布了 Mobile AssistanceUsing

Infrastructure 项目或 MAUI。

[0837] MAUI 研究人员 Cuervo 等人的文章“MAUI :Making SmartphonesLast Longer With Code Offload”(ACM MobiSys ‘10)的摘要介绍 MAUI 项目如下 :

[0838] 本文提出 MAUI,一种能够把移动代码以粒度精细、能量感知的方式卸载到基础设施的系统。以前对这些问题的解决方法或者严重地依赖程序员的支持来分割应用程序,或者这些解决方法在粒度方面较粗糙从而需要整个处理(或整个 VM)都被转移。MAUI 利用受管理的代码环境的益处来提供两个领域的最优特性 :它支持精细粒度的代码卸载以便在对程序员造成最小负担的情况下使能量节约最大化。在能够实现移动装置的当前连接性拘束下可能的最佳能量节约的最优化引擎的驱动下,MAUI 在运行时决定应该在远程执行哪些方法。在我们的评估中,我们展示出 MAUI 能够使下述得到实现 :1)资源密集型面部识别应用程序,其仅消耗小一个数量级的能量 ;2)对等待时间敏感的拱廊游戏应用程序,其使刷新速率加倍 ;以及 3)基于语音的语言翻译应用程序,其通过在远程执行不受支持的组件来绕过智能电话环境的限制。

[0839] MAUI 研究人员(包括来自 Duke、Carnegie Mellon、AT&T 研究院和 Lancaster 大学的个人)所提到的原理和概念重申了本申请人在当前和先前的文献中提出的许多原理和概念。例如,他们的工作是由于观察到这样的事实而被激发的 :电池约束是使用智能电话时的基本限制——这一事实在本申请人的文献中被反复提到。他们建议把与认知相关的应用程序分解为可在智能电话上运行或者可提交给云资源执行的子任务,本申请人也建议这样。他们还建议这种把不同的任务分配给不同的处理器的分配过程可以取决于动态环境(例如电池寿命、连接性、等等),这再次重申了本申请人的观点。这些研究人员还主张依赖附近的处理中心(“云块(cloudlet)”)来获得最少的等待时间,这正如本申请人出于该原因而建议使用无线网络边缘上的飞蜂窝处理节点那样(参见 2009 年 7 月 16 日提交的申请 61/226,195 和已公开的申请 WO2010022185)。

[0840] 考虑到 MAUI 项目和本申请人当前和先前的文献之间的许多共同的目标和原理,请读者参考 MAUI 文献以获得可结合到本申请人详述的方案中的特征和细节。类似地,来自本申请人的文献的特征和细节也可以结合到 MAUI 研究人员建议的方案中。通过这样的结合,会对各自的方案增加益处。

[0841] 例如,MAUI 采用 Microsoft.NET 公共语言运行时间(CLR),由此代码被写入一次,随后就可以在本地处理器(例如,ARM CPU)或远程处理器(通常为 x86 CPU)上运行。在该方案中,软件开发者对一应用程序的哪些方法可被卸载给远程执行做注释。在运行时,解算器模块基于以下因素来分析各方法应该在远程执行还是在本地执行 :(1)能量消耗特性 ;(2)程序特性(例如,运行时间和资源需求);和(3)网络特性(例如,带宽、等待时间和分组丢失)。特别地,解算器模块构造并解算代码卸载问题的线性规划公式,以便找到在等待时间约束的限制下使能量消耗最小化的最佳分割策略。

[0842] 类似地,MAUI 研究人员详述了可以有利地与申请人的工作相结合地使用的特定的云块架构和虚拟机综合技术。这些研究人员还详述了把云块在每次使用后恢复到其原始软件状态的瞬时专用化方法——把瞬时的客户软件环境从云块基础设施的永久主机软件环境中封装出来,并定义两者之间的稳定的普遍存在的接口。这些和其他 MAUI 技术可以在本申请人的技术的实施例中直接采用。

[0843] 在 Satyanarayanan 等人的文章“The Case for VM-based Cloudlets inMobile Computing”(IEEE Pervasive Computing,Vol.8,No.4,pp 14-23,Nov,2009)(其在通过引用结合在本文中的文献 61/318,217 中被附为附录 A,并且可在该申请公开后供公共查阅)中可以找到关于 MAUI 的额外信息。在 2010 年 3 月 4 日发布到网上的标题为“An EngagingDiscussion”的文章(其被作为附录 B 附到申请 61/318,217 中)中可以找到另外的信息。假定本领域技术人员熟悉这些先前的文献。

[0844] 关于声源定位的更多说明

[0845] 由于智能电话变得无处不在,所以它们可以以新颖的方式合作。一种合作是执行高级声源定位。

[0846] 如根据现有技术(例如,US20080082326 和 US20050117754)已知的那样,来自空间上分离的多个麦克风的信号可以用于基于感测到的音频信号中的关联特征之间的时间延迟来辨别音频的发出方向。由不同的个人携带的智能电话可以充当空间上分离的多个麦克风。

[0847] 声源定位的先决条件是理解组分音频传感器的位置。GPS 是一种可以使用的定位技术。然而,更精确的技术正在出现,其中的一些技术将在下面提到。通过使用这样的技术,移动电话的相对位置可以被确定到小于一米的准确度内(在一些情况下接近一厘米)。

[0848] 这种定位技术可以用于识别每个合作电话在三个空间维度中的位置。进一步的精炼结果可以来源于获悉电话主体上的传感器的位置和取向、以及获悉电话的取向。前一信息对于每个智能电话而言是特定的,并且可以从本地或远程数据存储库获得。电话中的传感器(如加速计、陀螺仪和磁力计)可以用于提供电话取向信息。最终,可以确定每个麦克风的 6D 姿态。

[0849] 然后智能电话与其他智能电话共享该信息。智能电话可以被编程为对由其麦克风感测到的音频的带时间戳的数字流进行广播。(对应于几个流的数据可以由具有几个麦克风的智能电话广播。)位置信息也可以由每个智能电话广播,或者一个智能电话可以使用下面提到的适合的技术辨别另一个智能电话的位置。广播可以通过短程无线电技术(如蓝牙或 Zigbee 或 802.11)。诸如 Bonjour 之类的服务发现协议可以用于在智能电话之间交换数据,或者也可以使用另一种协议。

[0850] 尽管 MP3 压缩通常被用于音频压缩,但是 MP3 压缩的使用在本环境中不是有利的。MP3 等按照采样窗口将音频表示为串行的多组频率系数。该采样窗口实际上是具有时间不确定性的窗口。该不确定性会限制声源被定位的准确度。为了使特征相关性准确地与时间延迟相关,优选的是使用未压缩的音频或者使用如实地保留时间信息的压缩(例如,无损数据压缩)。

[0851] 在一个实施例中,第一智能电话接收由一个或更多第二智能电话感测到并从所述一个或更多第二智能电话广播的音频数据,并且结合由自己的麦克风感测到的数据来判断声源方向。该确定结果然后可以与其他智能电话共享,使得其他智能电话不需要作出它们自己的确定。声源位置可以被表示为以第一智能电话为起点的指南针方向。合乎期望的是,第一智能电话的位置为其他智能电话所知,使得相对于第一智能电话的声源定位信息可以与其他智能电话的位置相关。

[0852] 在另一方案中,环境内的专用装置用来从附近的传感器采集音频流,作出声源定

位确定，并且将它的发现广播给参与的智能电话。该功能性可以构建到其他基础设施装置(诸如照明控制器、恒温器等)中。

[0853] 在两个维度中确定音频方向对于大多数应用场合都是足够的。然而，如果麦克风(智能电话)在三个维度中间隔开(例如，处于不同的高度)，那么声源方向可以在三个维度中确定。

[0854] 如果传感器间隔开数米而不是数厘米(如许多应用场合中常见的那样，诸如单个智能电话上的多个麦克风)，那么声源不仅可以通过其方向得到定位，而且可以通过其距离得到定位。两个或更多空间上分离的智能电话可以通过利用基于方向信息的三角剖分、并且知晓它们各自的位置，来确定各智能电话到声源的距离。相对于已知智能电话位置的距离和方向允许声源的位置得到确定。如前所述，如果传感器分布在三个维度中，则该位置信息可以在三个维度中得到解析。(同样，这些计算可以由一个智能电话使用来自另一智能电话的数据来执行。所得的信息随后可以被共享。)

[0855] 链接的数据

[0856] 根据本技术的另一方面，(例如，与链接的数据相关的)数据和资源的 Web 2.0 概念与有形对象和 / 或相关关键字向量数据以及相关联的信息一起使用。

[0857] 链接的数据是指由 Tim Berners Lee 爵士发起的用于经由万维网上的可解除引用的 URI 来发布、分享和连接数据的方案。(参看例如 T.B.Lee 的“Linked Data”(www<dot>w3<dot>org/DesignIssues/LinkedData.html。)

[0858] 简要地，URI 被用于识别有形对象和相关联的数据对象。使用 HTTP URI 使得这些对象可以被人们和用户代理查阅和查找(“解除引用”)。当对有形对象解除引用时，可以提供关于该有形对象的有用信息(例如，结构化的元数据)。该有用信息合乎期望地包括去往其他相关 URI 的链接以便改善对其他相关信息和有形对象的发现。

[0859] RDF (资源描述框架)通常被用于表示关于资源的信息。RDF 将资源(例如，有形对象)描述为许多由主语、谓语和宾语构成的三元组。这些三元组有时被称为声明。

[0860] 三元组的主语是标识所描述的资源的 URI。谓语表示主语和宾语之间存在着哪种关系。谓语通常也是 URI——从与特定领域相关的标准化词典中吸取。宾语可以是文字值(例如，名称或形容词)，或者宾语可以是以某种方式与主语相关的另一资源的 URI。

[0861] 不同的知识表示语言可被用于表示与有形对象和相关联的数据相关的本体论。万维网本体论语言(OWL)是一种这样的知识表示语言，并且使用提供与 RDF 纲要的兼容性的语义模型。SPARQL 是供 RDF 表达式使用的查询语言——允许查询由三元组样式以及逻辑“与”、逻辑“或”和任选样式构成。

[0862] 根据本技术的该方面，由移动装置拍摄并产生的数据条目各自被赋予唯一且持久的标识符。这些数据包括基本关键字向量、分割出的形状、识别出的对象、关于这些条目获得的信息、等等。这些数据中的每个数据都被登记到基于云的注册系统中，该注册系统也支持相关的路由功能。(数据对象自身也可以被推送到云中进行长期存储。)关于该数据的相关声明从移动装置被提供给注册系统。因此，本地装置知道的每个数据对象经由云中的数据被例示。

[0863] 用户可以摆动摄像机，从而拍摄图像。通过这样的动作被聚集、处理和 / 或识别的所有对象(和相关数据)被赋予标识符，并且继续存在于云中。一天或一年之后，另一用户可

以对这样的对象做出声明(例如,树是白橡、等等)。即使是在特定时间在特定地点的快速摄像机扫视,也会在长时期内记录在云中。这种基本的基于云的形式的这些内容可以是用于协作的组织构造。

[0864] 数据的命名可以由基于云的系统赋予。(基于云的系统可以将赋予的名称报告回给始发移动装置。)

[0865] 标识移动装置已经知道的数据的信息(例如,上面提到的聚簇 ID 或 UID)可以提供给基于云的注册系统,并且可以记录在云中作为关于该数据的另一声明。

[0866] 由基于云的注册系统保持的数据的部分视图可以包括 :

[0867]

主语	谓语	宾语
TangibleObject#HouseID6789	Has_the_Color	Blue (蓝色)
TangibleObject#HouseID6789	Has_the_Geolocation	45.51N 122.67W
TangibleObject#HouseID6789	Belongs_to_the_Neighborhood	Sellwood
TangibleObject#HouseID6789	Belongs_to_the_City	Portland (波特兰)
TangibleObject#HouseID6789	Belongs_to_the_Zip_Code	97211
TangibleObject#HouseID6789	Belongs_to_the_Owner	JaneA.Doe
TangibleObject#HouseID6789	Is_Physically_Adjacent_To	TangibleObject#HouseID6790
ImageData#94D6BDFA623	Was_Provided_From_Device	iPhone 3Gs DD69886
ImageData#94D6BDFA623	Was_Captured_at_Time	November 30,2009,8:32:16pm
ImageData#94D6BDFA623	Was_Captured_at_Place	45.51N 122.67W
ImageData#94D6BDFA623	Was_Captured_While_Facing	5.3 degree E of N
ImageData#94D6BDFA623	Was_Produced_by_Algorithm	Canny
ImageData#94D6BDFA623	Corresponds_to_Item	Barcode (条形码)
ImageData#94D6BDFA623	Corresponds_to_Item	Soup can (汤罐)

[0868] 因此,在该方面中,移动装置提供的数据允许基于云的注册系统为该移动装置处理的每个数据条目、和 / 或为在该移动装置的摄像机的视场中发现的每个物理对象或特征例示多个软件对象(例如,RDF 三元组)。可以关于每个数据条目和 / 或物理对象或特征做出许多声明(例如,我是 Canny 数据 ;我基于在某个地点和时间拍摄的图像 ;我是从纬度 X、经度 Y 向北方看时可以看到的高度纹理化的蓝色对象,等等。)

[0869] 重要的是，这些属性可以与其他装置所发布的数据链接在一起，从而允许获取并发现仅根据可获得的图像数据和背景环境无法由用户的装置辨别出的新信息。

[0870] 例如，John 的手机可以将形状识别为建筑物，但是不能够辨别出它的街道地址、或者不能够了解到它的租户。然而，Jane 可能在该建筑物中工作。由于她的特定背景环境和历史，她的手机先前关于与建筑物相关的图像数据而提供给注册系统的信息可能在关于该建筑物的信息方面更丰富，包括关于其地址和一些租户的信息。通过地理位置信息和形状信息的相似性，Jane 的手机提供的信息所关于的建筑物可以被识别为很可能是 John 的手机提供的信息所关于的同一建筑物。(新声明可以添加到云注册系统中，明确地将 Jane 的建筑物声明与 John 的建筑物声明相关，并且反之亦然。) 如果 John 的手机已经请求注册系统这样做(并且如果相关的隐私保护措施准许)，那么注册系统可以向 John 的手机发送由 Jane 的手机提供的关于该建筑物的声明。在这里运转的底层机制可以被认为是居间众包，其中所述声明在参与方支持的政策和商业规则框架内生成。

[0871] 具有与位置相关联的一组丰富的声明的所述位置(例如，通过地点确定，并且任选地也通过时间确定) 可以提供新的发现体验。移动装置可以提供诸如 GPS 位置和当前时间之类的简单声明，作为在链接的数据或其他数据储存库内开始搜索或发现体验的进入点。

[0872] 也应注意的是，在云中对声明进行的访问或导航可以受到移动装置上的传感器的影响。例如，仅当 John 处于由 GPS 或其他传感器确定的建筑物的特定邻近范围(例如，10m、30m、100m、300m、等等)内时，John 才可以被准许链接到 Jane 的关于该建筑物的声明。这可以进一步限制到这样的情况 :John 需要静止不动，或者需要以 GPS、加速计 / 陀螺仪或其他传感器所确定的行走速度行进(例如，小于每分钟 100 英尺或 300 英尺)。基于来自移动装置中的传感器的数据的这种限制可以减少不想要的或相关度较低的声明(例如，广告等兜售信息)，并且可以提供对数据的远程或路过式(或飞过式) 挖掘的某种防护。(可以采用各种方案来与 GPS 或其他传感器数据的电子欺骗作斗争。)

[0873] 类似地，仅当两个涉及的当事者共同拥有某种特性(诸如在地理位置、时间、社交网络联系等方面很接近) 时，才可以访问存储在云中的声明(或者才可以做出关于某主题的新声明)。(所述社交网络联系可以通过参考社交网络数据存储库(诸如 Facebook 或 LinkedIn，显示出 John 社交地联系到 Jane，例如作为朋友) 而被展示出来。) 对地理位置和时间的这种利用与社会惯例相似，即当大群的人聚集时，所发生的自发交互会是有价值的，因为存在着该群体的成员具有共同的兴趣、特征等的很大可能性。访问和发布声明的能力以及基于其他人的存在与否来实现新的发现体验遵循该模型。

[0874] 位置是多组图像数据相互关联的常见线索。也可以使用其他信息。

[0875] 考虑大象研究者。(例如，禁猎地中的)已知的大象通常有命名，并且通过面部特征(包括伤痕、皱纹和长牙) 而得以识别。研究者的智能电话可以把大象的面部特征矢量提交给使面部矢量与大象的名称相关联的大学数据库。然而，当这些面部矢量信息被提交给基于云的注册系统时，可能会揭示出更多的信息，例如先前观测的日期和位置、观察过该大象的其他研究者的姓名、等等。再一次，一旦辨别出数据集合之间的对应度，那么该事实就可以通过向注册系统添加另外的声明而得到记录。

[0876] 应认识到的是，关于由移动装置的摄像机、麦克风和其他传感器感测到的刺激的声明的这种基于云的储存库可以迅速包括全球有用信息的许多存储库，尤其是当与其他链

接数据系统(其中的一些被详细记述在 linkeddata<dot>org 中)中的信息相关时。由于存储的声明所表示的理解在某种程度上会反映贡献这样的信息的装置所属于的个体用户的概况和历史,所以该知识库特别丰富。(比较起来,Google 的万维网索引可能都显得小。)

[0877] (在有形对象的识别方面,潜在有用的词典是 AKT (先进知识技术)本体论。作为它的处于顶部的级别,它具有类别“东西(Thing)”,类别“东西”的下方是两个子类:“有形的东西(Tangible-Thing)”和“无形的东西(Intangible-Thing)”。“有形的东西”包括从软件到亚原子颗粒的任何东西,既包括真实的东西也包括假想的东西(例如米老鼠的汽车)。“无形的东西”具有的子类包括“位置”、“地理区域”、“人”、“交通装置”、和“承载有信息的对象”。该词典可以被扩展从而提供预期在本技术中会遇到的对象的标识。

[0878] 增强空间

[0879] 本技术的一个应用是在(真实的或合成的)关于夜空的图像上呈现信息的功能。

[0880] 用户可以将智能电话指向天空中的特定点,并拍摄图像。图像本身可以由于在小型手持成像装置中很难拍摄到星光而不被用于屏幕上的呈现。然而,地理位置、磁力计、加速计和/或陀螺仪数据可以被采样,以指示出用户把摄像机从什么位置指向什么方向。可以参考夜空数据库(诸如(可通过 Google Earth 界面获得的)Google Sky 项目),以获得与天空的该部分相对应的数据。然后,智能电话处理器可以在屏幕上再现该数据(例如直接从 Google 服务再现该数据)。或者智能电话处理器可以在屏幕上的与摄像机指向的天空部分中的星星的位置相对应的位置处叠盖图标、小玩意、或其他图形标记。指示出希腊的(和/或印度的、中国的、等等)星群的线可以在屏幕上绘出。

[0881] 尽管星星本身在摄像机拍摄的图像中可能并不可见,但是其它本地特征可能是明显的(树、房屋、等等)。星星和星群数据(图标、线、名称)可以显示在该实际图像之上——显示出星星相对于可见的周围环境位于哪里。这种应用程序还可以包括移动星星等经过它们的表观弧线的手段,例如采用滑块控制从而允许用户向前和向后改变所显示的观察时间(星星的位置对应于该观察时间)。用户因此可以发现北极星会在这天晚上的特定时间从特定一棵树的后面升起。

[0882] 其他评论

[0883] 尽管本说明书在前面提到了与本受让人的先前的专利申请的关系以及与 MAUI 项目的关系,但是这值得重复。这些材料应该前后一致地被解读并且被结合起来解释。本申请人期望每个公开文献中的特征与其他公开文献中的特征组合。因此,例如,本说明书中所述的方案和细节可以在申请 US12/271,772 和 US12/490,980 以及 MAUI 文献中所述的系统和方法的各变型实现方案中使用,而刚刚提到的文献的方案和细节也可以在本说明书中所述的系统和方法的各变型实现方案中使用。对于其它提到的文献而言,也类似是如此。因此,应理解的是,本申请中公开的方法、元素和概念可以与那些引用的文献中详述的方法、元素和概念组合。尽管在本说明书中已经特别详述了一些组合,但是许多组合由于大量置换和组合的存在以及描述简洁的需要而尚未被详述。然而,根据所提供的教导,所有这样的组合的实现方案对于本领域技术人员而言是直接明了的。

[0884] 尽管已经参考说明性特征和实例描述和举例说明了我们的创造性工作的原理,但应该认识到的是本技术并不局限于此。

[0885] 例如,尽管已经参考诸如智能电话之类的移动装置,但应该认识到的是,该技术也

适用于各式各样的便携式和固定式装置。PDA、组织器、便携式音乐播放器、台式计算机、膝上型计算机、平板计算机、上网本、超便携式计算机、可佩带式计算机、服务器等全都可以利用这里详述的原理。特别预期到的智能电话包括 Apple iPhone 和遵循 Google 的 Android 规范的智能电话(例如,由 HTC 公司为 T-Mobile 制造的 G1 手机,Motorola Droid 手机,和 Google Nexus 手机)。术语“智能电话”(或“手机”)应该被解释为包含所有这样的装置,甚至是严格地讲既不是蜂窝式电话、也不是电话机的那些装置(例如,苹果 iPad 装置)。

[0886] (包括 iPhone 的触摸界面在内的 iPhone 的细节在 Apple 的已公开的专利申请 20080174570 中有提供。)

[0887] 类似地,本技术也可以使用面部佩戴式装置(如增强现实(AR)眼镜)来实现。这样的眼镜包括显示器技术,通过该技术计算机信息能够由用户观看到——或者叠盖在用户前面的景象上,或者遮住该景象。虚拟现实护目镜是这种装置的一个实例。专利文献 7,397,607 和 20050195128 详述了该示例性技术。商业供给包括:Vuzix iWear VR920、Naturalpoint Trackir 5、和由 ezGear 提供的 ezVision X4 Video Glasses。即将出现的备选者是 AR 隐形眼镜。例如专利文献 20090189830 和 Parviz 的“Augmented Reality in a Contact Lens”(IEEE Spectrum,2009 年 9 月)详述了这种技术。一些或全部这样的装置可以例如无线地与(用户等携带的)其他计算装置通信,或者它们可以包括自含式处理能力。同样,它们可以包含根据现有的智能电话和专利文献已知的其他特征,包括电子罗盘、加速计、陀螺仪、摄像机、投影仪、GPS 等。

[0888] 进一步扩展地说,诸如激光测距(LIDAR)之类的特征可以变成智能电话(和相关装置)上的标准,并且可以结合本技术采用。任何其他传感器技术(例如触觉、嗅觉、等等)也同样如此。

[0889] 尽管详述的技术频繁提到小玩意,但是也可以(例如在用户界面方面)采用其他图形图标(不是必须服务于所详述的方案中的小玩意的目的)。

[0890] 本说明书详述了用于限制在用户屏幕上放置的小玩意的各种方案,例如冗长控制、评分方案、等等。在一些实施例中,有帮助的是,提供非可编程的固定限制(例如,三十个小玩意),以便防止基于病毒的拒绝服务攻击使屏幕被小玩意淹没、从而达到使该界面无用的程度。

[0891] 尽管本说明书中所述的小玩意最通常是与图像和音频特征相关联,但是它们也可以服务于其他目的。例如,它们可以向用户指示当前有哪些任务在工作,并提供其他状态信息。

[0892] 应注意的是,本技术的商业实现方案无疑将会采用与本说明说中呈现的用户界面完全不同的用户界面。本文中详述的那些用户界面是为了支持辅助说明相关联的技术(尽管在许多情况下,这些用户界面的原理和特征凭它们本身的资格应被认为是有创造性的)。以同样的方式,所详述的用户交互形式仅是说明性的;商业实现方案无疑将会采用其他用户交互形式。

[0893] 在本公开内容中提到的智能电话和其他计算机装置的设计是本领域技术人员所熟悉的。一般地说,各自包括一个或更多处理器(例如,Intel、AMD 或 ARM 种类的处理器)、一个或更多内存(例如,RAM)、存储器(例如,磁盘或闪存存储器)、用户界面(其可以包括例如键区、TFT LCD 或 OLED 显示屏、触摸或其他手势传感器、摄像机或其他光学传感器、罗盘

传感器、3D 磁力计、3 轴加速计、3 轴陀螺仪、麦克风、等等，以及用于提供图形用户界面的软件指令)、这些元件之间的互连装置(例如，总线)、以及用于与其他装置通信的接口(其可以是无线的(诸如 GSM、CDMA、W-CDMA、CDMA2000、TDMA、EV-DO、HSDPA、WiFi、WiMax、网状网络、Zigbee 和其他 802.15 方案、或蓝牙)，和 / 或有线的(诸如通过以太局域网、T-1 因特网连接、等等))。

[0894] 更一般地，本说明书中详述的处理和系统组件可以被实现为用于计算装置的指令，包括用于各种可编程处理器的通用处理器指令，所述可编程处理器包括微处理器、图形处理单元(GPU，诸如 nVidia TegraAPX 2600)、数字信号处理器(例如，Texas Instruments 的 TMS320 系列器件)、等等。这些指令可以被实现为软件、固件、等等。这些指令也可以被实现到各种形式的处理器电路中，包括可编程逻辑器件、FPGA (例如 Xilinx Virtex 系列器件)、FPOA (例如，PicoChip 品牌装置)、和专用电路——包括数字的、模拟的、和混合模拟 / 数字电路。指令的执行可以在处理器之间分配、和 / 或跨越一个装置内的多个处理器或者跨越装置网络并行地进行。内容信号数据的变换也可以在不同的处理器和存储器装置之间分配。对“处理器”或“模块”(诸如傅里叶变换处理器、或 FFT 模块等)的提及应该被理解为指代的是功能性、而不是需要特定的实现形式。

[0895] 用于实现详述的功能性的软件指令可以根据这里提供的描述由本领域技术人员容易地编写，例如用 C、C++、Visual Basic、Java、Python、Tcl、Perl、Scheme、Ruby 等编写。根据本技术的移动装置可以包括用于执行不同的功能和动作的软件模块。可以采用已知的人工智能系统和技术来做出上面提到的推断、结论和其它确定。

[0896] 通常，每个装置包括提供与硬件资源和通用功能的接口的操作系统软件，并且还包括可被选择性地调用以执行用户期望的特定任务的应用软件。已知的浏览器软件、通信软件和媒体处理软件可以适合于许多这里详述的用途。软件和硬件配置数据 / 指令通常被存储为可跨越网络访问的有形介质(诸如磁盘或光盘、存储卡、ROM、等等)所传递的一个或更多数据结构中的指令。一些实施例可以被实现为嵌入式系统——操作系统软件和应用软件对于用户而言无法区分的专用计算机系统(例如，基本的手机中的情况通常就是这种情况)。本说明书中详述的功能性可以以操作系统软件、应用软件和 / 或嵌入式系统软件来实现。

[0897] 除了存储软件之外，上面提到的各种存储器组件可以被用作用于本技术所利用的各种信息(例如，背景环境信息、表格、阈值、等等)的数据存储库。

[0898] 本技术可以在各种不同的环境中实现。一种环境是 Android (在 Linux 内核上运行的可从 Google 获得的开源操作系统)。Android 应用程序通常用 Java 编写，并且在其自己的虚拟机中运行。

[0899] 作为将应用程序构造为整体式大代码块这一方案的替代，Android 应用程序通常被实现为可根据需要有选择地加载的“活动”和“服务”的集合。在本技术的一个实现方案中，仅加载最基本的活动 / 服务。然后，根据需要来启动其它活动 / 服务。这些活动 / 服务可以相互间发送消息，例如相互唤醒。因此，如果一个活动寻找椭圆形，那么在有前途的椭圆形得到定位的情况下它可以激活面部检测器活动。

[0900] Android 活动和服务(以及 Android 的广播接收器)由传递消息(例如，请求服务，诸如生成特定类型的关键字向量)的“意图对象”激活。通过该构造，代码可以处于睡眠状

态，直到某些条件出现。面部检测器可能会需要椭圆形来启动。它处于空闲状态，直到发现椭圆形，此时它开始进入活动状态。

[0901] 为了在活动和服务之间共享信息(例如，充当先前提到的黑板的角色)，Android 利用“内容提供商”。这些内容提供商用来存储和检索数据，并使得该数据可由所有应用程序使用。

[0902] Android SDK和相关联的文献可从developer<dot>android<dot>com/index.html 获得。

[0903] 本说明书中所述的不同功能性可以在不同的装置上实现。例如，在智能电话与远程服务提供商处的服务器通信的系统中，不同的任务可以专门由一个装置或另一装置执行，或者执行可以在各装置之间分配。从图像中提取条形码或特征值数据只是这些任务中的两个实例。因此，应该理解的是，把一操作描述为由特定装置(例如，智能电话)执行这样的描述不是限制性的而是示例性的；该操作的执行由另一装置(例如，远程服务器或云)完成、或者在各装置之间分享也是可明确预期到的。(此外，多于两个装置可以共同地被采用。例如，服务提供商可以把一些任务(诸如图像搜索、对象分割、和 / 或图像分类)提交给专门用于执行这些任务的服务器。)

[0904] 以同样的方式，把数据描述为存储在特定装置上这样的描述也是示例性的；数据可以存储在任何地方：存储在本地装置中、存储在远程装置中、存储在云中、分布式的、等等。

[0905] 操作不需要专门由可具体识别的硬件执行。而是，一些操作可以向外提交给其他服务(例如，云计算)，这些其他服务通过另外的通常是匿名的系统来完成它们对所述操作的执行。这样的分布式系统可以是大规模的(例如，涉及全球范围的计算资源)，或者是本地的(例如，当便携式装置通过蓝牙通信识别出附近的装置、并且使一个或更多附近装置牵扯到一任务(诸如贡献来自本地地理位置的数据)中时；关于这一点参看 Beros 的专利 7,254,406)。

[0906] 类似地，尽管某些功能已经被详述为由某些模块、代理、处理等执行，但是在其他实现方案中这些功能也可以由其它这样的实体执行，或者以其它方式执行(或者一起被免除)。

[0907] 本说明书有时提到“识别代理”且有时提到“操作”，而另外一些时候提到“功能”且有时提到“应用程序”或“服务”或“模块”或“任务”或“阶段”、等等。在不同的软件开发环境中，这些术语可以具有不同的特定含义。然而，在本说明书中，这些术语通常可以互换地使用。

[0908] 如上所述，许多功能可以通过由多个组分阶段构成的顺序操作来实现。这些功能可以被认为是多阶段(级联)分类器，其中后面的阶段仅考虑前面的阶段已经处理的区域或值。对于这种类型的许多功能，可以存在着阈值或类似的判断，该类似的判断检查来自一个阶段的输出，并且仅在一定标准得到满足时才激活下一阶段。(仅在前一阶段输出的参数具有超过 15,000 的值时才触发的条形码解码器是这种类型的一个实例。)

[0909] 在许多实施例中，由各种组件执行的功能以及这些功能的输入和输出是以标准化的元数据的形式(由例如所述组件)指定或公开的，使得所述功能以及所述输入和输出可以被例如分派处理识别。基于 XML 的 WSDL 标准可以在一些实施例中使用。(参看例如 Web

ServicesDescription Language (WSDL) Version 2.0 Part 1:Core Language, W3C, 2007年6月。) WSDL的被称为WSDL-S的扩展把WSDL扩展成包括语义元素，所述语义元素通过便于服务的组成而提高可重复利用性。(备选的有语意能力的标准是万维网服务本体论语言:OWL-S。)为了与基于云的服务提供商通信，可以利用基于XML的简单对象访问协议(SOAP)——通常作为万维网服务协议栈的基础层。(其他基于服务的技术也是适合的，诸如Jini、公共对象请求代理架构(CORBA)、表象化状态转换(REST)和Microsoft的窗口通信基础(WCF)。)

[0910] 万维网服务的相互配合可以利用万维网服务业务处理执行语言2.0(WS-BPEL 2.0)来完成。编排可以采用W3C的万维网服务编排描述语言(WS-CDL)。JBoss的jBPM产品是适合于供WM-BPEL 2.0和WS-CDL这两者使用的开源平台。Active Endpoints提供了名称为ActiveBPEL的用于WS-BPEL 2.0的开源解决方案;SourceForge上的pi4SOA是WS-CDL的开源实现方案。万维网服务的安全性可以通过使用WS-Security(WSS)通信协议来提供，所述WS-Security通信协议的流行的Java库实现方案是Apache的WSS4J。

[0911] 本技术的某些实现方案利用现有的图像处理功能(软件)库。这些库包括CMVision(来自Carnegie Mellon大学——特别擅长彩色图像分割)、ImageJ(由国家卫生研究院开发的可自由分发的Java例程包;参看例如en<dot>Wikipedia<dot>org/wiki/ImageJ)、和OpenCV(由Intel开发的程序包;参看例如en<dot>Wikipedia<dot>org/wiki/OpenCV，以及Bradski的书“Learning OpenCV”(O'Reilly, 2008))。受好评的商用视觉库程序包包括:Cognex的Vision Pro，以及Matrox Imaging Library。

[0912] 重复操作被采取的刷新速率取决于具体情况，包括计算背景环境(电池容量、其他处理需求、等等)。可以对每个拍摄的帧或者几乎每个拍摄的帧采取一些图像处理操作(例如，检查镜头盖或其他障碍物是否遮蔽了摄像机的视图)。另外一些图像处理操作可以对每三帧中的第三帧、每十帧中的第十帧、每三十帧中的第三十帧、每一百帧中的第一百帧、等等采取。或者这些操作可以通过时间触发，例如在每十秒中的第十秒采取这些操作，每0.5秒、每一整秒、每三秒就执行一次这些操作，等等。或者这些操作可以通过所拍摄的景象中的变化等来触发。不同的操作可以具有不同的刷新速率——使简单的操作被频繁重复，并且使复杂的操作的重复频繁度较低。

[0913] 如前所述，可以将图像数据(或基于图像数据的数据)提交给云进行分析。在一些方案中，这是代替本地装置处理完成的(或者在某些本地装置处理已经完成之后完成的)。然而，有时，这样的数据可以传给云并且同时在云和本地装置中被处理。云处理的成本通常较小，因此主要成本可能只有一个，即带宽。如果有带宽可用，那么即使数据也可以在本地处理，也可能几乎没有原因不把数据发送给云。在一些情况下，本地装置可能会更快地返回结果;在另外一些情况下，云可能会赢得该竞赛。通过同时使用这两者，始终可以向用户提供这两个响应中较快速的一个。(并且，如上所述，如果本地处理陷入困境或者变得没有前途，那么可以提早结束该本地处理。同时，云处理可以继续运行——或许能产生本地装置根本无法提供的结果。)另外，诸如Google之类的云服务提供商可以搜集通过利用基于云的数据处理机会而获得的其它益处，例如了解这样的地理环境的细节，所述云服务提供商的关于所述地理环境的数据存储被相对耗尽(当然，要受到适当的隐私保护)。

[0914] 有时，本地图像处理可以被暂停，并在后来被恢复。一个这样的实例是如果在打电

话或接电话；装置的偏好可以是把它的资源专门用于为电话通话服务。手机也可以具有用户借以明确指引手机暂停图像处理的用户界面控制。在一些这样的情况下，相关数据被转移到云，由云来继续该处理并将结果返回给手机。

[0915] 如果本地图像处理不能产生迅速的令人满意的结果，并且图像的主题继续吸引用户的兴趣（或者如果用户不做相反指示），那么可以将图像提交给云进行更彻底且冗长的分析。书签等可以存储在智能电话上，从而允许用户核对并了解这种进一步分析的结果。或者如果这种进一步的分析达到了可引起行动得以采取的推断，那么可以提醒用户。

[0916] 应理解的是，所详述的技术的操作中所涉及的决策可以以许多不同的方式实现。一种方式是通过评分。提供与用于不同的备选者的相关输入相关联的参数，并且以不同的组合方式（例如根据多项式方程）对这些参数进行组合、加权、并求和。选择具有最大（或最小）分数的备选者，并且基于该备选者来采取行动。在其他方案中，可以采用基于规则的引擎。这样的方案可通过参考所存储的表示条件规则（例如，如果（条件）、那么行动，等等）的数据来实现。也可以采用自适应模型，其中规则例如基于使用情况的历史模式而进化。也可以采用直观推断方法。本领域技术人员将会认识到的是，仍然有另外的决定处理可以适合于特定情况。

[0917] 在许多实施例中可以包括基于位置的技术来获得有利效果。GPS 是一种这样的技术。其他技术依靠通常在各装置之间发生的那种无线电信号（例如，WiFi、蜂窝、广播电视）。专利公开 WO08/073347、US20090213828、US20090233621、US20090313370 和 US20100045531 描述了在给定若干装置的情况下，信号自身和控制这些信号的不完美的数字时钟信号如何形成一个参考系统，从该参考系统中可以抽取均高度准确的时间和位置信息。

[0918] 模板匹配方案可以用在本技术的许多不同的方面中。除了诸如基于某些背景环境数据来辨别可能的用户意图并确定适当的系统响应之类的应用之外，模板匹配还可以被用在诸如识别内容中的特征（例如，影像中的面部）之类的应用中。

[0919] 模板数据可以存储在云中，并通过使用而得到精炼。模板数据可以在若干个用户之间分享。根据本技术的系统可以在决定如何理解输入的数据或者如何鉴于输入的数据而采取行动时查阅多个模板（例如用户的几个朋友的模板）。

[0920] 在内容特征检测的特定应用中，模板可以采取掩模数据的形式，未知的影像在不同的位置与该掩模数据卷积以找到最大输出（有时被称作线性空间滤波）。当然，模板不需要在像素域中操作；所搜寻的特征图案可以在频域中定义，或者可以在对某些变换（例如，缩放、旋转、颜色）不敏感的其他域中定义。或者，可以尝试多个模板，每个模板对应不同的变换、等等。

[0921] 正如模板匹配可以在本技术的许多不同的方面中被使用那样，概率建模的相关科学也同样可以在本技术的许多不同的方面中被使用，例如在基于传感器数据来评估实际的用户背景环境时被使用（例如，眼 / 嘴图案更可能在面部上而不是在树上被发现）、在鉴于背景环境来确定适当的响应时被使用、等等。

[0922] 在某些实施例中，检查所拍摄的影像视彩度（例如，色饱和度）。这可以通过把来自摄像机的红 / 绿 / 蓝信号转换成颜色与亮度被分开表示的另一种表现形式（例如，CIELAB）来实现。在该后一种表现形式中，可以检查该影像以确定图像帧的全部或显著空间区域（例如，超过 40% 或 90%）的颜色是否显著地低（例如，饱和度小于 30% 或 5%）。如果该条件得到

满足，那么系统可以推断出正在察看的可能是印刷品(例如条型码或文本)，并且可以激活裁制成适合于这样的印刷品的识别代理(例如，条形码解码器、光学字符识别处理、等等)。类似地，这种低颜色饱和度的情况可以发出装置不需要应用某些其他识别技术(例如面部识别和水印解码)的通知。

[0923] 对比度是可以类似地应用的另一种图像度量(例如，印刷的文本和条形码是高对比度的)。在这种情况下，超过阈值的对比度度量(例如，RMS 对比度、Weber 对比度、等等)可以触发与条形码和文本相关的代理的激活，并且可以将其他识别代理(例如，面部识别和水印解码)偏置到不激活。

[0924] 相反，如果拍摄的影像在颜色饱和度方面较高或者对比度较低，那么这会把条形码和 OCR 代理偏置到不激活，而是会把面部识别和水印解码代理偏置到激活。

[0925] 因此，在帮助决定应该对拍摄的影像应用什么不同类型的处理时，粗略的图像度量会是有用的判别式或滤波器。

[0926] 根据本说明书来实现各系统的本领域技术人员被假定熟悉所涉及的各种技术。

[0927] 无线电技术的新兴领域被称为“认知无线电”。通过该镜头观察，本技术可能被命名为“认知成像”。通过改编来自认知无线电的描述，认知成像的领域可以被认为是“要点在于，无线成像装置和相关网络在计算方面充分智能地提取成像结构来支持语义提取和计算机间的通信，以检测随用户背景环境而变的用户成像需求并以最适合于那些需求的方式无线地提供成像服务”。

[0928] 尽管本公开内容已经在说明性实施例中详述了动作的特定排序和元素的特定组合，但应认识到的是，其它方法可以对各动作进行重新排序(可能省略一些动作并添加另外一些动作)，并且其它组合可以省略一些元素并增加另外一些元素，等等。

[0929] 尽管是作为完整系统公开的，但是所详述的方案的子组合也是可分别预期到的。

[0930] 在某些实施例中提及因特网。在另外一些实施例中，还可以采用包括专用计算机网络在内的其它网络或者可以采用所述其它网络作为替代。

[0931] 尽管主要是在执行图像拍摄和处理的系统的背景环境中详述，但是相应的方案也同等地适用于获取和处理音频或其他刺激(例如，触摸、气味、动作、方向、温度、湿度、大气压力、痕量化学品、等等)的系统。一些实施例可以对多种不同类型的刺激做出响应。

[0932] 考虑图 18，其示出音频场景分析器的一些方面(来自 Kubota 等人的“Design and Implementation of 3D Auditory Scene Visualizer-TowardsAuditory Awareness With Face Tracking”(10th IEEE Multimedia Symp., pp. 468-476, 2008))。Kubota 系统采用麦克风阵列获取 3D 声音，定位并分离各声音，并通过语音识别技术识别分离出的声音。Java 可视化软件呈现许多显示。图 8 中的第一个框沿着时间线示出来自人们的讲话事件和背景音乐。第二个框示出在所选的时间点，声源相对于麦克风阵列的布置。第三个框允许定向滤波以便去除不想要的声源。第四个框允许选择特定的说话者并对该说话者所说的话进行转录。用户与这些显示的交互是通过面部跟踪实现的，例如靠近屏幕并朝向期望的说话者移动会允许用户选择并过滤该说话者所说的话。

[0933] 在本技术的背景环境中，系统可以利用类似于基于摄像机的系统的空间模型组件的方案来提供 3D 听觉场景的一般可视化。小玩意可以随位置、时间和 / 或类别而变地被放置在识别出的音频源上。用户可以通过与系统进行交互来参与到对音频源进行分割的过程

中——使用户能够把他们想要了解更多信息的那些声音隔离出来。可以提供例如关于背景音乐的信息，从而识别麦克风、定位音频源、按照风格进行分类、等等。现有的基于云的服务(例如，流行的音乐识别服务如来自 Shazam、Gracenote 和 Midomi 的音乐识别服务)可以适合于在这些方案中提供一些音频识别 / 分类。

[0934] 在大学讲座的背景环境中，学生的移动装置可以获取教授的声音和附近学生的一些偶尔在旁边的谈话。由于分心于谈话的有趣细节，学生可能会随时错过一部分讲座。通过将手指扫过手机屏幕，学生在时间上向后倒退大约 15 秒钟(例如，每帧 5 秒钟)，到达显示出各种面部小玩意的屏幕。通过识别与教授相对应的面部小玩意，学生轻拍它，仅根据教授的声音转录出来的文本于是被呈现(和 / 或被可听地呈递) ——允许学生了解已经错过了什么。(为了快速回顾，在教授的讲话的呈递过程中，可以跳过一些内容、或者缩短、暂停所述呈递。缩短可以通过百分比(例如 50%) 来实现，或者可以把每个长于 0.5 秒钟的暂停裁减为 0.5 秒钟来实现缩短。) 或者，学生可以简单地把教授的小玩意扫到屏幕的顶部——存储去往所存储的麦克风音频数据中的该位置的书签，学生于是可以在之后回顾其内容。

[0935] 为了执行声源定位，合乎期望地使用两个或更多麦克风。虽然不是为了该目的，但是 Google 的 Nexus 手持电话、摩托罗拉的 Droid 手持电话和苹果公司的 iPhone 4 配备有两个麦克风。(多个麦克风在主动噪声消除方案中采用。)因此，这些手持电话能够适于通过使用适当的软件、结合第二音频传感器来执行声源定位(以及声源识别)。(各手持电话中的第二音频传感器是微机械 MEM 麦克风。这种装置在手持电话中正变得日益普及。说明性的多麦克风声源定位系统被详述在已公开的 US20080082326 和 US20050117754 中。)

[0936] 关于声源识别的额外信息可以在例如 Martin 的“Sound SourceRecognition:A Theory and Computational Model”(PhD Thesis,MIT,1999 年 6 月)中得到。关于声源定位的额外信息可在例如已公开的 US20040240680 和 US20080181430 中得到。这样的技术在某些实施例中可以与面部识别和 / 或语音识别技术组合使用。

[0937] 关于例如把讲话与音乐和其他音频区分开的额外信息被详述在美国专利 6,424,938 和已公开的 PCT 专利申请 WO08143569 (基于特征提取) 中。

[0938] 尽管详述的实施例被描述为是相对通用的，但是另外一些实施例也可以专门服务于特定目的或知识领域。例如，一种这样的系统可以被裁制成适合于鸟类观察者，具有一套特别地被设计成识别鸟类及其叫声并更新鸟类观测等的众包数据库的图像和声音识别代理。另一种系统可以提供多样化但是专门的功能性的集合。例如，装置可以包括 Digimarc 提供的用于读取印刷的数字水印的识别代理、用于读取条形码的 LinkMe 移动识别代理、用于解码来自包装上的认证标记的 AlpVision 识别代理、用于识别歌曲的 Shazam 或 Gracenote 音乐识别代理、用于识别电视广播的 Nielsen 识别代理、用于识别无线电广播的 Arbitron 识别代理、等等。(关于识别出的媒体内容，这样的系统还可以提供其他功能性，诸如申请 US12/271,772 (公开号为 US20100119208) 和 US12/490,980 中详述的功能性。)

[0939] 详述的技术可以结合从万维网获得的视频数据(诸如从 YouTube<dot>com 获得的用户原创内容(UGC)) 一起使用。通过类似于这里所述的方案的方案，可以辨别出视频的内容，使得适当的广告 / 内容配对能够得以确定，并且可以提供对用户体验的其他增强。特别地，本申请人预期，这里公开的技术可以用于增强并扩展以下文献中详述的与 UGC 相关的系统：已公开的专利申请 20080208849 和 20080228733 (Digimarc)，20080165960

(TagStory),20080162228 (Trivid),20080178302 和 20080059211 (Attributor),20080109369 (Google),20080249961 (Nielsen),以及 20080209502 (MovieLabs)。

[0940] 应认识到的是,对内容信号(例如,图像信号、音频信号、等等)的所详述的处理包括以各种物理形式对这些信号进行变换。图像和视频(通过物理空间传播并描绘物理对象的电磁波的形式)可以使用照相机或其它拍摄设备从物理对象拍摄,或者通过计算装置产生。类似地,通过物理媒介传播的声压波可以使用音频换能器(例如麦克风)来捕获并转换成电子信号(数字或模拟形式)。尽管这些信号典型地以电子和数字形式被处理以实现上面描述的组件和处理,但是它们也可以以其它物理形式(包括电子的、光的、磁的和电磁波形式)被捕获、处理、转移和存储。内容信号在处理期间以各种方式并出于各种目的被变换,从而产生信号和相关信息的各种数据结构表示。继而,存储器中的数据结构信号被变换以便在搜索、分类、读取、写入和检索期间被操作。信号也被变换以便被捕获、转移、存储并经由显示器或音频换能器(例如扬声器)输出。

[0941] 读者将会注意到,当提及相似或相同的组件、处理等时,有时会使用不同的术语。这部分地是由于本技术是随时间发展的,并且该发展过程会涉及好几个人。

[0942] 本说明书中公开的不同实施例内的要素和教导也可以交换和组合。

[0943] 对 FFT 的提及应该被理解为也包括反向 FFT 和相关变换(例如,DFT、DCT、它们各自的反向变换、等等)。

[0944] 已经提到 SIFT,如通过引用结合在本文中的某些文献详述的那样,SIFT 基于缩放不变特征来执行模式匹配操作。SIFT 数据基本上充当借以识别对象的指纹。

[0945] 以类似的方式,发布到黑板(或其他共享数据结构)上的数据也可以充当指纹——包括借以识别图像或景象的表征该图像或景象的显著可见的信息。对于视频序列而言也同样如此,所述视频序列可以产生由关于用户装置正在感测的刺激的数据(既包括时间数据又包括体验数据)的集合构成的黑板。或者这些情况下的黑板数据可以通过对其应用指纹算法,生成借以识别最近捕获的刺激并将其匹配到其他模式的刺激的、总体唯一的一组识别数据,而得到进一步精炼。(毕加索(Picasso)很久以前就预见到时间空间混合的一组图像元素可以提供与一景象相关的认识,由此可以理解其本质。)

[0946] 如上所述,人工智能技术会在本技术的实施例中起到重要的作用。该领域的最近加入者是由 Wolfram Research 提供的 Alpha 产品。Alpha 通过参考所组织的数据的知识库来计算响应于构造出的输入的回答和可视化。从这里详述的方案搜集的信息可以提供给 Wolfram 的 Alpha 产品,以把响应信息提供回给用户。在一些实施例中,用户被牵扯到该信息提交过程中,诸如通过从系统搜集的词语和其他基元构造出一查询,通过从系统编制的不同查询的菜单中选择,等等。在其他方案中,这由系统来处理。附加地或备选地,来自 Alpha 系统的响应信息可以被提供为对其它系统(如 Google)的输入,以进一步识别响应信息。Wolfram 的专利公开 20080066052 和 20080250347 进一步详述了 Alpha 技术的一些方面,现在可作为 iPhone 应用软件获得。

[0947] 另一辅助技术是 Google Voice,其向传统的电话系统提供了大量改进。这样的特征可以与本技术结合使用。

[0948] 例如,由 Google Voice 提供的语音到文本转录服务可以被采用以便使用用户的智能电话中的麦克风从说话者的环境中捕获环境音频,并产生相应的数字数据(例如 ASCII

信息)。系统可以将这样的数据提交给服务(诸如 Google 或 Wolfram Alpha)以获得相关信息,系统可以随后将该相关信息提供回给用户(或者通过屏幕显示,或者通过语音(例如,通过已知的文本到语音系统),或者通过其他方式)。类似地,由 Google Voice 提供的语音识别可以用来向智能电话装置提供对话用户界面,由此这里详述的技术的一些特征可以通过说出的语言而被选择性地调用和控制。

[0949] 在另一方面中,当用户用智能电话装置捕获内容(听觉或视觉内容)并且采用本公开技术的系统返回响应时,响应信息可以从文本转换成语音,并被递送给用户,例如递送给用户在 Google Voice 中的语音邮件账户。用户可以从任何手机或从任何计算机访问该数据储存库。所存储的语音邮件可以以其听得见的形式回顾,或者用户可以选择回顾例如呈现在智能电话或计算机屏幕上的文字对应物作为替代。

[0950] (Google Voice 技术的各方面在专利申请 20080259918 中有详述。)

[0951] 音频信息有时可以帮助理解视觉信息。不同的环境可通过充当关于该环境的线索的不同声音现象来表征。轮胎噪声和发动机声音可以表征车内或路边环境。HVAC 鼓风机的嗡嗡声或键盘声音可以表征办公室环境。鸟叫声和树中的风声可能会表示户外。频带受限的、展缩扩展器处理过的、很少无声的音频可能会暗示附近在播放电视——或许在家里。水波的反复出现的冲击声暗示着位置在海滩。

[0952] 这些音频位置线索在视觉图像处理方面可以起各种作用。例如,这些音频位置线索可以帮助识别视觉环境中的对象。如果是在存在类似办公室的声音的情况下拍摄的,那么描绘看起来像圆柱形对象的图像很可能是咖啡杯或水瓶,而不是树干。海滩音频环境中的略圆的对象可能是轮胎,但更可能是海贝壳。

[0953] 对这种信息的利用可以采取许多形式。一个特定实现方案设法在可以识别出的特定对象与不同的(音频)位置之间建立关联。可以识别出音频位置的有限集合,例如室内或户外、或者海滩 / 汽车 / 办公室 / 家 / 不确定。随后可以对不同的对象赋予表示该对象在这样的环境中被发现的相对可能性的分数(例如,在 0-10 的范围内)。这种歧义消除数据可以保存在数据结构中,诸如保存在因特网(云)上的可公开访问的数据库中。这里是对应于室内 / 户外情况的简单实例 :

[0954]

	室内分数	户外分数
海贝壳	6	8
电话	10	2
轮胎	4	5
树	3	10
水瓶	10	6
…	…	…

[0955] (应注意的是,室内和户外分数不是必须逆相关 ;一些对象可以是可能在两种环境

中被发现的种类。）

[0956] 如果在图像帧中辨别出看上去像是圆柱形的对象，并且根据可用的图像分析不能明确确定该对象是树干还是水瓶，那么可以参考歧义消除数据和关于听觉环境的信息。如果听觉环境具有“户外”的属性（和／或缺乏“室内”的属性），那么检查候选对象“树”和“水瓶”所对应的户外歧义消除分数。“树木”所对应的户外分数是 10；“水瓶”所对应的户外分数是 8，因此确定“树木”的可能性更大。

[0957] 可以利用作为本说明书中其它地方描述的图像分析方案的音频对应方案的技术和分析来执行听觉环境的识别。或者可以使用其他技术。然而，听觉环境的识别结果常常是不确定的。这种不确定性可以作为因素计入歧义消除分数的使用中。

[0958] 在刚刚给出的实例中，从环境捕获的音频可能会具有与室内环境相关联的一些特征和与户外环境相关联的一些特征。因此音频分析可能会得出模糊的结果，例如 60% 的可能性是户外，40% 的可能性是室内。（这些百分比可以相加起来等于 100%，但这不是必需的；在一些情况下，它们的总和可以更多或更少。）这些评估可以用于影响对象歧义消除分数的评估。

[0959] 尽管存在许多这样的方法，但是一种方法是通过简单乘法利用音频环境不确定性对各候选对象的对象歧义消除分数进行加权，诸如通过下面的表格示出的那样：

[0960]

	室内分数*室内概率（40%）	户外分数*户外概率（60%）
树	3 * 0.4 = 1.2	10 * 0.6 = 6
水瓶	10 * 0.4 = 4	6 * 0.6 = 3.6

[0961] 在这种情况下，即使对听觉环境的了解没有高度确定性，歧义消除数据在识别对象的过程中仍很有用。

[0962] 在刚刚给出的实例中，仅进行视觉分析会暗示两个候选识别结果具有相等的概率：该对象可以是树，可以是水瓶。视觉分析常常会为一对象确定几个不同的可能识别结果——一个识别结果比其余识别结果更可能。最可能的识别结果可以被用作最终识别结果。然而，这里提到的概念可以帮助精炼这样的识别结果——有时会导致不同的最终结果。

[0963] 考虑视觉分析得出的结论是，所描绘的对象有 40% 的可能是水瓶并且有 30% 的可能是树（例如，基于在圆柱形状上缺少视觉纹理）。该评估可以与上面提到的计算级联——通过进一步与单独通过视觉分析确定的对象概率相乘：

[0964]

	室内分数*室内概率（40%）*对象概率	户外分数*户外概率（60%）*对象概率
树（30%）	3 * 0.4 * 0.3 = 0.36	10 * 0.6 * 0.3 = 1.8
水瓶（40%）	10 * 0.4 * 0.4 = 1.6	6 * 0.6 * .4 = 1.44

[0965] 在这种情况下，对象可以被识别为树（1.8 是最高分数）——即使单独运用图像分析得出的结论是该形状更可能是水瓶。

[0966] 这些实例有点过分简单化，以便举例说明在起作用的原理；在实际应用中，无疑将会使用更复杂的数学和逻辑运算。

[0967] 尽管这些实例已经简单地示出两个备选对象识别结果，但是在实际的实现方案中，可以类似地执行从许多可能的备选者的范围中识别一种类型的对象的操作。

[0968] 还没有给出关于歧义消除数据的编辑的说明，例如使不同的对象与不同的环境相关联。尽管这可能是大型的任务，但是存在着许多备选的方法。

[0969] 考虑诸如 YouTube 之类的视频内容站点和诸如 Flickr 之类的图像内容站点。服务器可以从这些源下载静止图像文件和视频图像文件，并应用已知的图像分析技术来识别出各静止图像文件和视频图像文件内示出的某些对象——即使许多对象可能尚未识别出来。可以进一步分析各文件以便在视觉上猜测在其中找到这些对象的环境的类型（例如，室内 / 户外；海滩 / 办公室 / 等等）。即使仅有很小百分比的视频 / 图像给出了有用信息（例如，在一个室内视频中识别出床和书桌；在户外照片中识别出花朵，等等）、并且即使一些分析是错误的，但是就合计值来说，以这种方式可以产生统计上有用的信息选集。

[0970] 应注意的是，在刚刚讨论的方案中，环境可以仅仅参考视觉信息来分类。墙壁会指示室内环境；树会指示户外环境，等等。声音可以形成数据挖掘的一部分，但这不是必须的。在其他实施例中，类似的方案可以可替换地（或附加地）采用声音分析以便对内容和环境进行表征。

[0971] YouTube、Flickr 和其他内容站点还包括描述性元数据（例如，关键字、地理位置信息、等等），所述描述性元数据也可以被挖掘以获得关于所描绘的图像的信息，或者帮助识别所描绘的对象（例如，在可能的对象识别结果之间做出决定）。前面引用的文献（包括 PCT/US09/54358（公开号为 WO2010022185））详述了各种这样的方案。

[0972] 音频信息也可以被用于帮助决定（即，在一组例程操作之外）应该采取什么类型的进一步图像处理操作。如果音频暗示办公室环境，那么这可以暗示与文本 OCR 相关的操作可能是相关的。装置因此可以采取这样的操作，而如果在另一音频环境中（例如，户外），那么装置可能不会采取这样的操作。

[0973] 对象与其典型环境之间的额外关联可以通过对百科全书（例如，Wikipedia）和其他文本的自然语言处理来搜集。如其他地方提到的那样，专利 7,383,169 描述了词典和其他大型语言著作如何通过 NLP 技术被处理，从而编译出充当关于这个世界的这种“常识”信息的丰富来源的词汇知识库。通过这样的技术，系统可以使例如主题“蘑菇”与环境“森林”（和 / 或“超级市场”）相关联；使“海星”与“海洋”相关联，等等。另一个资源是 Cyc——一种已经汇编出常识知识的大型本体论和知识库的人工智能项目。（OpenCyc 可在开源软件许可下获得。）

[0974] 对环境歧义消除数据的编辑也可以利用人类的参与。视频和图像可以被呈现给人类观察者进行评估，诸如通过使用 Amazon 的 Mechanical Turk 服务。特别是发展中国家的许多人愿意为了获得支付而提供图像的主观分析（例如识别所描绘的对象和发现这些对象的环境）。

[0975] 可以采用相同的技术来使不同的声音与不同的环境相关联（使青蛙的叫声与池塘相关联；使飞机发动机的声音与机场相关联；等等）。也可以采用诸如由 Google Voice、Dragon Naturally Speaking、ViaVoice 等执行的语音识别（包括 Mechanical Turk）来识别环境或环境属性。（“请将您的座椅靠背和托盘返回其竖直锁定位置…”指示飞机环境。）

[0976] 尽管刚刚详述的特定方案使用音频信息来消除备选对象识别结果的不明确性，但

是在图像分析方面也可以以许多其他不同的方式来使用音频信息。例如，胜于标识出在不同环境中遇到不同对象的可能性评分的数据结构，可以把音频简单地用来选择 SIFT 特征（SIFT 在其它地方讨论）的几个不同的词汇表之一（或者汇编出一个词汇表）。如果音频包含海滩噪声，那么对象词汇表可以只包含在海滩附近发现的对象（海贝壳，而不是订书机）的 SIFT 特征。图像分析系统所要寻找的候选对象的范围因此可以根据音频刺激而得到限制。

[0977] 音频信息因此可以以许多方式被用来帮助图像分析——取决于特定应用场合的要求；上述内容只是一些实例。

[0978] 正如音频刺激可以帮助向图像的分析 / 理解提供信息，视觉刺激也可以帮助向音频的分析 / 理解提供信息。如果摄像机感测到明亮的日光，那么这暗示户外环境，并且对所捕获音频的分析可以因此参考与户外相对应的参考数据库继续进行。如果摄像机感测到具有作为荧光灯照明的特征的色谱的有规则地闪烁的照明，那么可以假定是室内环境。如果图像帧被拍摄成使蓝色横跨顶部并且下面有高度纹理的特征，那么可以假定是户外背景环境。对在这些情况下捕获的音频进行的分析可以利用这样的信息。例如，低水平背景噪声不是 HVAC 鼓风机——它可能是风；大的卡嗒声不是键盘噪声；它更可能是咆哮的松鼠。

[0979] 正如 YouTube 和 Flickr 提供图像信息的来源，在因特网上存在着许多可免费获得的音频信息的来源。再一次，一个来源是 YouTube。也存在着提供声音效果的零售出售物的免费的低保真度的对应物的在线声音效果库（例如，soundeffect<dot>com、sounddog<dot>com、soundsnap<dot>com、等等）。这些声音效果通常以良好组织的分类法来呈现（例如，自然界：海洋：冲浪海鸥和轮船汽笛；天气：雨：城市混凝土上的暴雨；运输：列车：拥挤的列车内部；等等）。可以挖掘描述性文本数据来确定相关联的环境。

[0980] 尽管上面的讨论聚焦于音频和视觉刺激之间的相互作用，但是根据本技术的装置和方法可以对各式各样的刺激和感测数据（温度、位置、磁场、气味、痕量化学品感测、等等）采用这样的原理。

[0981] 关于磁场，应认识到的是，智能电话正日益设置有磁力计，例如用于电子罗盘目的。这些装置相当敏感——因为它们需要对地球的微妙磁场做出响应（例如，30-60 微特拉斯，0.3-0.6 高斯）。调制磁场的发射器可以用于向手机的磁力计发送信号，以便例如传递信息给手机。

[0982] Apple iPhone 3Gs 具有 3 轴霍尔效应磁力计（被理解为由 AsahiKasei 制造），该磁力计使用固态电路产生与所施加的磁场和极性成比例的电压。当前的装置并没有被优化来进行高速数据通信，尽管将来的实现方案可以把这种特征列入优先。尽管如此，可以容易地实现有用的数据速率。不同于音频和视觉输入，手机不需要以特定的方向被定向以便最优化（由于 3D 传感器的）磁性输入的接收。手机也甚至不需要从用户的口袋或钱包中取出。

[0983] 在一个实现方案中，零售店可以具有视觉推销显示器，该显示器包括用随时间变化的信号驱动的隐藏的电磁铁。所述随时间变化的信号用来发送数据给附近手机。数据可以是任何类型的数据。所述显示器可以提供信息给磁力计驱动的智能电话应用程序，该应用程序呈现可由接收者使用的优待券（例如对于所推销的条目优惠一美元）。

[0984] 磁场数据可以简单地向手机提醒通过不同的通信媒介发送的相关信息的可用性。在初步的应用中，磁场数据可以简单地向移动装置发信号以开启指定的输入组件（例如蓝牙、NFC、WiFi、红外线、摄像机、麦克风、等等）。磁场数据也可以提供密钥、通道、或对该媒介

有用的其它信息。

[0985] 在另一方案中，不同的产品（或与不同的产品相关联的安装在货架上的装置）可以发出不同的磁性数据信号。用户通过将智能电话靠近特定产品移动，而从相互竞争的发送中进行选择。由于磁场与距发射器的距离成指数比例地减小，所以手机有可能把最强（最接近）的信号与其它信号区别开来。

[0986] 在又一方案中，安装在货架上的发射器通常不处于活跃状态，但是会响应于感测到用户或用户意图而变得活跃。该发射器可以包括将磁性发射器激活五至十五秒钟的按钮或运动传感器。或者该发射器可以包括响应于照明变化（更亮或更暗）的光电池。用户可以将手机的照亮的屏幕呈现给光电池（或用手遮蔽它），使得磁性发射器开始五秒钟广播。等等。

[0987] 一旦被激活，可以利用磁场来向用户通知如何利用需要被布置或瞄准以便得到使用的其它传感器（例如摄像机、NFC、或麦克风）。对距离固有的方向性和灵敏性使得磁场数据在建立目标的方向和距离（例如，用于把摄像机指向某物并聚焦）的过程中很有用。例如，发射器可以生成坐标系，该坐标系具有处于已知位置（例如原点）的组件，从而为移动装置提供地面实况数据。把该发射器与（通常存在的）移动装置加速计 / 陀螺仪组合，使得能够准确地进行姿势估计。

[0988] 用于从产品读取条形码或其他机器可读数据并且基于此来触发响应的各种应用程序已经可由智能电话使用（并且可从例如 US20010011233、US20010044824、US20020080396、US20020102966、US6311214、US6448979、US6491217 和 US6636249 的专利文献获悉）。相同的方案可以使用磁性感测信息、使用智能电话的磁力计来实现。

[0989] 在其他实施例中，可以在提供微方向方面使用磁场。例如，在商店内，来自发射器的磁性信号可以将微方向传送给移动装置用户，例如“去走廊 7，向上看你的左边寻找产品 X，现在以 Y 美元出售，并且对前三个拍摄该条目（或相关的推销显示器）的照片的人给予 2 美元的额外折扣”。

[0990] 相关应用程序提供去往商店内的特定产品的方向。用户可以键入或讲出期望的产品的名称，该名称使用各种信号发送技术中的任何一种技术被发送给商店计算机。计算机识别商店内的期望产品的位置，并将方向数据公式化以便引导用户。所述方向可以磁性地或以其它方式传送给移动装置。磁性发射器或几个发射器构成的网络帮助引导用户去往期望的产品。

[0991] 例如，处于期望的产品处的发射器可以充当归航信标。每个发射器可以在多个帧或分组中发送数据，每个帧或分组都包含产品标识符。提供给用户的最初方向（例如，向左走以找到走廊 7，然后在半路向右转）也可以提供用户期望的产品所对应的该商店的产品标识符。用户的移动装置可以使用这些标识符来“调谐”到来自期望产品的磁性发射。罗盘或其他这样的用户界面可以帮助用户找到由该方向指示的总体区域内的该产品的精确位置。当用户找到每个期望的产品时，移动装置可以不再调谐到与该产品相对应的发射。

[0992] 商店中的走廊和其他位置可以具有它们自己各自的磁性发射器。提供给用户的方向可以具有由自动导航系统普及的“分路段显示”类型。（也可以在其他实施例中采用这些导航技术。）移动装置可以通过感测来自沿着该路线的各种路点的发射器，来跟踪用户在该方向上的进展，并向用户提示下一步。继而，发射极可以感测移动装置的接近（诸如通过蓝

牙或其他信号发送技术)，并且根据用户和用户的位置来使其发送的数据得到适应。

[0993] 为了服务于多个用户，来自发射器(例如，导航发射器，而不是产品识别发射器)的某些网络的发送可以被时分复用，从而在多个分组或帧中发送数据，每个分组或帧包含指示意图的接收者的标识符。该标识符可以响应于对方向的请求而提供给用户，并且允许用户的装置把意图用于该装置的发送与其他发送区别开来。

[0994] 来自这些发射器的数据也可以被频分复用，例如为一个应用程序发射高频数据信号，并且为另一应用程序发射低频数据信号。

[0995] 可以使用任何已知的方案来调制磁性信号，包括但不限于频移键控、幅移键控、最小移动键控或相移键控、正交调幅、连续相位调制、脉冲位置调制、网格调制、线性调频扩频或直接序列扩频、等等。可以采用不同的前向纠错编码方案(例如，turbo、Reed-Solomon、BCH)来保证准确、鲁棒的数据发送。为了帮助区别来自不同发射器的信号，调制域可以在不同的发射器、或不同种类的发射器之间以类似于不同的无线电台对频谱的共享的方式分割。

[0996] 移动装置可以设置有尤其适合于把装置的磁力计用于这里详述的应用程序的用户界面。该用户界面可以类似于所熟悉的WiFi用户界面——向用户呈现有关可用通道的信息，并且允许用户指定要利用的通道和/或要避免的通道。在上面详述的应用程序中，用户界面可以允许用户指定要调谐到哪些发射器、或者要收听什么样的数据而忽略其它数据。

[0997] 参考触摸屏界面——一种形式的手势界面。可以在本技术的实施例中使用的另一种形式的手势界面是通过感测智能电话的移动来工作的——通过跟踪所拍摄的图像内的特征的移动来工作。关于这种手势界面的进一步的信息在Digimarc的专利6,947,571中有详述。每当用户输入将要被提供给系统时，可以采用手势技术。

[0998] 进一步向前看，也可以采用响应于从用户检测到的面部表情(例如，眨眼等)和/或生物统计信号(例如，脑电波或EEG)的用户界面。这样的方案正日益为人们熟知；一些方案被详细记述在专利文献20010056225、20020077534、20070185697、20080218472和20090214060中。智能电话的摄像机系统(和辅助云资源)可以被用来识别这样的输入，并相应地对操作进行控制。

[0999] 本受让人在内容识别技术(包括数字水印法和基于指纹的技术)方面有广泛的历史。这些技术对某些视觉查询有重要作用。

[1000] 例如，水印法是可用来识别分发网络内的离散的媒体/实体对象的唯一的独立于容器的技术。水印法被广泛地使用：基本上美国所有的电视和无线电都带有数字水印，如同不计其数的歌曲、电影和印刷品那样。

[1001] 水印数据可以充当一种用于计算机的布莱叶盲文——利用关于标记的对象(物理对象或电子对象)的信息来引导计算机。将模式识别技术应用于一图像可能会在较长的等待之后输出这样的假设：图像可能描绘的是鞋子。相反，如果该鞋子承载着数字水印数据，那么在更短的时间内，可以获得更可靠且准确的一组信息，例如该图像描绘的是2009年五月在印尼制造的尺寸为11M、型号为“Zoom Kobe V”的Nike篮球鞋。

[1002] 通过提供对象身份的指示作为对象本身的固有的一部分，数字水印基于对象的身份来极大地促进移动装置与对象间的交互。

[1003] 用于对水印进行编码/解码的技术被详细记述在例如Digimarc的专利6,614,914

和 6,122,403、Nielsen 的专利 6,968,564 和 7,006,555、以及 Arbitron 的专利 5,450,490、5,764,763、6,862,355 和 6,845,360 中。

[1004] Digimarc 具有与本主题相关的各种其他专利申请。参看例如专利公开 20070156726、20080049971 和 20070266252。

[1005] 音频指纹法的实例被详细记述在专利公开 20070250716、20070174059 和 20080300011（Digimarc）、20080276265、20070274537 和 20050232411（Nielsen）、20070124756（Google）、7,516,074（Auditude）、以及 6,990,453 和 7,359,889（二者都属于 Shazam）中。图像 / 视频指纹法的实例被详细记述在专利公开 7,020,304（Digimarc）、7,486,827（Seiko-Epson）、20070253594（Vobile）、0080317278（Thomson）和 20020044659（NEC）中。

[1006] Nokia 在湾区有 Philipp Schloter 建立的研究视觉搜索技术（Pixto）的新成立部门，并且在其“Point & Find”项目的该领域中有持续的研究。该研究工作被详细记述在例如已公开的专利申请 20070106721、20080071749、20080071750、20080071770、20080071988、20080267504、20080267521、20080268876、20080270378、20090083237、20090083275、和 20090094289 中。这些文献中详述的特征和教导适合于与本申请中详述的技术和方案组合，并且反之亦然。

[1007] 为了简明，所描述的技术的不计其数的变型和组合并没有编入本文件的目录中。本申请人认识到并且期望本说明书的各概念可以被组合、替换和互换——在这些概念本身之间，以及在这些概念与根据所引用的现有技术而已知的那些概念之间。此外，应认识到的是，所详述的技术可以与其他当前和即将出现的技术一起被包括在内，从而获得有利效果。

[1008] 为了提供全面的公开而不过渡加长本说明书，本申请人通过引用将上面提到的文献和专利公开结合在本文中。（这些文献的全部内容被结合在本文中，即使在上文中仅是关于这些文献的特定教导而引用这些文献的。）这些参考文献公开的技术和教导可以结合到这里详述的方案中，并且这里详述的技术和教导也可以结合到这些参考文献公开的技术和教导中。

[1009] 应认识到的是，本说明书已经详述了各种各样的有创造性的技术。其不完全的概要包括：

[1010] 在包括以下步骤的方法中使用具有处理器和一个或更多麦克风的便携式用户装置：把与麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与之相对应的识别出的用户讲话数据。然后，通过参考识别出的用户讲话数据，推断一个或更多信号处理操作，或者推断用于信号处理操作的参数，以便与麦克风所接收的音频相关地被应用。

[1011] 在包括以下步骤的方法中使用包括用于接收语音刺激的麦克风并且包括用于感测第二刺激的至少一个其他传感器的另一装置：把与麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块；从语音识别模块接收与动词相对应的识别出的动词数据；通过参考识别出的动词数据，确定哪种传感器刺激类型是用户所感兴趣的；从语音识别模块接收与用户环境中的对象相对应的识别出的名词数据；以及通过参考识别出的名词数据，确定一个或更多信号处理操作或用于信号处理操作的参数，以便与所确定的类型的刺激相关地被应用。

[1012] 在包括以下步骤的方法中使用包括用于接收第一种类型和不同的第二种类型的刺激的至少第一和第二传感器的又一装置：在装置处接收帮助识别用户环境中的用户感兴趣的对象的非触觉用户输入；以及通过参考指示感兴趣的对象的所述输入，配置相关的传感器数据处理系统来提取与该对象相关的信息。

[1013] 在包括以下步骤的方法中使用又一这样的装置：把与一个或更多麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与之相对应的识别出的用户讲话数据；以及通过参考所述识别出的用户讲话数据来建立参数，所述参数至少部分地定义将要与第二种类型的刺激相关地被应用的处理。

[1014] 又一这样的装置包括配置成执行以下动作的处理器：把与一个或更多麦克风所接收的用户讲话相对应的音频数据应用于语音识别模块，并接收与之相对应的识别出的用户讲话数据；通过参考所述识别出的用户讲话数据来建立参数，所述参数至少部分地定义将要与第二种类型的刺激相关地被应用的处理；以及根据所述建立的参数来处理第二种类型的刺激。

[1015] 又一装置包括多个传感器、一处理器、和一存储器，其中存储器包含黑板数据结构，并且所述处理器在以传感器数据作为输入并产生输出的多个识别代理服务的执行过程中被使用。在包括以下步骤的方法中使用这样的装置：取决于(a)一服务在性质上是否是商用的、和/或(b)从与所述服务相关的外部提供商提供的信任标志是否满足标准，向所述服务授予在黑板数据结构中发布、编辑或删除数据的特权。

[1016] 又一装置包括图像和音频传感器、处理器和存储器。存储器存储使装置执行包括以下内容的动作的指令：处理图像数据以产生对象识别数据；处理音频数据以产生识别出的讲话数据；以及与产生识别出的讲话数据的过程中的模糊话语的解析相关地使用对象识别数据。

[1017] 又一这样的装置包括位置和音频传感器、处理器和存储器。存储器存储使装置执行以下动作的指令：通过参考来自位置传感器的数据，来获得关于装置位置的位置描述信息；处理音频数据以产生识别出的讲话数据；以及与产生识别出的讲话数据的过程中的模糊话语的解析相关地使用位置描述信息。

[1018] 本技术的另一创造性方面是包括以下步骤的方法：分析所接收的图像数据以确定视彩度度量或对比度度量；以及在决定多个不同的图像识别处理中的哪一个应该应用于移动电话摄像机所捕获的图像数据、或者多个不同的图像识别处理应该以什么顺序应用于所述图像数据的过程中使用所确定的度量，以便从移动电话向用户呈现根据不同类型的影像得到的信息。

[1019] 又一这样的方法包括：分析所接收的图像数据以确定颜色饱和度度量；将所确定的度量与阈值进行比较；如果所确定的度量低于阈值，则应用来自第一组处理的一个或更多识别处理；以及如果所确定的度量高于阈值，则应用来自与第一组处理不同的第二组处理的一个或更多识别处理。

[1020] 又一这样的方法包括：分析第一组图像数据以计算颜色饱和度度量；将所计算的颜色饱和度度量作为输入应用于基于规则的处理，以确定应该应用多个不同的识别处理中的哪一个或者应该以什么顺序应用多个不同的识别处理；以及将所确定的识别处理应用于一组图像数据。

[1021] 又一创造性方法涉及基于传感器的沿一路线的人力导航，并且包括：确定去往目的地的路线；使用由用户携带的电子装置中的一个或更多传感器来感测用户沿着所确定的路线的进展；以及向用户提供反馈以帮助导航；其中所述反馈包括由滴答声构成的样式，所述样式随着用户朝向目的地前进而变得更频繁。

[1022] 又一这样的方法涉及操作配备有摄像机的处理图像数据的便携式装置，并且包括：执行初始的一组多个不同的图像处理操作；以及无需明确的用户命令，在环境准许的限度内调用额外的图像处理操作；其中所述装置自发地行动从而满足推断出的或预见到的用户需求。

[1023] 又一创造性方法涉及配备有磁性传感器的智能电话，并且包括：感测由零售环境中的多个电磁发射器发射的磁性信号，并且基于磁性信号向用户提供导航或产品信息。

[1024] 又一这样的方法包括：在操作的第一阶段，从用户的环境捕获一图像序列；处理所述序列以识别所述序列中的特征并鉴别相关信息，所述处理至少部分地由用户携带的便携式装置执行；以及在跟随在第一阶段之后的操作的第二阶段中，使用与便携式装置相关联的输出装置将所述相关信息呈现给用户。

[1025] 又一方法限制用户访问关于一物理对象的声明的能力或者限制用户做出关于一物理对象的声明的能力(例如，在链接数据系统中)，除非用户与所述对象或者与先前做出这样的声明的另一用户具有可证明的关系。

[1026] 一种相关的链接数据方法，其特征在于，基于由用户携带的传感器产生的数据检查运动信息，并且在运动信息指示出用户以某种方式(例如，以高于阈值的速度)移动的情况下，限制用户访问与一物理对象相关的声明的能力或者限制用户做出与一物理对象相关的声明的能力。

[1027] 所详述的技术的另一创造性方面是一种处理装置，其包括处理器、存储器、触摸屏、位置确定模块和至少一个音频或图像传感器。存储器存储将处理器配置成在触摸屏上呈现用户界面的指令，所述用户界面的第一部分呈现来自传感器的信息，并且同时，所述用户界面的第二部分呈现与所述装置的位置相关的信息。

[1028] 另一装置包括处理器、存储器、屏幕和图像传感器。存储器存储将处理器配置成在触摸屏上呈现与由图像传感器感测的图像相对应的数据的指令，所述处理器还在触摸屏上呈现雷达扫描线的扫掠效果以指示处理图像数据的过程中的装置活动。

[1029] 这里详述的又一创造性方法是一种进行声源定位的方法，其包括：利用环境内的多个无线电话对环境音频进行采样；将由第一电话感测的音频信息发送给第二电话；辨别使第一电话的位置与第二位置相关的位置数据；以及在第二电话中，处理所述位置数据、从第一电话接收的音频信息、以及由第二电话采样的音频，以辨别相对于第二电话的声源方向。

[1030] 自然地，与上述方法相对应的装置和软件以及与上述装置相对应的方法和软件也是本申请人的创造性工作的一部分。此外，被描述为由便携式装置中的处理器执行的方法也可以由远程服务器执行、或者可以由若干个单元以分布式方式执行。

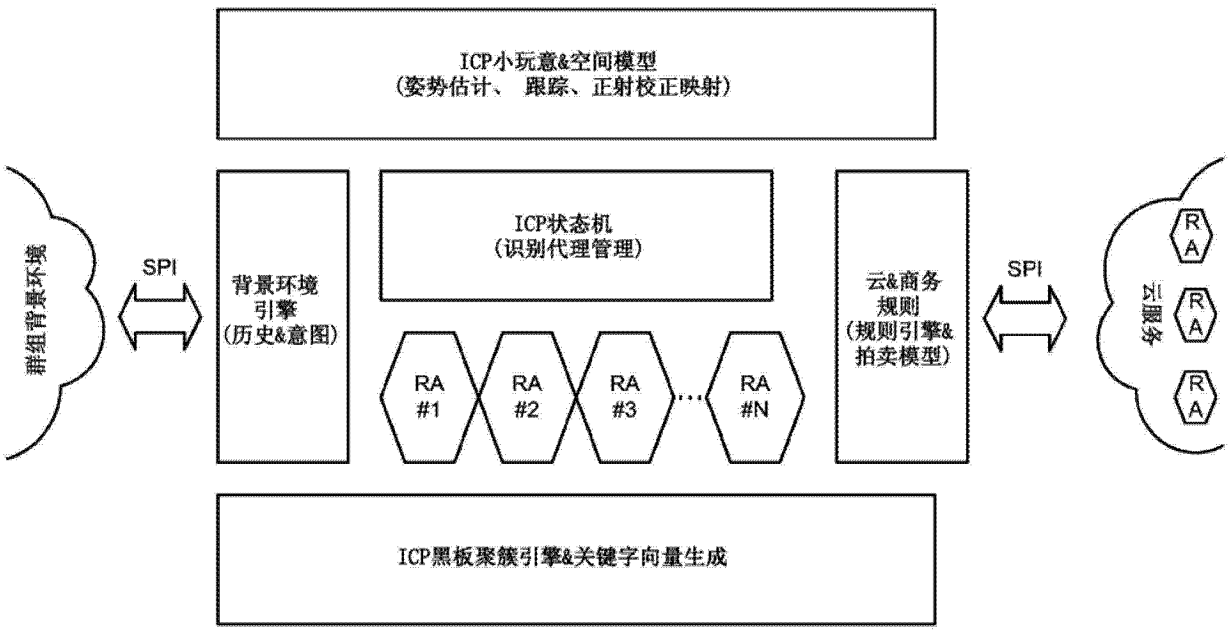

ICP小玩意&空间模型
(姿势估计、 跟踪、正射校正映射)
群组背景环境
SPI
背景环境
引擎
(历史&意图)
ICP状态机
(识别代理管理)
RA
#1
RA
#2
RA
#3
RA
#N
云&商务
规则
(规则引擎&
拍卖模型)
SPI
云服务
RA
RA
RA
ICP黑板聚簇引擎&关键字向量生成

图 1


本地装置
云管道
云

图 2

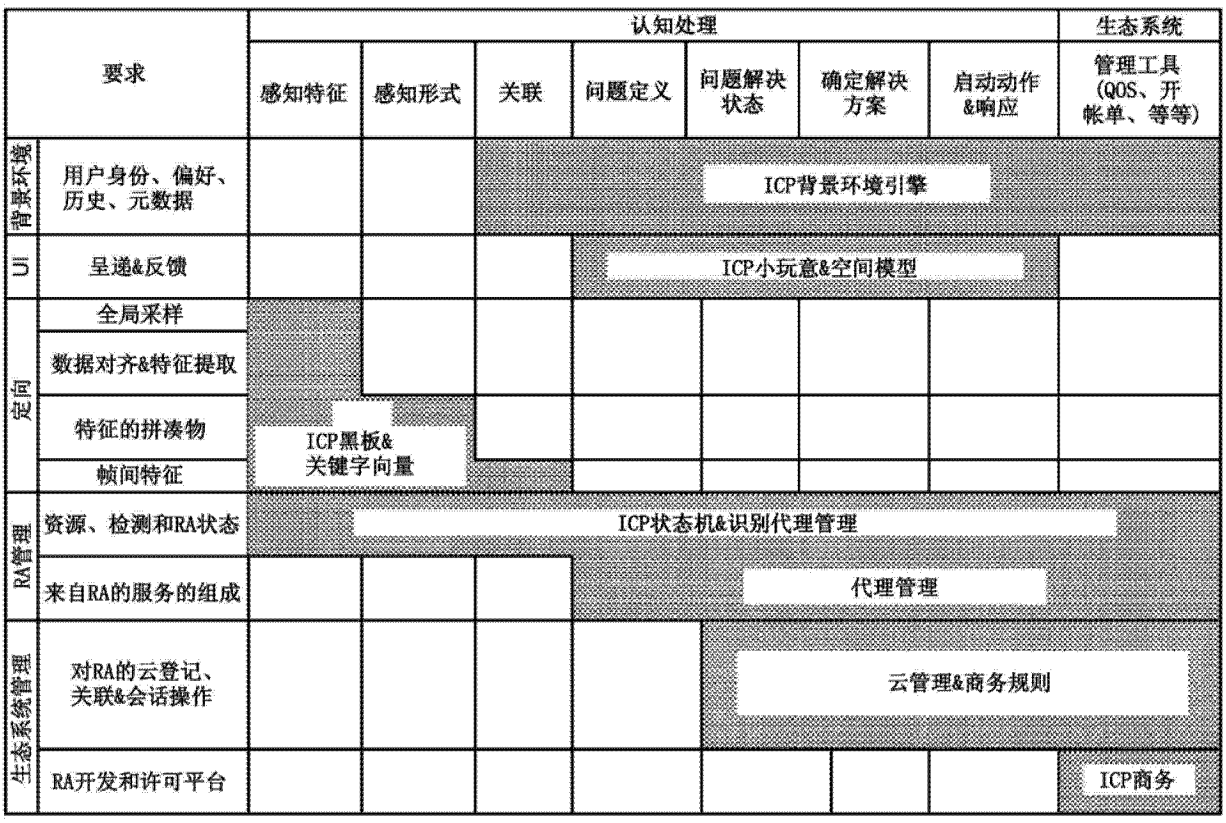

要求
认知处理
生态系统
感知特征
感知形式
关联
问题定义
问题解决状态
确定解决方案
启动动作&响应
管理工具(QOS、开帐单、等等)
背景环境
用户身份、偏好、历史、元数据
ICP背景环境引擎
UI
呈递&反馈
ICP小玩意&空间模型
定向
全局采样
数据对齐&特征提取
特征的拼凑物
帧间特征
ICP黑板&关键字向量
RA管理
资源、检测和RA状态
ICP状态机&识别代理管理
来自RA的服务的组成
代理管理
生态系统管理
对RA的云登记、关联&会话操作
云管理&商务规则
RA开发和许可平台
ICP商务

图 3

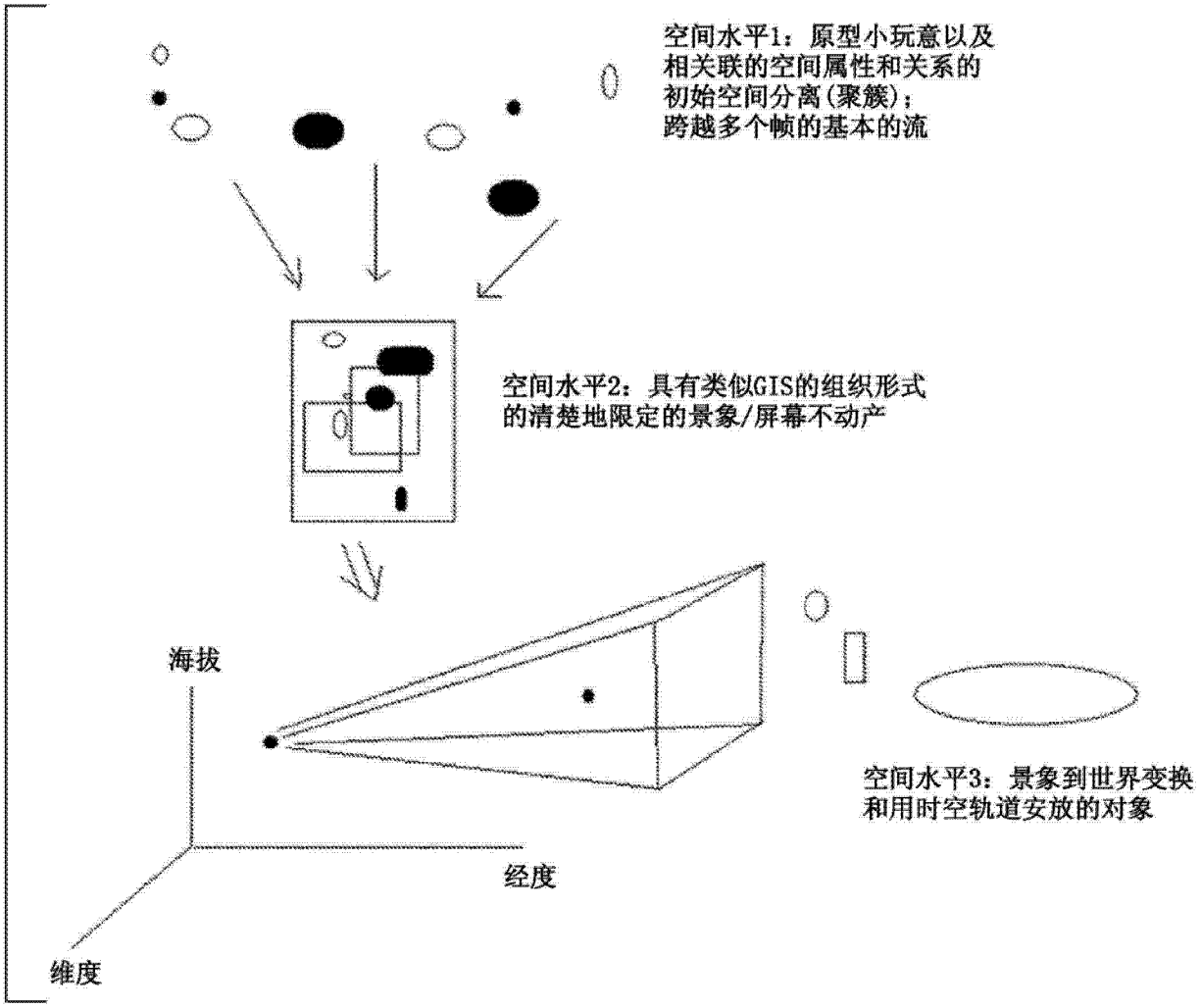

空间水平1：原型小玩意以及
相关联的空间属性和关系的
初始空间分离(聚簇)；
跨越多个帧的基本的流
空间水平2：具有类似GIS的组织形式
的清楚地限定的景象/屏幕不动产
海拔
空间水平3：景象到世界变换
和用时空轨道安放的对象
经度
维度

图 4

	无限对称指数滤波器（高斯）	坎尼（高斯）	马尔-希尔德雷斯（拉普拉斯高斯）	索贝尔（梯度）	克希霍夫（梯度）	拉普拉斯算子（零交叉）
无限对称指数滤波器	X					
坎尼	95	X	90	70	70	80
马尔-希尔德雷斯			X			
索贝尔		90		X		
克希霍夫					X	
拉普拉斯算子						X

图 5

	无限对称指数滤波器	拉普拉斯算子
坎尼	95	80

图 5A

	CPU	内存
无限对称指数滤波器(高斯)	25	15
坎尼(高斯)	22	20
马尔-希尔德雷斯(拉普拉斯高斯)	25	20
索贝尔(梯度)	10	5
克希霍夫(梯度)	12	10
拉普拉斯算子(零交叉)	20	10

图 6

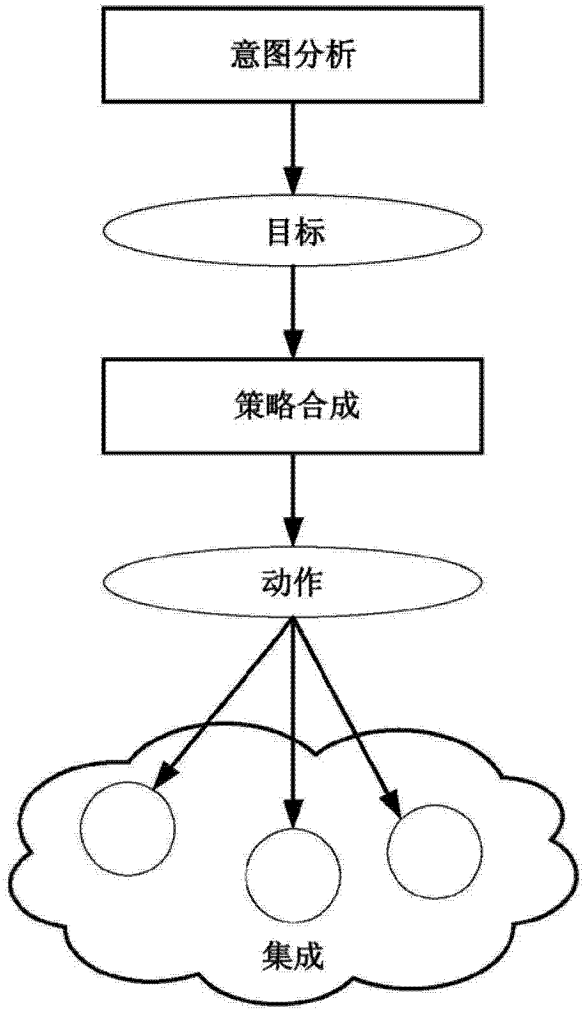


图 7

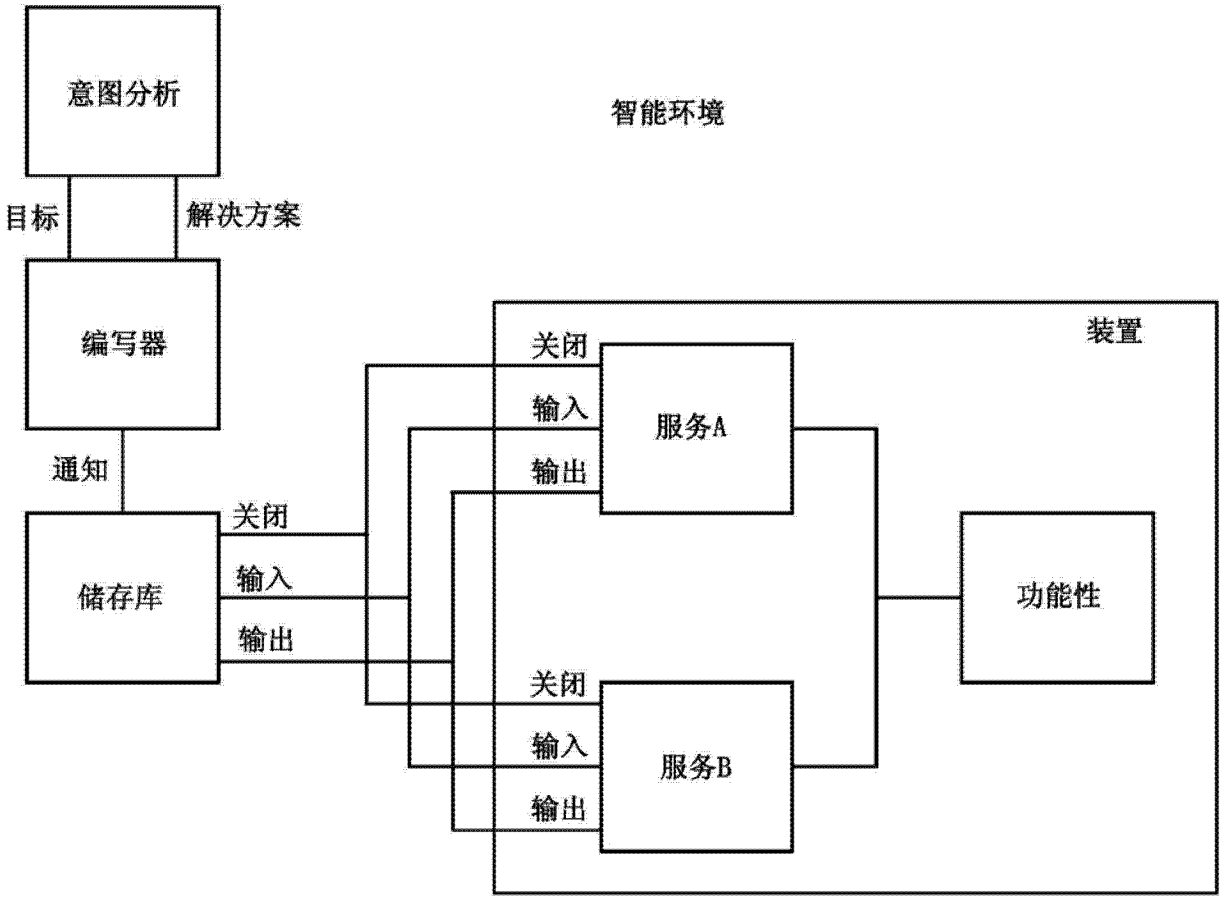

意图分析
智能环境
目标
解决方案
编写器
通知
储存库
关闭
输入
输出
装置
关闭
输入
输出
服务A
关闭
输入
输出
服务B
功能性

图 8

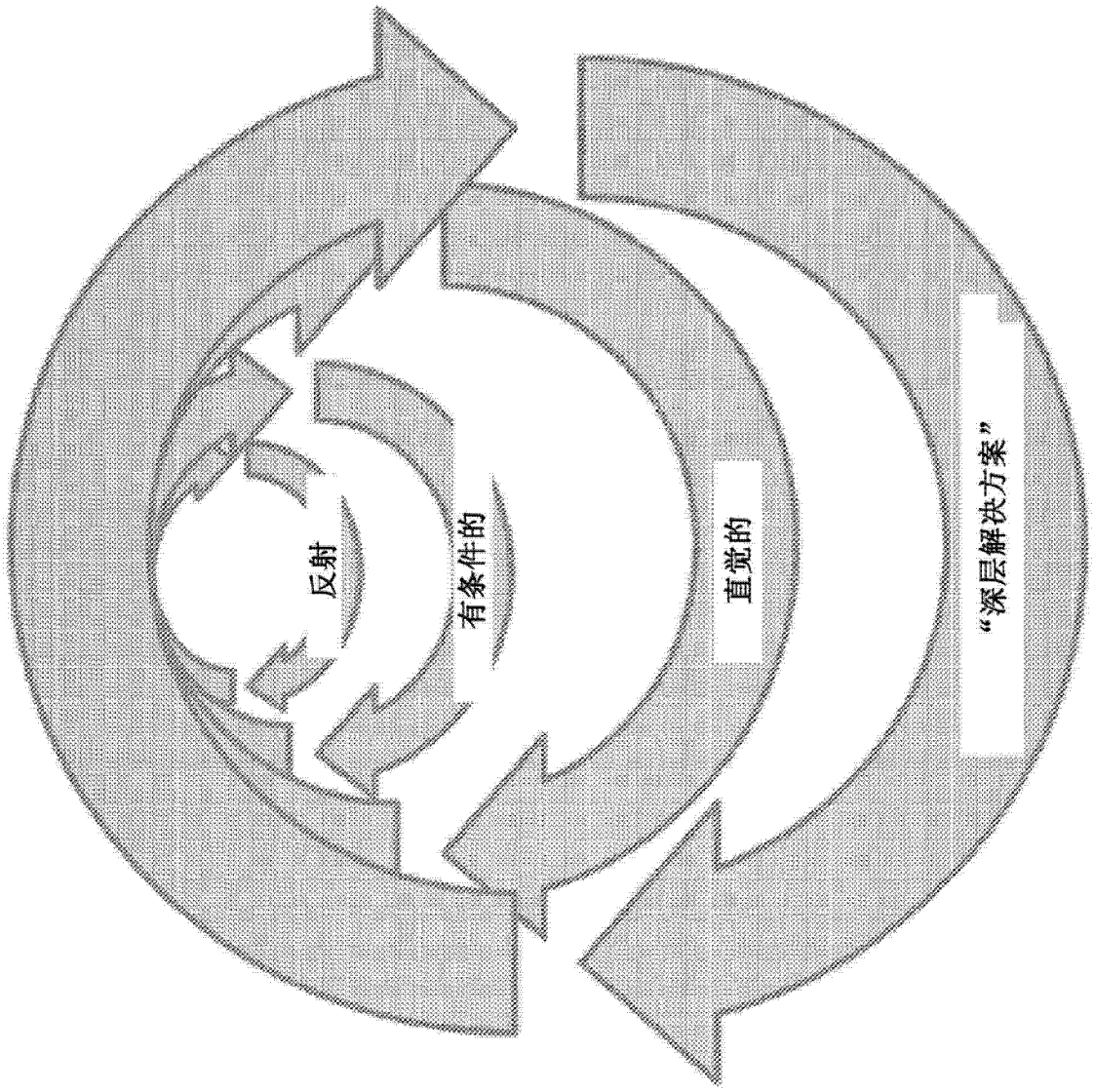


图 9

状态机步调	查询/背景环境	传感器(例如像素)	通信	用户界面
反射 0至0.5秒钟 每0.05秒 刷新一次	背景环境： * 用户ID * 摄像机类型 * 分辨率	镜头盖是否 在上面？ 分割 流 焦点	分组因特网探测 (ping)	2D映射 模拟小玩意 小玩意确认
有条件的 0.5至1秒钟 每0.5秒 刷新一次	代理清单 装置状态 地理位置 历史 用户反馈 代理间的 互关联	定向解析 为各代理所共有 的基本操作	对代理服务器进行 分组因特网探测 公告进入市场的条目 服务质量	代理小玩意 启动代理会话 冗长
直觉的 1至10秒钟	语义提取	使关键字向量丰富 对各代理调谐(关键 字向量) 支持正射映射	将关键字向量发给 意图的云接收者 (代理服务器) 预期来自云的早期 周转时间 规则引擎	充分交互性 小玩意会话 (菜单，窗口) 第三方库
深层解决方案 5+秒钟	为服务竞争 社交背景环境 * 历史 * 群体 * 等等 关联	基于云反馈优化 关键字向量 为了云的稳定状态 而使关键字向量 统一化 +代理	通过会话进行通信 管理服务质量 规则引擎	充分实现& 货币化的UI

图 10

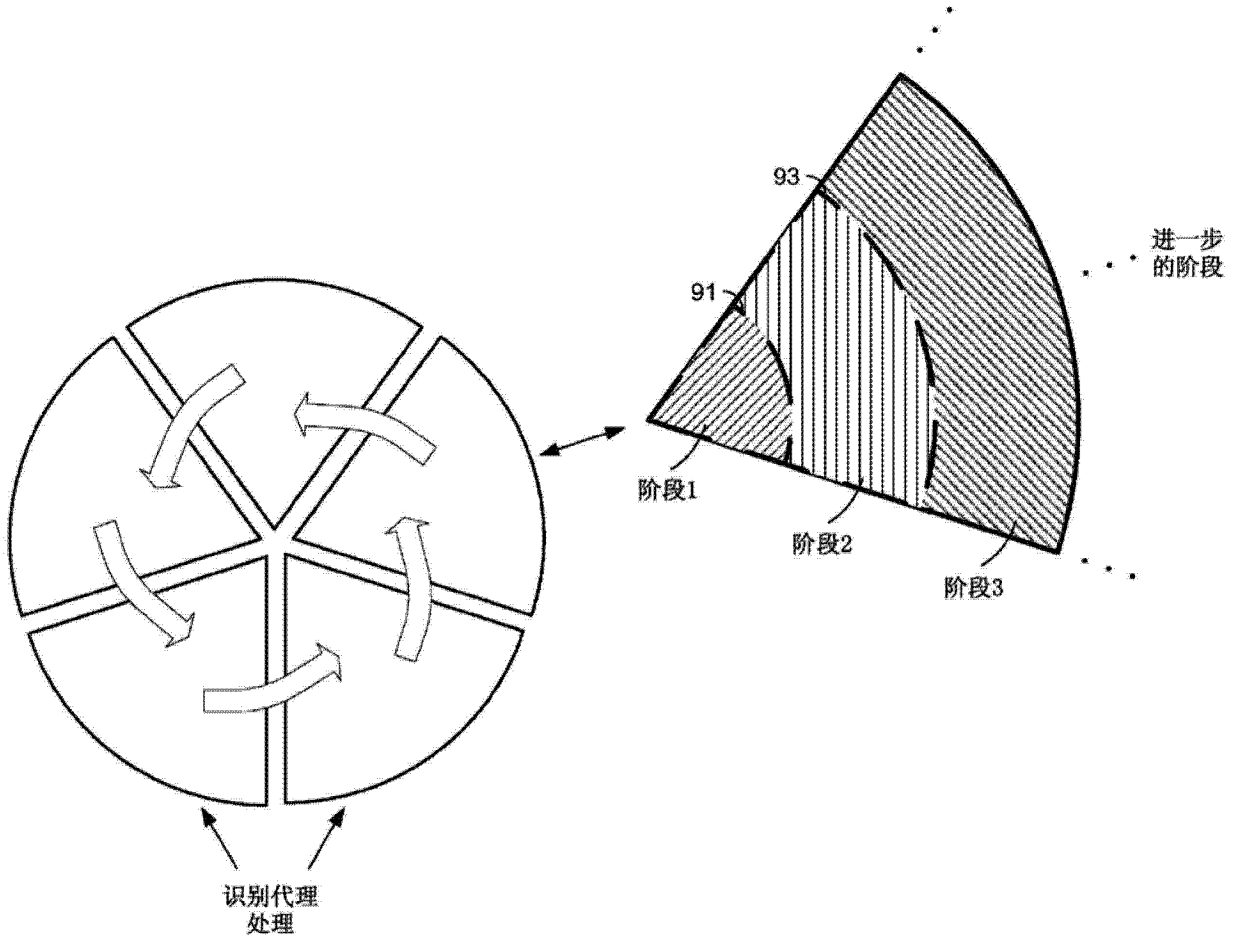


图 12

	意图	历史
个体用户	用户想要做什么？	用户在过去做了什么？
群体	在相似情况下该群体想要做什么？	该群体在过去做了什么？

图 11

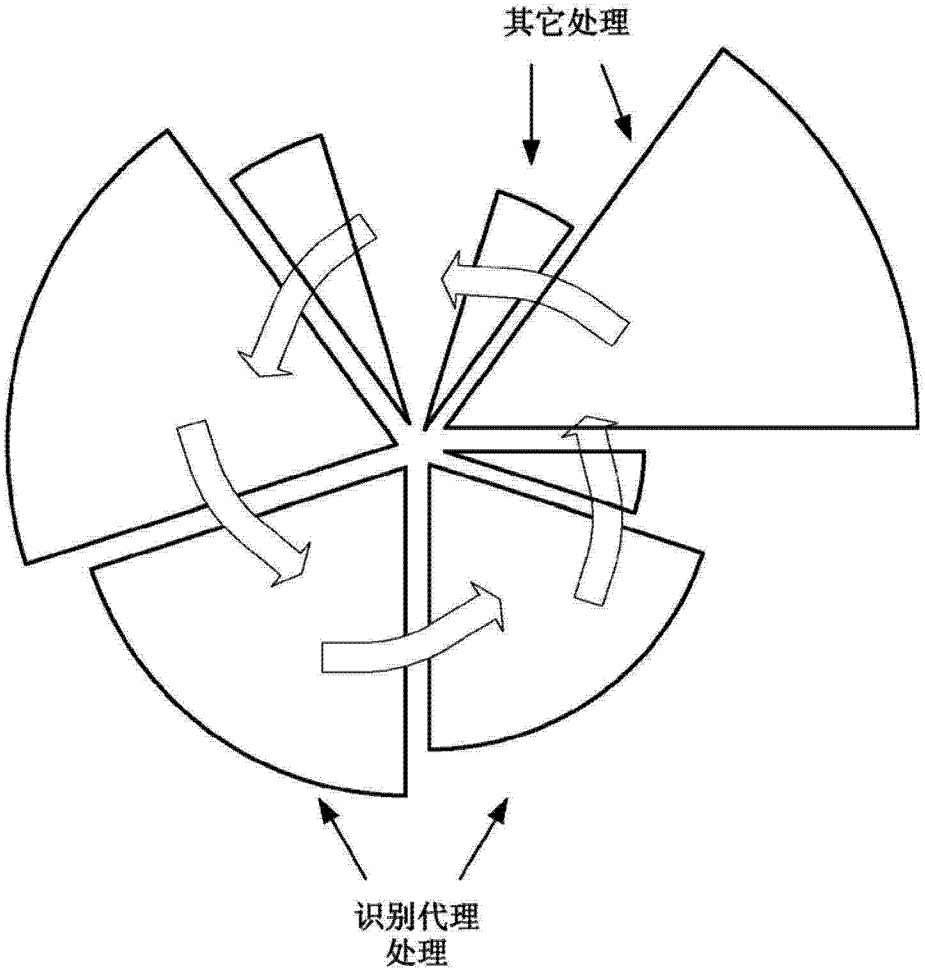


图 13

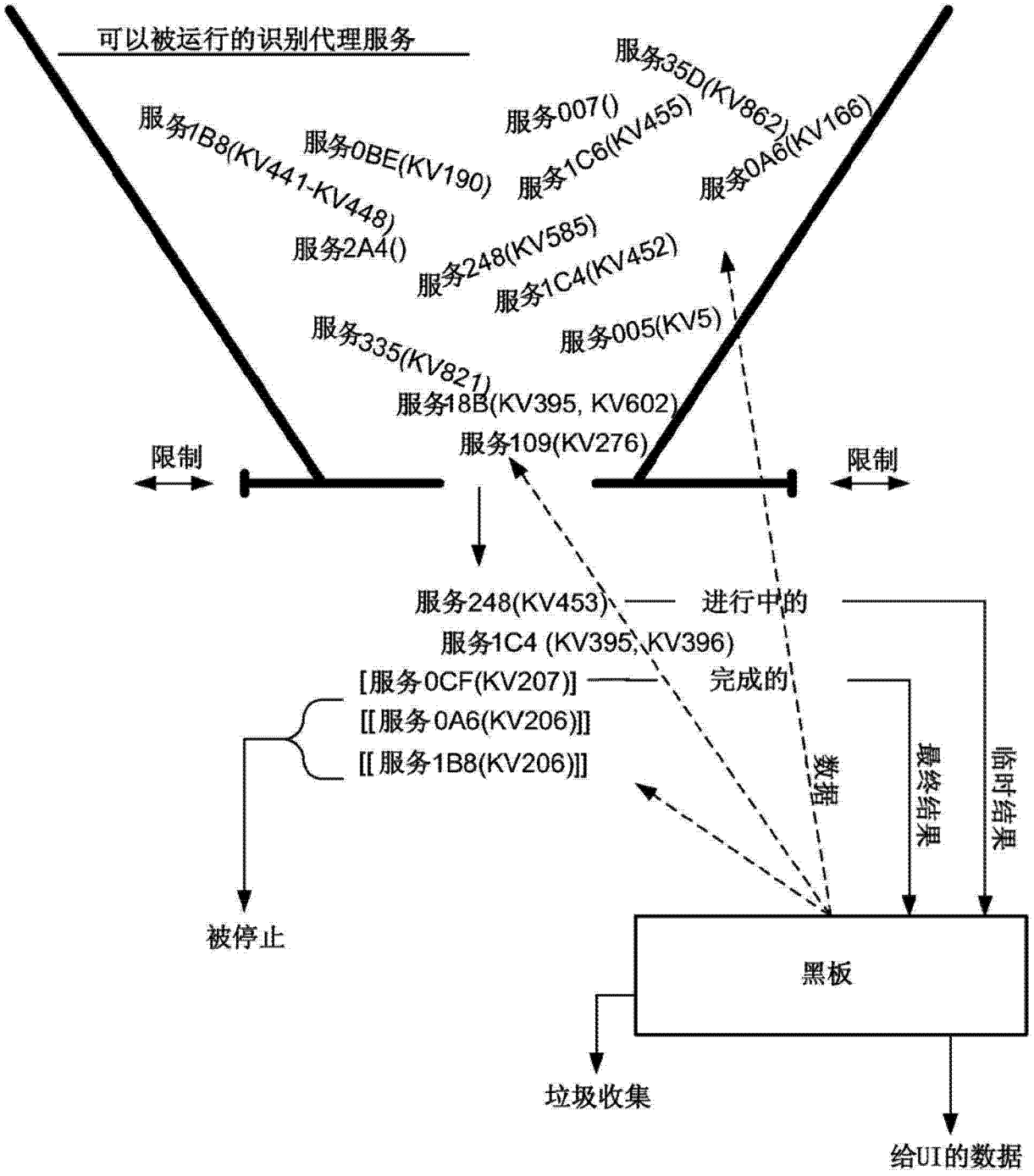

可以被运行的识别代理服务
服务35D(KV862)
服务007()
服务1B8(KV441-KV448)
服务0BE(KV190)
服务1C6(KV455)
服务0A6(KV166)
服务2A4()
服务248(KV585)
服务1C4(KV452)
服务005(KV5)
服务335(KV821)
服务18B(KV395, KV602)
服务109(KV276)
限制
限制
服务248(KV453)
进行中的
服务1C4 (KV395, KV396)
[服务0CF(KV207)]
完成的
[[服务0A6(KV206)]]
[[服务1B8(KV206)]]
数据
最终结果
临时结果
被停止
黑板
垃圾收集
给UI的数据

图 14

识别代理服务的队列

服务	相关度分数	成本分数	净值	条件	偏置
服务19F(KV415)	46	{37,64,15}	-70	...	{100,110,200}
服务2E7(KV755,KV745)	14	{3,99,4}	-92		{150,110,100}
服务1E4(KV485)	12	{12,1,2}	-3		{100,100,100}
服务19F(KV416)	11	{57,12,6}	-64	...	{100,110,120}
服务2D1(KV722)	8	{24,1,1}	-18		{105,100,200}
服务191()	6	{8,1,2}	-5		{100,100,110}
服务32F(KV815)	4	{4,3,1}	-4	... ...	{100,110,105}
服务154(KV416)	3	{2,1,1}	-1		{100,107,100}
⋮	⋮	⋮	⋮	⋮	⋮

图 15

资源跟踪

	CPU	GPU	功率	带宽	内存
最大	100	200	50	250	300
被使用的	93	80	46	100	250
可用的	7	120	4	150	50

图 16

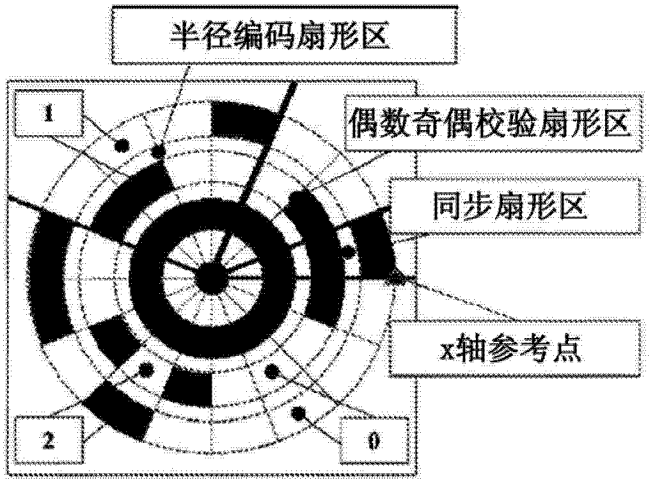


图 17

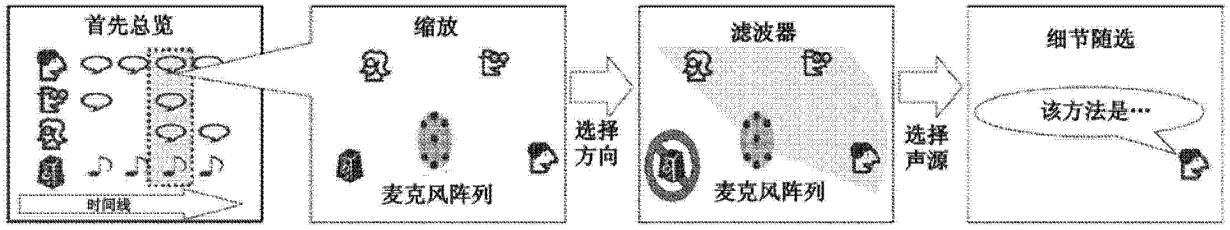


图 18

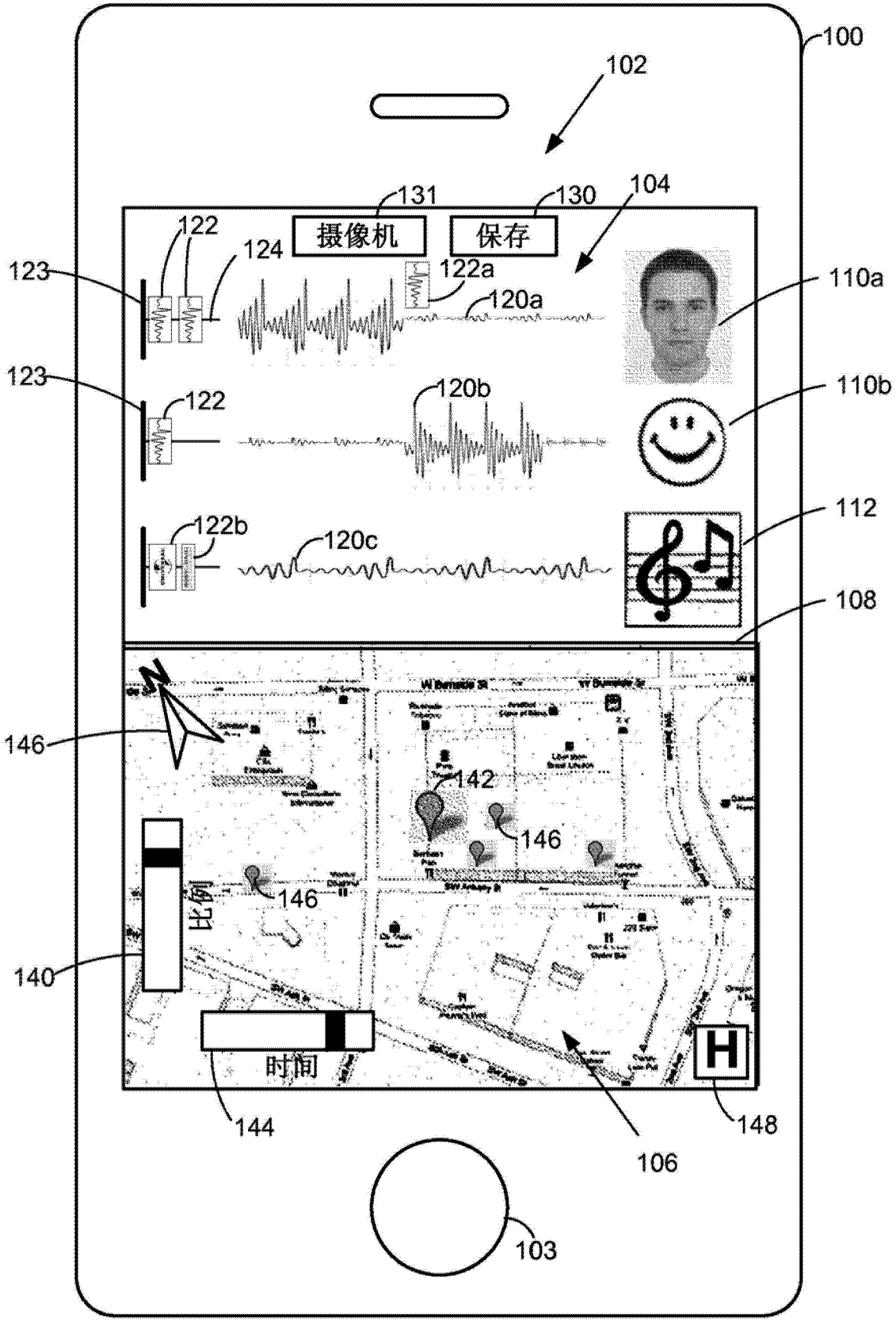

100
102
104
131
130
摄像机
保存
122
124
123
122a
120a
110a
123
122
120b
110b
122b
120c
112
108
146
142
146
146
比例
140
时间
144
106
148
H
103

图 19

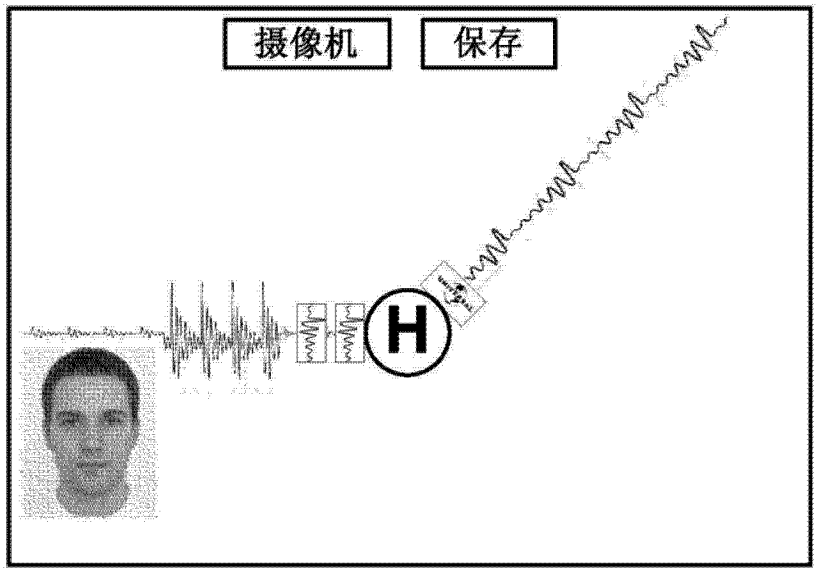


图 19A

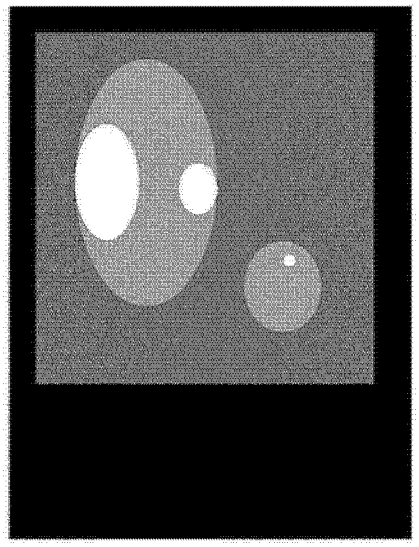


图 20A

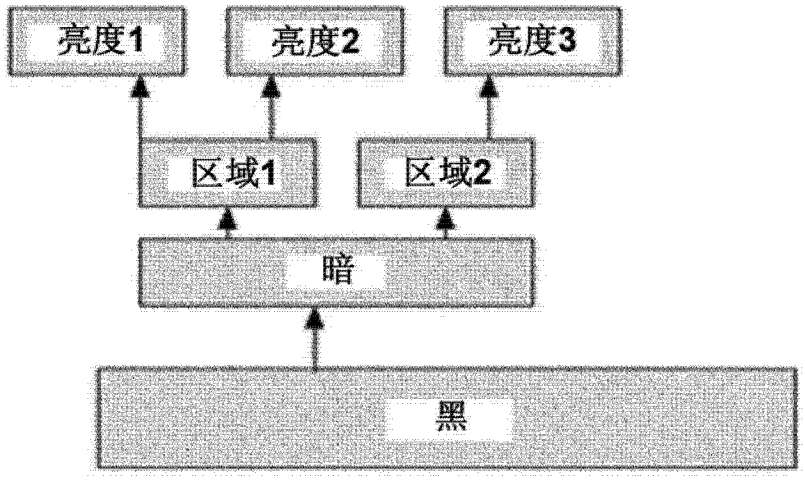


图 20B

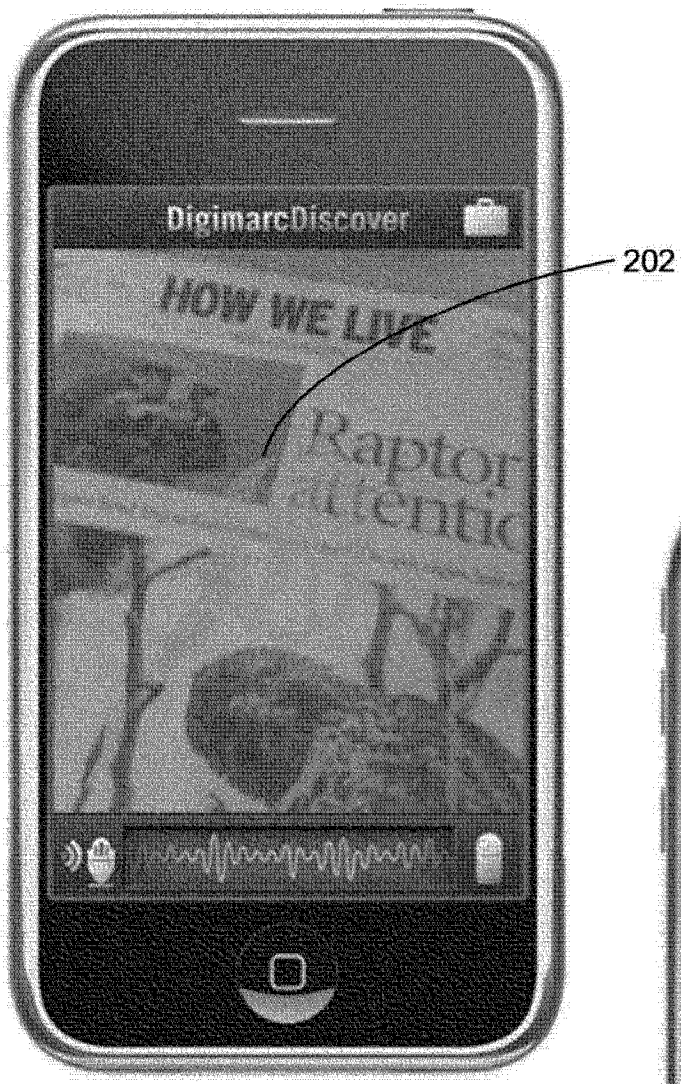

DigimarcDiscover
HOW WE LIVE
202

图21A


DigimarcDiscover
The Oregonian

图21B

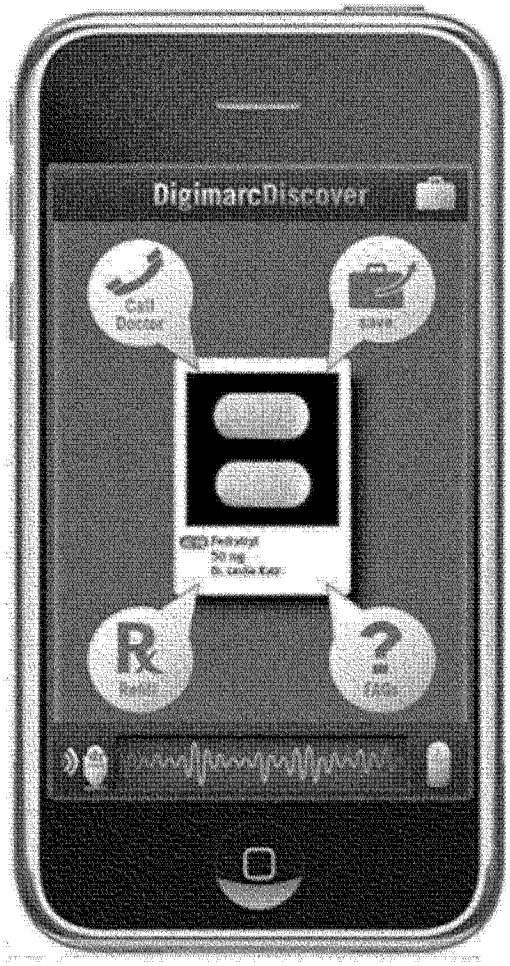


图 22

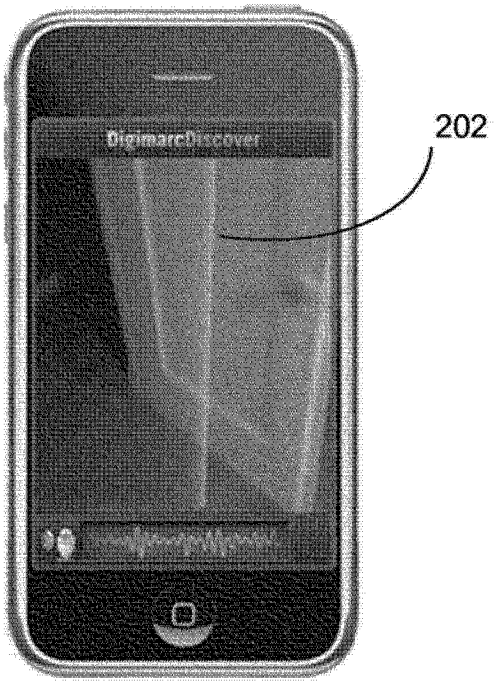


图 23A

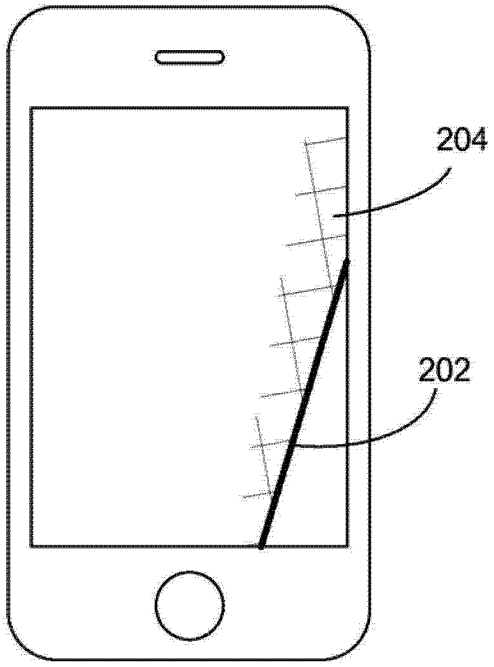


图 23B

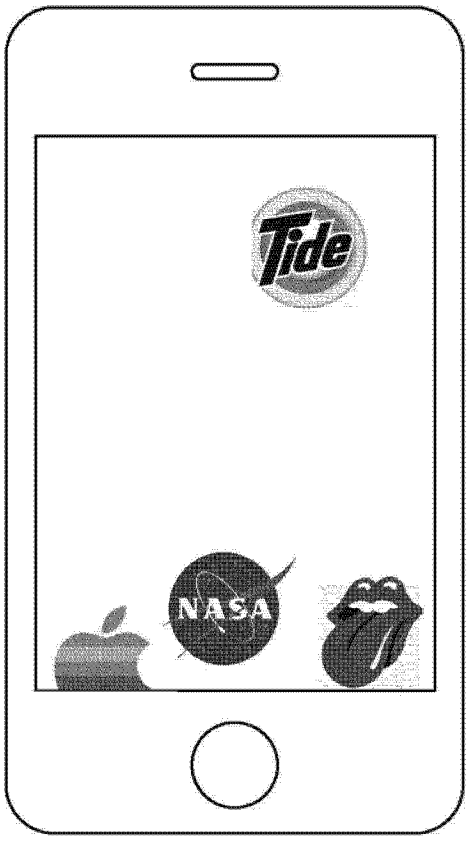


图 24

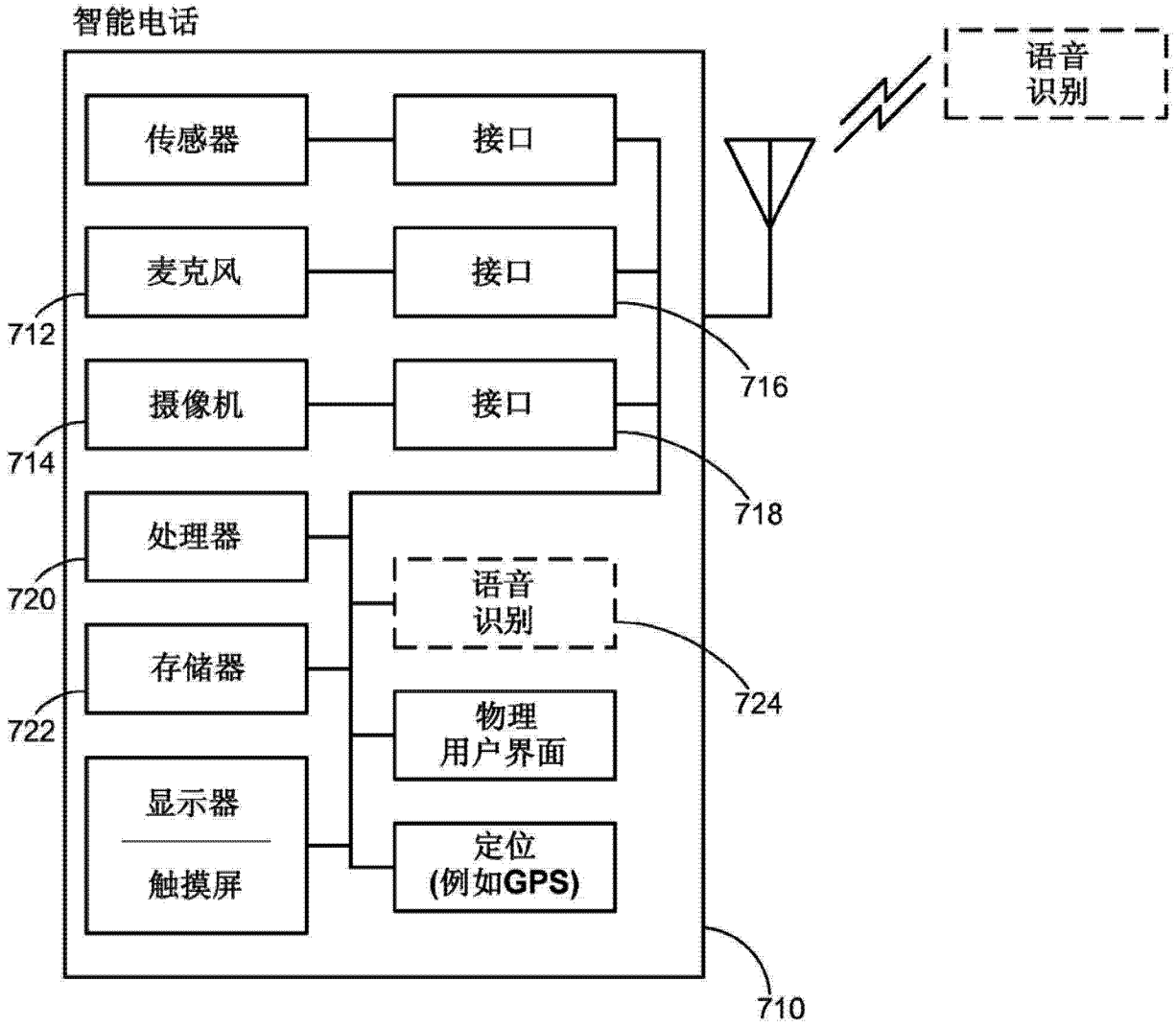

智能电话
传感器
接口
麦克风
接口
712
摄像机
接口
714
处理器
720
语音
识别
存储器
722
物理
用户界面
显示器
触摸屏
定位
(例如GPS)
语音
识别
716
718
724
710

图 25

对象	信号处理指令
男人	以5KHz进行采样；带通滤波器，85-2500Hz；应用语音识别；在屏幕上呈现文本
女人	以5KHz进行采样；带通滤波器，165-2500Hz；应用语音识别；在屏幕上呈现文本
无线电广播	以6KHz进行采样；带通滤波器，1-3Khz；解码Arbitron水印；在ARB数据库中查找；在屏幕上呈现元数据；否则，按照"歌曲"进行重新处理
电视	带通滤波器，1-4KHz；解码Nielsen水印；在Nielsen数据库中查找；在屏幕上呈现元数据；否则，按照"歌曲"进行重新处理
歌曲	低通滤波器；计算Shazam指纹；在Shazam数据库中查找；在屏幕上呈现元数据
音乐	按照"歌曲"进行处理

图 27

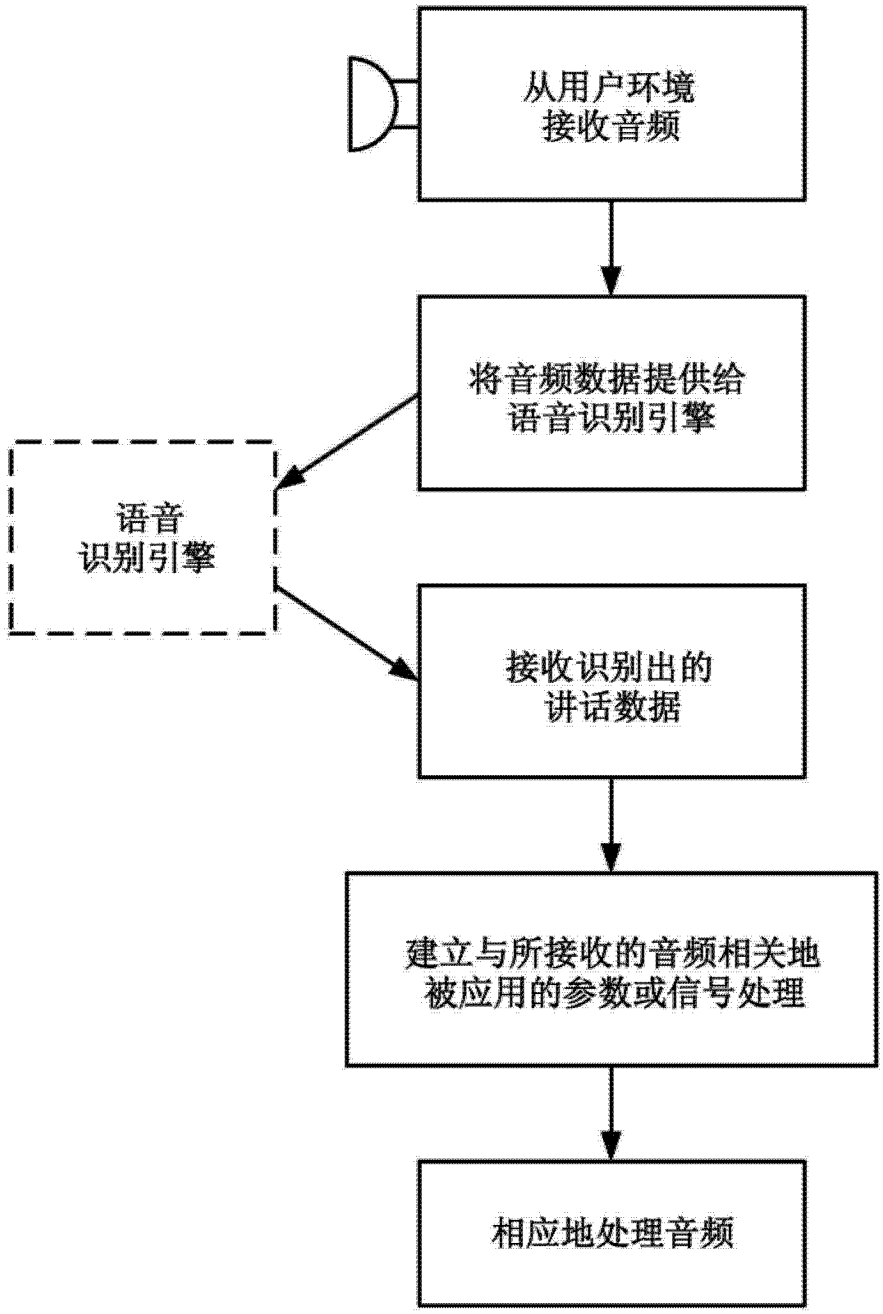

从用户环境
接收音频
将音频数据提供给
语音识别引擎
语音
识别引擎
接收识别出的
讲话数据
建立与所接收的音频相关地
被应用的参数或信号处理
相应地处理音频

图 26

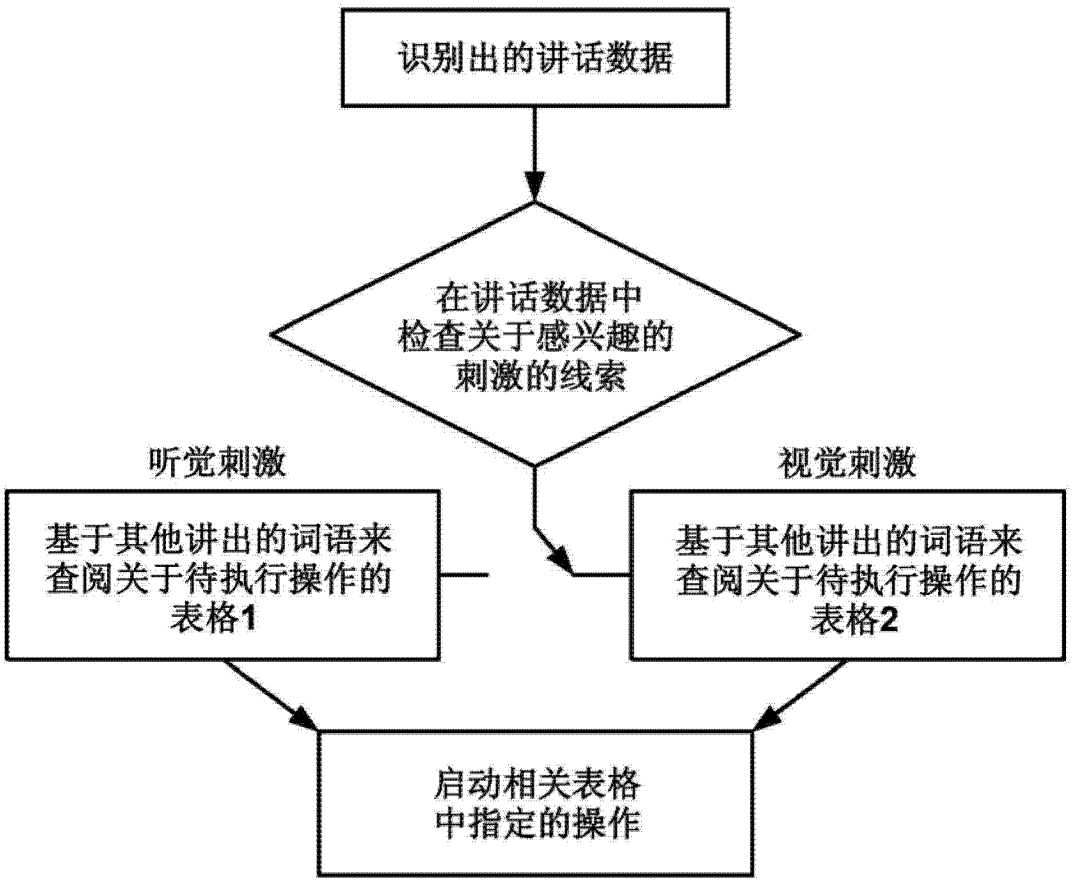

识别出的讲话数据
在讲话数据中
检查关于感兴趣的
刺激的线索
听觉刺激
基于其他讲出的词语来
查阅关于待执行操作的
表格1
视觉刺激
基于其他讲出的词语来
查阅关于待执行操作的
表格2
启动相关表格
中指定的操作

图 28

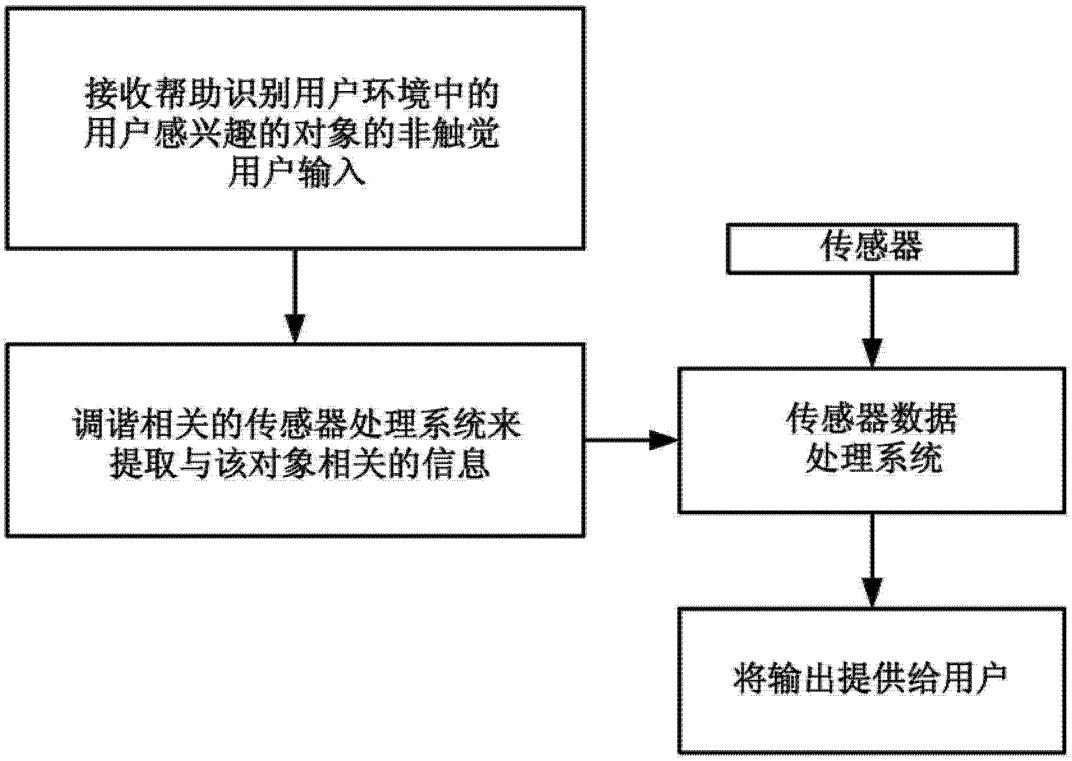

接收帮助识别用户环境中的
用户感兴趣的对象的非触觉
用户输入
传感器
调谐相关的传感器处理系统来
提取与该对象相关的信息
传感器数据
处理系统
将输出提供给用户

图 30

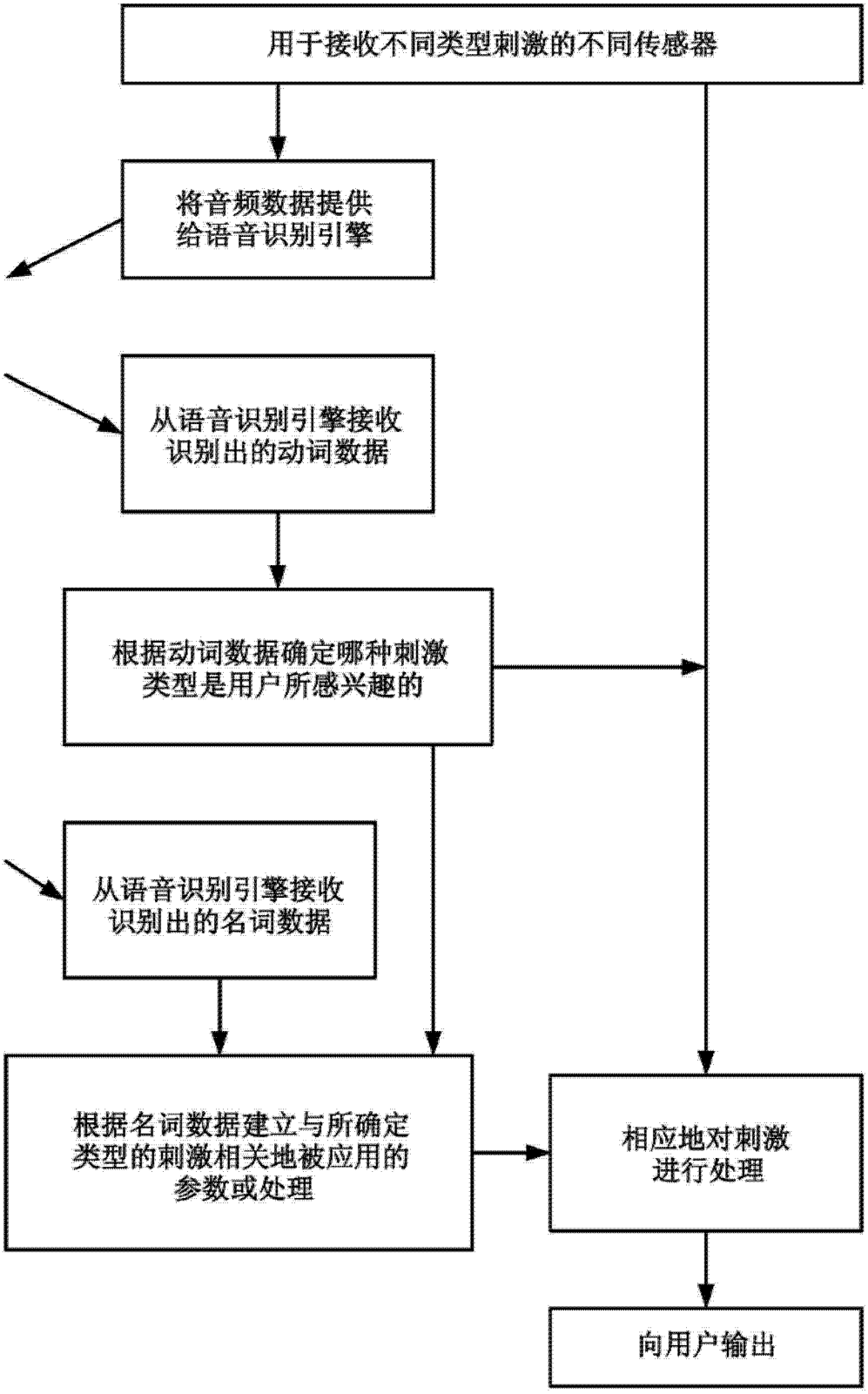

用于接收不同类型刺激的不同传感器
将音频数据提供给语音识别引擎
从语音识别引擎接收识别出的动词数据
根据动词数据确定哪种刺激类型是用户所感兴趣的
从语音识别引擎接收识别出的名词数据
根据名词数据建立与所确定类型的刺激相关地被应用的参数或处理
相应地对刺激进行处理
向用户输出

图 29