(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 110914872 A
(43)申请公布日 2020.03.24

(21)申请号 201880045644.9

(22)申请日 2018.07.05

(30)优先权数据
15/657,626 2017.07.24 US

(85)PCT国际申请进入国家阶段日
2020.01.08

(86)PCT国际申请的申请数据
PCT/IB2018/054963 2018.07.05

(87)PCT国际申请的公布数据
WO2019/021088 EN 2019.01.31

(71)申请人 国际商业机器公司
地址 美国纽约

(72)发明人 R·汉密尔顿二世 夏音 翟毓琳
G·博斯

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 李永敏 于静

(51)Int.Cl.
G06T 17/30(2006.01)

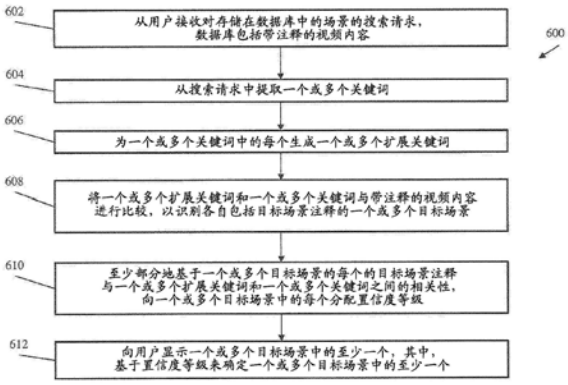
权利要求书4页 说明书12页 附图7页

(54)发明名称

用认知洞察力导航视频场景

(57)摘要

本发明的实施例包括用于从数据库获取场景的方法、系统和计算机程序产品。本发明的各方面包括接收对存储在包括带注释的视频内容的数据库中的场景的搜索请求。从搜索请求中提取一个或多个关键词。为每个关键词生成扩展关键词。将扩展关键词和关键词与带注释的视频内容进行比较，识别出包含目标场景注释的目标场景。至少部分地根据每个目标场景的目标场景注释与扩展关键词和关键词之间的相关性，为每个目标场景分配置信度等级。以及向用户显示至少一个目标场景，其中，基于置信度等级确定所述至少一个目标场景。



CN 110914872 A

1.一种用于从数据库获取场景的计算机实现的方法,该方法包括:

从用户接收对存储在数据库中的场景的搜索请求,该数据库包括带注释的视频内容;

从搜索请求中提取一个或多个关键词;

为所述一个或多个关键词中的每一个生成一个或多个扩展关键词;

将所述一个或多个扩展关键词和所述一个或多个关键词与所述带注释的视频内容进行比较,以识别一个或多个目标场景,所述一个或多个目标场景各自包括目标场景注释;

至少部分基于一个或多个目标场景的每个的目标场景注释与所述一个或多个扩展关键词和一个或多个关键词之间的相关性,为所述一个或多个目标场景的每个指定置信度等级;以及

向用户显示所述一个或多个目标场景中的至少一个,其中,基于所述置信度等级确定所述一个或多个目标场景中的所述至少一个。

2.根据权利要求1所述的方法,还包括:

从用户接收对所述一个或多个目标场景中的一个目标场景的选择的指示;以及

至少部分基于所述选择更新所述目标场景注释。

3.根据权利要求1所述的方法,还包括:

向用户显示所述一个或多个目标场景;

从用户接收对所述一个或多个目标场景中的一个目标场景的拒绝的指示;以及

至少部分基于所述拒绝更新所述目标场景注释。

4.根据权利要求2所述的方法,还包括:

监视用户以确定用户在观看目标场景时的情绪反应;

将用户的情绪反应与所述目标场景注释进行比较,以确定情绪反应与目标场景注释之间的相关性;以及

至少部分基于所述情绪反应与所述目标场景注释之间的相关性更新所述目标场景注释。

5.根据权利要求1所述的方法,其中所述搜索请求是用户音频输入,所述提取一个或多个关键词包括:

对所述搜索请求执行自然语言处理,以将用户音频输入转换为文本;

将文本分割成一个或多个单独的短语;以及

分析所述一个或多个单独的短语以确定关键词。

6.根据权利要求1所述的方法,还包括:

向用户显示所述一个或多个目标场景;以及

按照所述置信度等级排定显示所述一个或多个目标场景的顺序。

7.一种用于从数据库获取场景的计算机系统,所述计算机系统包括与存储器可通信地耦合的处理器,所述处理器被配置为:

从用户接收对存储在数据库中的场景的搜索请求,该数据库包括带注释的视频内容;

从搜索请求中提取一个或多个关键词;

为所述一个或多个关键词中的每一个生成一个或多个扩展关键词;

将所述一个或多个扩展关键词和所述一个或多个关键词与所述带注释的视频内容进行比较,以识别一个或多个目标场景,所述一个或多个目标场景各自包括目标场景注释;以

及

至少部分基于一个或多个目标场景的每个的目标场景注释与所述一个或多个扩展关键词和一个或多个关键词之间的相关性，为所述一个或多个目标场景的每个指定置信度等级。

8.根据权利要求7所述的计算机系统，其中，所述处理器被进一步配置为：

向用户显示所述一个或多个目标场景；

从用户接收对所述一个或多个目标场景中的一个目标场景的选择的指示；以及

至少部分基于所述选择更新所述目标场景注释。

9.根据权利要求7所述的计算机系统，其中，所述处理器被进一步配置为：

向用户显示所述一个或多个目标场景；

从用户接收对所述一个或多个目标场景中的一个目标场景的拒绝的指示；以及

至少部分基于所述拒绝更新所述目标场景注释。

10.根据权利要求8所述的计算机系统，其中，所述处理器被进一步配置为：

监视用户以确定用户在观看目标场景时的情绪反应；

将用户的情绪反应与所述目标场景注释进行比较，以确定情绪反应与目标场景注释之间的相关性；以及

至少部分基于所述情绪反应与所述目标场景注释之间的相关性更新所述目标场景注释。

11.一种用于从数据库获取场景的计算机程序产品，所述计算机程序产品包括具有其中体现程序指令的计算机可读存储介质，所述程序指令可由处理器执行以使所述处理执行：

从用户接收对存储在数据库中的场景的搜索请求，该数据库包括带注释的视频内容；

从搜索请求中提取一个或多个关键词；

为所述一个或多个关键词中的每一个生成一个或多个扩展关键词；

将所述一个或多个扩展关键词和所述一个或多个关键词与所述带注释的视频内容进行比较，以识别一个或多个目标场景，所述一个或多个目标场景各自包括目标场景注释；以及

至少部分基于一个或多个目标场景的每个的目标场景注释与所述一个或多个扩展关键词和一个或多个关键词之间的相关性，为所述一个或多个目标场景的每个指定置信度等级。

12.根据权利要求11所述的计算机产品，进一步包括：

向用户显示所述一个或多个目标场景；

从用户接收对所述一个或多个目标场景中的一个目标场景的选择的指示；以及

至少部分基于所述选择更新所述目标场景注释。

13.根据权利要求11所述的计算机产品，进一步包括：

向用户显示所述一个或多个目标场景；

从用户接收对所述一个或多个目标场景中的一个目标场景的拒绝的指示；以及

至少部分基于所述拒绝更新所述目标场景注释。

14.根据权利要求12所述的计算机产品，进一步包括：

监视用户以确定用户在观看目标场景时的情绪反应；

将用户的情绪反应与所述目标场景注释进行比较，以确定情绪反应与目标场景注释之间的相关性；以及

至少部分基于所述情绪反应与所述目标场景注释之间的相关性更新所述目标场景注释。

15.根据权利要求12所述的计算机产品，其中所述搜索请求是用户音频输入，所述提取一个或多个关键词包括：

对所述搜索请求执行自然语言处理，以将用户音频输入转换为文本；

将文本分割成一个或多个单独的短语；以及

分析所述一个或多个单独的短语以确定关键词。

16.一种用于注释视频场景的计算机实现的方法，该方法包括：

由处理器接收一个或多个视频；

将所述一个或多个视频分割成场景集合；

分析所述场景集合中的第一场景以确定第一场景的标签集合；

用所述标签集合注释第一场景；以及

将第一场景存储在数据库中。

17.根据权利要求16所述的方法，还包括：

至少部分基于第一场景的所述标签集合确定扩展关键词；以及

用第一场景注释所述扩展关键词。

18.根据权利要求16所述的方法，其中确定第一场景的标签集合包括：

对第一场景进行图像识别以为第一场景中的一个或多个对象生成标签。

19.根据权利要求16所述的方法，其中确定第一场景的标签集合还包括：

对第一场景进行语音分析以提取一个或多个关键词；

至少部分基于所述一个或多个关键词为第一场景生成标签；

用所述一个或多个关键词注释第一场景。

20.根据权利要求19所述的方法，还包括：

至少部分基于第一场景的所述一个或多个关键词确定扩展关键词；以及

用第一场景注释所述扩展关键词。

21.一种用于注释视频场景的计算机系统，所述计算机系统包括通信地耦合到存储器的处理器，所述处理器被配置为：

由处理器接收一个或多个视频；

将所述一个或多个视频分割成场景集合；

分析所述场景集合中的第一场景以确定第一场景的标签集合；

用所述标签集合注释第一场景；以及

将第一场景存储在数据库中。

22.根据权利要求21所述的计算机系统，所述处理器被进一步配置为：

至少部分基于第一场景的所述标签集合确定扩展关键词；以及

用第一场景注释所述扩展关键词。

23.根据权利要求21所述的计算机系统，其中确定第一场景的标签集合包括：

由所述处理器对第一场景进行图像识别以为第一场景中的一个或多个对象生成标签。

24.根据权利要求21所述的计算机系统,其中确定第一场景的标签集合还包括:

由所述处理器对第一场景进行语音分析以提取一个或多个关键词;

由所述处理器至少部分基于所述一个或多个关键词为第一场景生成标签;

由所述处理器用所述一个或多个关键词注释第一场景。

25.根据权利要求24所述的计算机系统,所述处理器被进一步配置为:

由所述处理器至少部分基于第一场景的所述一个或多个关键词确定扩展关键词;以及

由所述处理器用第一场景注释所述扩展关键词。

用认知洞察力导航视频场景

技术领域

[0001] 本发明涉及导航视频场景，更具体地说，涉及使用认知洞察力导航视频场景的方法和系统。

背景技术

[0002] 随着计算机性能的不断提升，数字视频的使用变得越来越普遍。观看视频已经成为我们日常生活中最受欢迎的娱乐活动之一，正如研究表明的那样，超过50％的互联网带宽消耗于视频服务。有了现成的摄像机、智能手机和其他手持设备，人们正在记录越来越多的旅行、重要的庆祝活动和家庭时光。此外，数字视频技术在大多数监控系统中都得到了应用。

[0003] 然而，使用和管理这些海量视频数据确实会出现问题。人们可能希望观看电影或家庭视频的特定场景，而不是整个视频。同样，警察或安全人员可能希望从大量的视频数据中找到并收集证据。

[0004] 在一个或多个视频文件中查找和导航到特定场景通常非常耗时且困难。使用现有技术，用户必须通过使用快进模式观看视频，或者反复拖动滑块，直到找到期望的场景为止。此外，一些视频按章节进行分解，用户可以导航到场景可能位于的特定章节。这些技术需要从用户的视频内存中进行大量导航才能找到场景。此外，用户可能以前没有观看过视频，可能在基于来自另一个观看者的建议来搜索场景。

[0005] 这一领域有一些解决方案，但它们并不能直接满足需求。例如，许多现代视频播放器“记住”播放电影的上一个断点，以便下一次从其停止的位置自动恢复电影。但是这个解决方案是基于历史记录的，并且是为单个用户和客户机设计的，因此，如果没有给定用户的以前的记录，或者在多用户的情况下，该解决方案不会很好地工作。

[0006] 其他系统采用面部识别来在视频流中查找特定的人；然而，这并不能解决使用一般描述来查找特定场景的问题，也不能解决在多个场景中查找特定场景的问题。在这种情况下使用这种技术将是一个问题，因为提供给用户的大量结果将需要额外的手工过滤。因此，需要一种据以能够从单个视频或大量视频数据中快速、准确地发现所需场景的新方法。

发明内容

[0007] 实施例包括一种用于从数据库获取场景的计算机实现的方法。该计算机实现的方法的非限制性示例包括接收对存储在数据库中的场景的搜索请求，该数据库中存储有带注释的视频内容。从搜索请求中提取一个或多个关键词。为每个关键词生成扩展关键词。将扩展关键词和关键词与带注释的视频内容进行比较，识别出包含目标场景注释的目标场景。至少部分地根据每个目标场景的目标场景注释与扩展关键词和关键词之间的相关性，为每个目标场景分配置信度等级。以及向用户显示至少一个目标场景，其中，基于置信度等级确定所述至少一个目标场景。

[0008] 实施例包括一种用于从数据库获取场景的计算机系统，该计算机系统具有处理

器，处理器被配置为执行一种方法。该系统的非限制性示例包括接收对存储在数据库中的场景的搜索请求，该数据库中存储有带注释的视频内容。从搜索请求中提取一个或多个关键词。为每个关键词生成扩展关键词。将扩展关键词和关键词与带注释的视频内容进行比较，识别出包含目标场景注释的目标场景。至少部分地根据每个目标场景的目标场景注释与扩展关键词和关键词之间的相关性，为每个目标场景分配置信度等级。以及向用户显示至少一个目标场景，其中，基于置信度等级确定所述至少一个目标场景。

[0009] 实施例还包括一种用于从数据库获取场景的计算机程序产品，该计算机程序产品包括具有计算机可读程序代码的非暂时性计算机可读存储介质。计算机可读程序代码包括被配置成执行一种方法的程序指令。该方法的非限制性示例包括接收对存储在数据库中的场景的搜索请求，该数据库中存储有带注释的视频内容。从搜索请求中提取一个或多个关键词。为每个关键词生成扩展关键词。将扩展关键词和关键词与带注释的视频内容进行比较，识别出包含目标场景注释的目标场景。至少部分地根据每个目标场景的目标场景注释与扩展关键词和关键词之间的相关性，为每个目标场景分配置信度等级。以及向用户显示至少一个目标场景，其中，基于置信度等级确定所述至少一个目标场景。

[0010] 实施例包括一种用于注释视频场景的计算机实现方法。计算机实现方法的非限制性示例包括由处理器接收一个或多个视频。将该一个或多个视频分割成场景集合。分析该场景集合中的第一场景以确定第一场景的标签集合。用该标签集合注释第一场景并将其存储在数据库中。

[0011] 实施例包括一种用于注释视频场景的计算机系统，该计算机系统具有处理器，该处理被配置为执行一种方法。该系统的非限制性示例包括由处理器接收一个或多个视频。将该一个或多个视频分割成场景集合。分析该场景集合中的第一场景以确定第一场景的标签集合。用该标签集合注释第一场景并将其存储在数据库中。

[0012] 通过本发明的技术实现了其它特征和优点。本文中详细描述了本发明的其他实施例和各个方面，它们被认为是所要求保护的本发明的一部分。为了更好地理解，请参阅说明和附图。

附图说明

[0013] 结尾处的权利要求书中特别指出并明确声明了本文所述专有权利的具体内容。以下结合附图的详细描述清楚地表明了本发明实施例的上述和其他特点和优点，其中：

[0014] 图1描绘根据本发明的一个或多个实施例的云计算环境；

[0015] 图2描绘根据本发明的一个或多个实施例的抽象模型层；

[0016] 图3示出了用于实践本文中的教导的计算机系统的框图；

[0017] 图4描绘了根据一个或多个实施例的用于导航数字视频的系统的框图；

[0018] 图5描绘了根据一个或多个实施例的用于导航数字视频的系统的示例；

[0019] 图6描绘了根据一个或多个实施例的用于从数据库获取场景的方法的流程图；

[0020] 图7描绘了根据一个或多个实施例的用于注释视频场景的方法的流程图。

[0021] 本文中所示的附图是说明性的。在不脱离本发明精神的情况下，附图或附图中描述的操作可以有许多变化。例如，可以按不同的顺序执行操作，也可以添加、删除或修改操作。此外，术语“耦合”及其变体描述了在两个元件之间具有通信路径，并且不意味着元件之

间的没有介于它们之间的中间元件/连接的直接连接。所有这些变化都被视为说明书的一部分。

[0022] 在附图以及以下对所公开的实施例的详细描述中，附图中所示的各种元件具有两位或三位数字的标记号。除了一些小的例外，每个标记号的最左边数字对应于其元素首次被示出的图。

具体实施方式

[0023] 本文参考相关附图描述本发明的各种实施例。在不脱离本发明范围的情况下，可以设计本发明的替代实施例。在以下描述和附图中的元素之间列出各种连接和位置关系(例如，上方、下方、相邻等)。除非另有规定，否则这些连接和/或位置关系可以是直接的或间接的，并且本发明并不旨在对这方面进行限制。因此，实体的耦合可以指直接或间接的耦合，实体之间的位置关系可以是直接或间接的位置关系。此外，本文所描述的各种任务和过程步骤可以并入具有本文未详细描述的其它步骤或功能的更全面的程序或过程。

[0024] 以下定义和缩写用于解释权利要求和说明书。如本文所使用的，术语“包括”、“包含”、“具有”、“有”、“带有”“含”或其任何其他变体旨在涵盖非排他性的包含。例如，包含一系列元素的成分、混合物、工艺、方法、物品或装置不一定仅限于这些元素，而是可以包括未明确列出或此类成分、混合物、工艺、方法、物品或装置固有的其他元素。

[0025] 此外，本文中的术语“示例性”是指“作为示例、实例或说明”。本文中描述为“示例性”的任何实施例或设计不一定被解释为优选的或优于其他实施例或设计。术语“至少一个”和“一个或多个”可理解为包括大于或等于一的任何整数，即一、二、三、四等。“多个”可理解为包括大于或等于二的任何整数，即二、三、四、五，等。术语“连接”可包括间接“连接”和直接“连接”

[0026] 术语“约”、“基本上”、“近似”及其变体意在包括与基于在提交申请时可用的设备的特定数量的测量相关联的误差程度。例如，“大约”可以包括给定值的±8％或5％或2％的范围。

[0027] 为了简洁起见，与本发明的制造和使用方面相关的传统技术在这里可能作了也可能没有作详细描述。特别地，计算系统的各个方面和实现本文所描述的各种技术特征的特定计算机程序是众所周知的。因此，为了简洁起见，许多传统的实现细节仅在本文中简要提及或完全省略，而不提供众所周知的系统和/或过程细节。

[0028] 首先应当理解，尽管本公开包括关于云计算的详细描述，但其中记载的技术方案的实现却不限于云计算环境，而是能够结合现在已知或以后开发的任何其它类型的计算环境而实现。

[0029] 云计算是一种服务交付模式，用于对共享的可配置计算资源池进行方便、按需的网络访问。可配置计算资源是能够以最小的管理成本或与服务提供者进行最少的交互就能快速部署和释放的资源，例如可以是网络、网络带宽、服务器、处理、内存、存储、应用、虚拟机和服务。这种云模式可以包括至少五个特征、至少三个服务模型和至少四个部署模型。

[0030] 特征包括：

[0031] 按需自助式服务：云的消费者在无需与服务提供者进行人为交互的情况下能够单方面自动地按需部署诸如服务器时间和网络存储等的计算能力。

[0032] 广泛的网络接入:计算能力可以通过标准机制在网络上获取,这种标准机制促进了通过不同种类的瘦客户机平台或厚客户机平台(例如移动电话、膝上型电脑、个人数字助理PDA)对云的使用。

[0033] 资源池:提供者的计算资源被归入资源池并通过多租户(multi-tenant)模式服务于多重消费者,其中按需将不同的实体资源和虚拟资源动态地分配和再分配。一般情况下,消费者不能控制或甚至并不知晓所提供的资源的确切位置,但可以在较高抽象程度上指定位置(例如国家、州或数据中心),因此具有位置无关性。

[0034] 迅速弹性:能够迅速、有弹性地(有时是自动地)部署计算能力,以实现快速扩展,并且能迅速释放来快速缩小。在消费者看来,用于部署的可用计算能力往往显得是无限的,并能在任意时候都能获取任意数量的计算能力。

[0035] 可测量的服务:云系统通过利用适于服务类型(例如存储、处理、带宽和活跃用户帐号)的某种抽象程度的计量能力,自动地控制和优化资源效用。可以监测、控制和报告资源使用情况,为服务提供者和消费者双方提供透明度。

[0036] 基础架构即服务(IaaS):向消费者提供的能力是消费者能够在其中部署并运行包括操作系统和应用的任意软件的处理、存储、网络和其他基础计算资源。消费者既不管理也不控制底层的云基础架构,但是对操作系统、存储和其部署的应用具有控制权,对选择的网络组件(例如主机防火墙)可能具有有限的控制权。

[0037] 部署模型如下:

[0038] 私有云:云基础架构单独为某个组织运行。云基础架构可以由该组织或第三方管理并且可以存在于该组织内部或外部。

[0039] 共同体云:云基础架构被若干组织共享并支持有共同利害关系(例如任务使命、安全要求、政策和合规考虑)的特定共同体。共同体云可以由共同体内的多个组织或第三方管理并且可以存在于该共同体内部或外部。

[0040] 公共云:云基础架构向公众或大型产业群提供并由出售云服务的组织拥有。

[0041] 混合云:云基础架构由两个或更多部署模型的云(私有云、共同体云或公共云)组成,这些云依然是独特的实体,但是通过使数据和应用能够移植的标准化技术或私有技术(例如用于云之间的负载平衡的云突发流量分担技术)绑定在一起。

[0042] 云计算环境是面向服务的,特点集中在无状态性、低耦合性、模块性和语意的互操作性。云计算的核心是包含互连节点网络的基础架构。

[0043] 现在参考图1,其中显示了示例性的云计算环境50。如图所示,云计算环境50包括云计算消费者使用的本地计算设备可以与其相通信的一个或者多个云计算节点10,本地计算设备例如可以是个人数字助理(PDA)或移动电话54A,台式电脑54B、笔记本电脑54C和/或汽车计算机系统54N。云计算节点10之间可以相互通信。可以在包括但不限于如上所述的私有云、共同体云、公共云或混合云或者它们的组合的一个或者多个网络中将云计算节点10进行物理或虚拟分组(图中未显示)。这样,云的消费者无需在本地计算设备上维护资源就能请求云计算环境50提供的基础架构即服务(IaaS)、平台即服务(PaaS)和/或软件即服务(SaaS)。应当理解,图1显示的各类计算设备54A-N仅仅是示意性的,云计算节点10以及云计算环境50可以与任意类型网络上和/或网络可寻址连接的任意类型的计算设备(例如使用网络浏览器)通信。

[0044] 现在参考图2,其中显示了云计算环境50(图1)提供的一组功能抽象层。首先应当理解,图2所示的组件、层以及功能都仅仅是示意性的,本发明的实施例不限于此。

[0045] 硬件和软件层60包括硬件和软件组件。硬件组件的例子包括:主机61;基于RISC(精简指令集计算机)体系结构的服务器62;服务器63;刀片服务器64;存储设备65;网络和网络组件66。软件组件的例子包括:网络应用服务器软件67以及数据库软件68。

[0046] 虚拟层70提供一个抽象层,该层可以提供下列虚拟实体的例子:虚拟服务器71、虚拟存储72、虚拟网络73(包括虚拟私有网络)、虚拟应用和操作系统74,以及虚拟客户端75。

[0047] 在一个示例中,管理层80可以提供下述功能:资源供应功能81:提供用于在云计算环境中执行任务的计算资源和其它资源的动态获取;计量和定价功能82:在云计算环境内对资源的使用进行成本跟踪,并为此提供帐单和发票。在一个例子中,该资源可以包括应用软件许可。安全功能:为云的消费者和任务提供身份认证,为数据和其它资源提供保护。用户门户功能83:为消费者和系统管理员提供对云计算环境的访问。服务水平管理功能84:提供云计算资源的分配和管理,以满足必需的服务水平。服务水平协议(SLA)计划和履行功能85:为根据SLA预测的对云计算资源未来需求提供预先安排和供应。

[0048] 工作负载层90提供云计算环境可能实现的功能的示例。在该层中,可提供的工作负载或功能的示例包括:地图绘制与导航91;软件开发及生命周期管理92;虚拟教室的教学提供93;数据分析处理94;从数据库获取视频95;以及注释视频场景96。

[0049] 参考图3,示出了用于实现本文的教导的处理系统100的实施例。在本实施例中,系统100具有一个或多个中央处理单元(处理器)101a、101b、101c等(统称或一般称为处理器101)。在一个或多个实施例中,每个处理器101可以包括精简指令集计算机(RISC)微处理器。处理器101经由系统总线113耦合到系统存储器114和各种其它组件。只读存储器(ROM)102耦合到系统总线113,并且可以包括控制系统100的某些基本功能的基本输入/输出系统(BIOS)。

[0050] 图3进一步描绘了耦合到系统总线113的输入/输出(I/O)适配器107和网络适配器106。I/O适配器107可以是与硬盘103和/或磁带存储驱动器105或任何其他类似组件通信的小型计算机系统接口(SCSI)适配器。这里,I/O适配器107、硬盘103和磁带存储设备105统称为大容量存储104。用于在处理系统100上执行的操作系统120可以存储在大容量存储器104中。网络适配器106将总线113与外部网络116互连,使得数据处理系统100能够与其他这样的系统通信。显示屏(例如,显示监视器)115通过显示适配器112连接到系统总线113,显示适配器112可以包括用于提高图形密集型应用的性能的图形适配器和视频控制器。在一个实施例中,适配器107、106和112可以通过中间总线桥(未示出)连接到一个或多个连接到系统总线113的I/O总线。用于连接外围设备(例如硬盘控制器、网络适配器和图形适配器)的合适I/O总线通常包括诸如外围组件互连(PCI)的公共协议。其它输入/输出设备被示为经由用户接口适配器108和显示适配器112连接到系统总线113。键盘109、鼠标110和扬声器111都经由用户接口适配器108互连到总线113,用户接口适配器108可以包括例如将多个设备适配器集成到单个集成电路中的超级I/O芯片。

[0051] 在示例性实施例中,处理系统100包括图形处理单元130。图形处理单元130是专用的电子电路,其设计用于操纵和改变存储器以加速在帧缓冲器中创建用于输出到显示器的图像。一般而言,图形处理单元130在处理计算机图形和图像处理方面非常有效,并且具有

高度并行的结构，使得其对于并行处理大块数据的算法比通用中央处理单元更有效。

[0052] 因此，如图3所配置，系统100包括处理器101形式的处理能力、包括系统存储器114和大容量存储器104的存储能力、诸如键盘109和鼠标110等的输入装置以及包括扬声器111和显示屏115的输出能力。在一个实施例中，系统存储器114和大容量存储器104的一部分共同存储操作系统，以协调图3所示的各种组件的功能。

[0053] 本发明的一个或多个实施例提供用于注释视频场景和从数据库获取视频场景的系统、方法和计算机程序产品。本发明的各方面包括利用技术分析视频内容以识别视频场景中的对象、实体、动作、概念和情感，并提供与视频场景相关联的标签。这些标签可以是关键词的形式，也可以是自然语言的描述(即，描述视频场景的句子)。除了关键词标签之外，还会创建与提取的关键词关联的扩展关键词。例如，诸如“棒球”之类的关键词可以与诸如运动场、投手、体育场、土墩、垒等扩展关键词相关联。这些扩展关键词也可以与视频场景一起注释。这些带注释的视频场景可以存储在数据库中用于搜索。

[0054] 在本发明的一个或多个实施例中，用户可以在数据库中搜索视频场景。用户可以提交对存储在数据库中的视频场景的搜索请求。搜索请求可以是用户的音频输入，也可以是用户对特定视频场景的文本输入。从搜索请求中，可以提取关键词并与视频场景的注释进行比较，以确定匹配场景的列表。此外，可以从搜索请求的提取关键词创建扩展关键词。扩展关键词可以与视频场景的注释进行比较，以确定匹配场景的列表。

[0055] 图4描绘了根据一个或多个实施例的用于导航数字视频的系统400的框图。系统400包括服务器401、用于视频场景的参考数据库402、外部库405和客户端406。服务器401包括图像识别模块403、自然语言处理(NLP)模块404和通信模块408。服务器401还包括分析模块420，分析模块420包括概念标记、情绪(emotion)分析、情感(sentiment)分析和关系提取。客户端406包括用户输入模块407、显示模块410和与客户端406电子通信的传感器430。

[0056] 在本发明的一个或多个实施例中，可以在图3中的处理系统100上实现服务器401、客户端406、图像识别模块403、自然语言处理模块404、通信模块408和分析模块420。此外，云计算系统50可以与系统400的一个或所有元素进行有线或无线电子通信。云50可以补充、支持或替换系统400的部分或全部元素的功能。另外，系统400的元素的部分或全部功能可以实现为云50的节点10(如图1和2所示)。云计算节点10仅是合适的云计算节点的一个示例，并且无意建议对本文所描述的本发明实施例的使用范围或功能性的任何限制。

[0057] 在本发明的一个或多个实施例中，系统400可用于注释视频数据。注释视频数据包括以与相应视频场景相关联的语义属性(单词形式或句子形式，例如标记、描述等)的形式应用注释。系统400包括服务器401和存储视频数据的参考数据库402。视频数据可以包括电影、电视节目、互联网视频等。视频数据可以分割成不同长度的视频场景。场景的长度可以与主题或概念相关联。例如，视频中的完整婚礼场景可能会持续几分钟。但是，可以根据发生的动作或诸如沿过道走动或婚礼招待等情节，将整个婚礼场景进一步分为较短的场景。视频数据可以根据视频场景的概念、情绪和情感进行进一步划分。

[0058] 服务器401利用分析模块420来注释(有时称为“标签”或“标记”)参考数据库402上的视频场景，该模块利用概念标签、情绪分析、情感分析和关系提取。服务器401还利用图像识别模块403来识别视频场景内的对象。NLP模块404用于分析和识别每个视频场景中的音频，以提取视频场景的概念、情绪、情感、关系分析和注释。

[0059] 情感分析技术包括(通过NLP)对以文本和音频表达的观点进行识别和分类,以确定说话者或其他主体(subject)对主题、产品或整体语境极性或对对象、互动或事件的情绪反应的态度。可以提取的情感包括但不限于积极、消极和中性。态度可以是判断或评估、情感状态(即情绪状态)或预期的情绪交流(即说话者预期的情感效果)。除了如上所述通过情感分析提取的情绪外,情绪分析还可以包括通过诸如面部表情识别之类的技术分析来个体的面部,以确定个体的一种或多种情绪。情绪分析还可至少部分基于面部表情识别来确定个体的情绪变化。

[0060] 图像识别模块403用于确定从参考数据库402获取的各种视频场景中的对象。执行图像识别以识别和标识场景中的多个图像中的形状和对象。在图像识别的执行期间使用的特定图像识别算法可以是可用于特定应用或处理约束的任何合适的图像或图案识别算法。图像识别算法可能受到用于提供一个或多个场景中的对象与已知对象的匹配的可用数据库的限制。作为一个示例,图像识别算法可以涉及图像的预处理。预处理可以包括但不限于调整图像的对比度、转换为灰度和/或黑白、裁剪、调整大小、旋转以及它们的组合。根据某些图像识别算法,可以选择诸如颜色、大小或形状的区分特征,用于检测特定对象。可以使用提供对象的区分特征的多个特征。可以执行边缘检测以确定视频场景中对象的边缘。可以在图像识别算法中执行形态学,以对像素集进行动作,包括去除不需要的成分。另外,可以执行降噪和/或区域填充。另外,一种图像识别算法,一旦在图像中找到/检测到一个或多个对象(及其关联的属性),则该一个或多个对象各自在视频场景中被定位,然后被分类。通过根据与区分特征有关的特定规格评估所定位的对象,可以对所定位的物体进行分类(即,识别为特定形状或物体)。特定规格可以包括数学计算或关系。在另一示例中,除了在视频场景中定位可识别对象之外或作为替代,还可以执行模式匹配。可以通过将图像中的元素和/或对象与“已知”(先前识别或分类的)对象和元素(例如,标记的训练数据)进行比较来进行匹配。图像识别模块403可以通过将视频场景中的识别出的对象与标记的训练数据进行比较来利用机器学习,以验证识别的准确性。图像识别模块403可以利用神经网络(NN)和其他学习算法。图像识别模块403的标识过程可以包括标识的置信度,例如阈值置信度。置信度低于阈值的任何对象的标识都可以丢弃。如果某对象的标识的置信度高于某个阈值,则可以为场景添加注释(标记)该对象。例如,某场景可能包括背景中的车辆,并且图像识别模块403可以将识别出该车辆为摩托车,则标签或注释可以包括该场景的标签“摩托车”。图像识别模块403还可以识别关于摩托车的特征,例如颜色、位置、正在运行还是停泊、品牌等。应当理解,尽管所描述的实施例和示例中的某些可以参考图像识别,但是这不应被解释为将所描述的实施例和示例限制为仅图像。例如,视频信号可以被系统400接收并且经历根据本发明的一个或多个实施例所描述的自动标签生成过程。可以接收来自参考数据库402的一个或多个视频帧,其中该视频帧可以包括图像并且可以执行图像识别。

[0061] 在本发明的一个或多个实施例中,图像识别模块403被用于识别视频场景中的人物、物体、实体和其他特征。分析模块420用于确定这些人物、对象、实体和其他特征之间的关系,以用于视频场景的注释。例如,对法庭中某角色的识别连同关系提取可以将该角色识别为法官。对该场景的注释可以用与法律诉讼或判决等相关的关键词标记。角色一旦被识别,就可以与外部库405进行交叉参考。外部库405包括但不限于互联网电影数据库(IMDB)、电子节目指南(EPG)以及与视频场景有关的其他类似的外部库。通过交叉参考外部库405中

的角色描述来确认上述示例中的角色是法官，可以进一步改进对角色的识别。

[0062] 通过NLP模块404的文本和语音分析被用来分析与视频场景相关的字幕/子标题(subtitles)和对话。分析模块420用于确定情感、实体、动作和概念。例如，可以由NLP模块404来分析与视频场景相关联的音频数据，以通过诸如语音到文本(STT)的技术将音频数据转换成文本。文本格式格式的关键词可以被提取，用于注释视频场景。

[0063] 在一个或多个实施例中，服务器401可以通过通信模块408与客户端406通信。客户端406包括与客户端406进行电子通信的一个或多个传感器430。客户端406可以是用于通过显示模块410为观众显示视频场景的任何类型的计算机或接口(interface)。例如，客户端406可以是在智能手机上显示视频场景以供智能手机周围的一群人(即，观众)观看的智能手机。传感器430可以包括照相机和/或麦克风。在该示例中，传感器430可以是智能手机中的记录观众对视频场景的反应的嵌入式麦克风和照相机。可以使用服务器401上的分析模块420来分析包括情绪、情感等的观众反应。这些情绪、情感等可以进一步注释参考数据库402上的视频场景。

[0064] 观众可以是任何类型的观众，包括电影院中的个人，观看个人视频的家庭成员等。客户端406可以包括智能电视和可以与传感器430通信并将传感器数据从传感器430传输到服务器以由分析模块420分析的其他系统。系统400记录听众在观看视频场景期间的口头评论，并利用NLP模块404提取短语和关键词进行分析。另外，系统400将利用传感器430来记录观众在观看视频场景时观众的面部表情和身体姿势。可以由分析模块420利用情绪分析和情感分析来分析该传感器数据。例如，可能被记录的反应可以包括诸如惊讶、害怕、哭泣、高兴等反应。在一个或多个实施例中，至少部分地基于观众的反应，系统400可以注释观众正在观看的视频场景。注释可以包括观众的情绪反应。

[0065] 包括观众对视频场景的反应的优点包括确认场景的情感分析。可以使用诸如机器学习之类的任何合适的学习算法来执行情感分析和情绪分析。观众的反应可以确认或拒绝由学习算法创建的标签，并辅助教导该学习算法。例如，通过情感分析，学习算法可以基于场景中的实体、音频内容和关系来识别情感。识别出的情感可以具有与其相关的置信度。至少部分地基于该置信度，可以获得观众的反应以确认或拒绝场景的情绪。如果将场景的情感标记为“悲伤”场景，但观众的反应包括“欢乐”和“笑声”，则可以调整标签以匹配观众的反应并训练机器学习算法。

[0066] 在本发明的一个或多个实施例中，系统400可用于导航数字视频。用户通过客户端406上的用户输入模块407请求视频中的特定场景。客户端406可以是个人计算设备、电视、智能手机或其他智能设备等。用户输入模块407被配置成通过用户的音频输入、文本输入或图形输入来接收用户对视频中的场景的查询。例如，用户可以在客户端406的外围设备上键入其请求作为文本输入，例如用键盘或鼠标在屏幕上选择字母。此外，用户可以通过可通信地耦合到客户端406的外围设备(例如，与客户端406电子通信的麦克风等)口头地发送请求。用户还可以通过用户输入模块407选择查询的图形表示。例如，查询可以包括表示视频中的动作场景或视频中的爱情场景的图标。在这种情况下，用户可以选择该图标来向服务器401提交查询。

[0067] 客户端406通过用户输入模块407接收用户输入，并查询服务器401以找到最符合用户请求的一个或多个视频场景。服务器401具有存储在参考数据库402中的视频数据。视

频数据包括具有相应注释(有时称为“标记”或“关键词”)的视频场景。注释包括描述视频场景中的对象、情感、动作、实体和概念的标签或标记。对象可以包括演员、车辆、位置、建筑物、动物等。这些概念可以包括场景的情感,如浪漫、快乐、幸福等等。概念还可以包括场景的类别,如动作、惊险、恐怖等。

[0068] 在本发明的一个或多个实施例中,用户输入模块407从用户接收查询并发送到通信模块408。该查询被发送到NLP模块404进行分析。NLP模块404可以利用诸如语音到文本(STT)等技术将音频查询转换为文本格式。NLP模块404分析文本格式的查询以从查询语言中提取关键词。例如,一个查询请求″the scene where Bill hits a home run″(比尔击中本垒打的场景)。提取的关键词包括“Bill”(比尔)、“home run”(本垒打)和“hits”(击中)。提取的关键词用于识别扩展关键词。在前面的示例中,扩展关键词可以包括“棒球”、“棒球场”、“球棒”、“球衣”、“得分”等。扩展关键词还可以包括演员姓名,其中关键词“Bill”将扩展到演员的全名或电影中角色的全名。由场景查询模块409将关键词和扩展关键词与参考数据库402中的注释场景进行比较,以找到与关键词和扩展关键词最匹配的一个或多个匹配场景。该一个或多个匹配场景可以在客户端406的显示模块410上呈现给用户。

[0069] 使用NLP模块404接收用户输入的优点包括为用户访问视频内容创建用户友好的方式。客户端406可以包括任何类型的电子设备,包括智能手机。智能手机通常没有输入搜索请求的简单方法,不像电脑有键盘来输入搜索请求。接收音频请求并且能够提取关键词对于各种类型的客户端406来说都具有优点和灵活性。

[0070] 采用情感和情感分析为标记视频场景以赋予用户提供更多的搜索选项创造了优势。通常,用户在寻找场景时会试图描述位置或角色。相反,通过包含情感和情感的能力,用户能够对场景进行更广泛的搜索,特别是当用户不记得场景中的角色名称或描述时。此外,用户可能不是在寻找特定场景。取而代之的是,用户可能想要一个带有某种情绪的场景。例如,如果用户正在准备一个讲演,希望找到要包括在讲演中的“令人振奋的”场景,则用户可以搜索“令人振奋的场景”,并且系统400可以返回与该描述匹配的多个场景。

[0071] 图5描绘了根据本发明的一个或多个实施例的用于导航数字视频的系统的说明性示例。系统500包括用户输入“I want to watch the scene when Bob is proposing to Alice in a park”(“我想观看鲍勃在公园向爱丽丝求婚时的场景”)的用户输入501。用NLP模块404来移除用户输入501中与场景描述无关的内容,以创建用户语句503。从用户语句503中提取关键词502以及扩展关键词。抽取的关键词502包括“Bob”(鲍勃)、“Alice”(爱丽丝)、“Proposing”(求婚)和“Park”(公园),“Park”的扩展关键词包括“Tree”(树木)、“Grass”(草地)和“Pond”(池塘),“Proposing”的扩展关键词包括“Ring”(耳环)、“Diamond”(钻石)和“Flowers”(鲜花),(图4的)场景查询模块409将关键词和扩展关键词与参考数据库402中的视频场景进行比较,以确定与关键词和扩展关键词匹配的一个或多个视频场景。如图所示,参考数据库402中的视频场景包括以场景标记和场景描述505为形式的注释。这些注释是运用了上述技术而产生的。在说明性示例中,场景标记包括“Bob”、“Alice”、“A rose”(玫瑰花)、“Tree”、“Grass”和“Propose”(求婚)。此外,场景描述505包括“Bob is proposing to Alice in the park”(鲍勃在公园里向爱丽丝求婚)。至少部分基于用户语句的关键词与视频场景的注释的比较,场景查询模块409选择该视频场景,并将其呈现给客户端406以供显示。

[0072] 在本发明的一个或多个实施例中，场景查询模块409可以基于用户输入关键词和视频场景的注释的比较来确定所识别视频场景的置信值。当呈现可能匹配的场景的列表时，该置信值可以显示在客户端设备406上。至少部分地基于用户选择特定场景，服务器401可以增加视频场景的注释的置信度并更新视频场景的注释。例如，如果5个关键词中有4个匹配视频场景的注释，并且用户随后选择该特定场景，则可以使用第5个关键词更新视频场景，以获得经用户确认的更好的注释。在本发明的一个或多个实施例中，可以基于查询语言和系统的多个用户的随后选择而连续地更新视频场景的注释。

[0073] 在本发明的一个或多个实施例中，可以使用图像识别模块403来识别视频场景中的对象，并且可以使用机器学习技术来验证该识别。例如，在图5中，场景包括玫瑰，图像识别模块403可以返回诸如“花”、“玫瑰”或“红花”之类的标签。图像识别模块403可以将这些标签与已知的(标记了的)花、玫瑰和/或红花的图像进行比较，以验证场景中标识的该对象的标签。

[0074] 现在参考图6，示出了根据本发明一个或多个实施例的用于从数据库获取场景的方法600的流程图。如方框602所示。方法600包括从用户接收对存储在数据库中的场景的搜索请求，该数据库包括带注释的视频内容。在方框604，方法600包括从搜索请求中提取一个或多个关键词。在方框606，方法600包括为该一个或多个关键词中的每一个生成一个或多个扩展关键词。在方框608，方法600包括将该一个或多个扩展关键词和该一个或多个关键词与带注释的视频内容进行比较以识别一个或多个目标场景，该一个或多个目标场景各自包括目标场景注释。如方框610所示，方法600包括至少部分地基于该一个或多个目标场景的每一个的目标场景注释与该一个或多个扩展关键词和一个或多个关键词之间的相关性，向该一个或多个目标场景中的每一个分配一个置信度等级。在方框612，方法600包括向用户显示该一个或多个目标场景中的至少一个，其中，基于所述置信度等级来确定该一个或多个目标场景中的该至少一个。

[0075] 还可以包括其他过程。应该理解的是，图6所示的过程代表示例，在不脱离本公开的范围和精神的情况下，可以添加其它过程，或者可以删除、修改或重新排列现有的过程。

[0076] 现在参考图7，示出了根据本发明的一个或多个实施例的用于注释视频场景的方法700的流程图。方法700包括由处理器接收一个或多个视频，如方框702所示。在方框704，方法700包括将一个或多个视频中的每一个分割成一组场景。在方框706处，方法700包括分析场景集合中的第一场景以确定第一场景的标签集合。在方框708，方法700包括用该标签集合注释第一场景。在方框710，方法700包括将第一场景存储在数据库中。

[0077] 还可以包括其他过程。应该理解的是，图7所示的过程代表示例，在不脱离本公开的范围和精神的情况下，可以添加其它过程，或者可以删除、修改或重新排列现有的过程。

[0078] 本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质，其上载有用于使处理器实现本发明的各个方面的计算机可读程序指令。

[0079] 计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是——但不限于——电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括：便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式

压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身,诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如,通过光纤电缆的光脉冲)、或者通过电线传输的电信号。

[0080] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备,或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令,并转发该计算机可读程序指令,以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0081] 用于执行本发明操作的计算机程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等,以及过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中,通过利用计算机可读程序指令的状态信息来个性化定制电子电路,例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA),该电子电路可以执行计算机可读程序指令,从而实现本发明的各个方面。

[0082] 这里参照根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本发明的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0083] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,从而生产出一种机器,使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制造品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0084] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上,使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0085] 附图中的流程图和框图显示了根据本发明的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用

于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

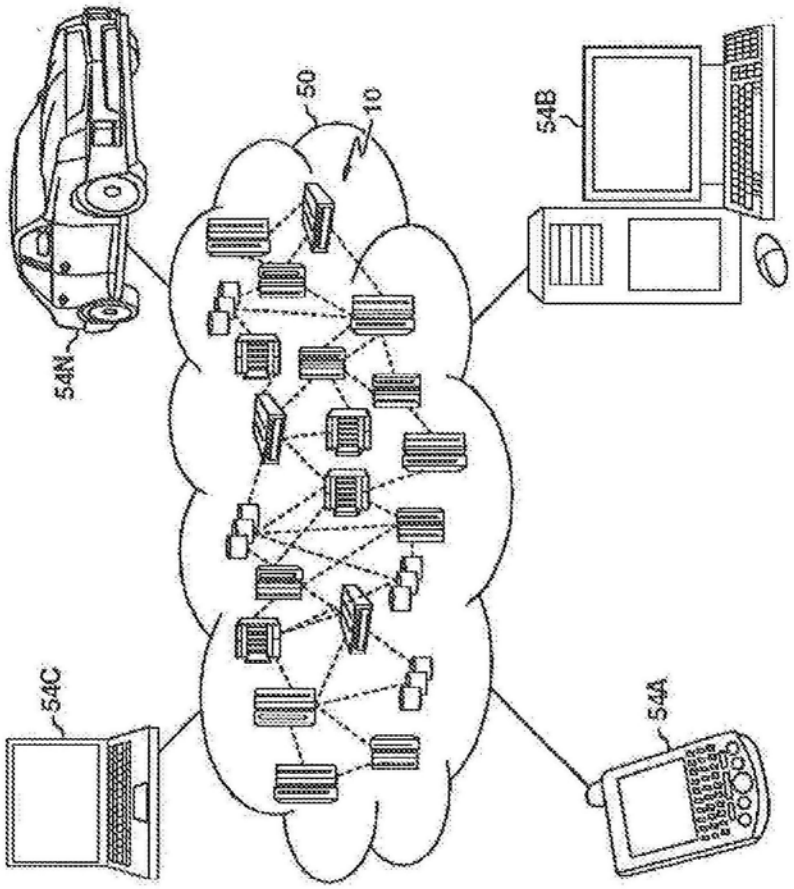


图1

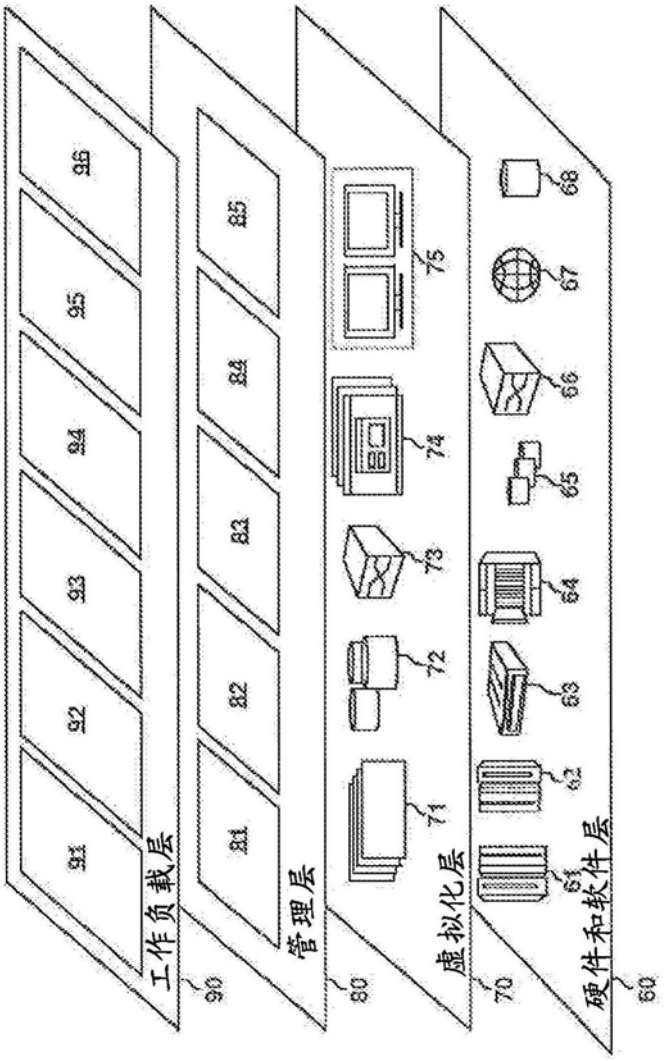

工作负载层
管理层
虚拟化层
硬件和软件层
90
80
70
60
91
92
93
94
95
96
81
82
83
84
85
71
72
73
74
75
61
62
63
64
65
66
67
68

图2

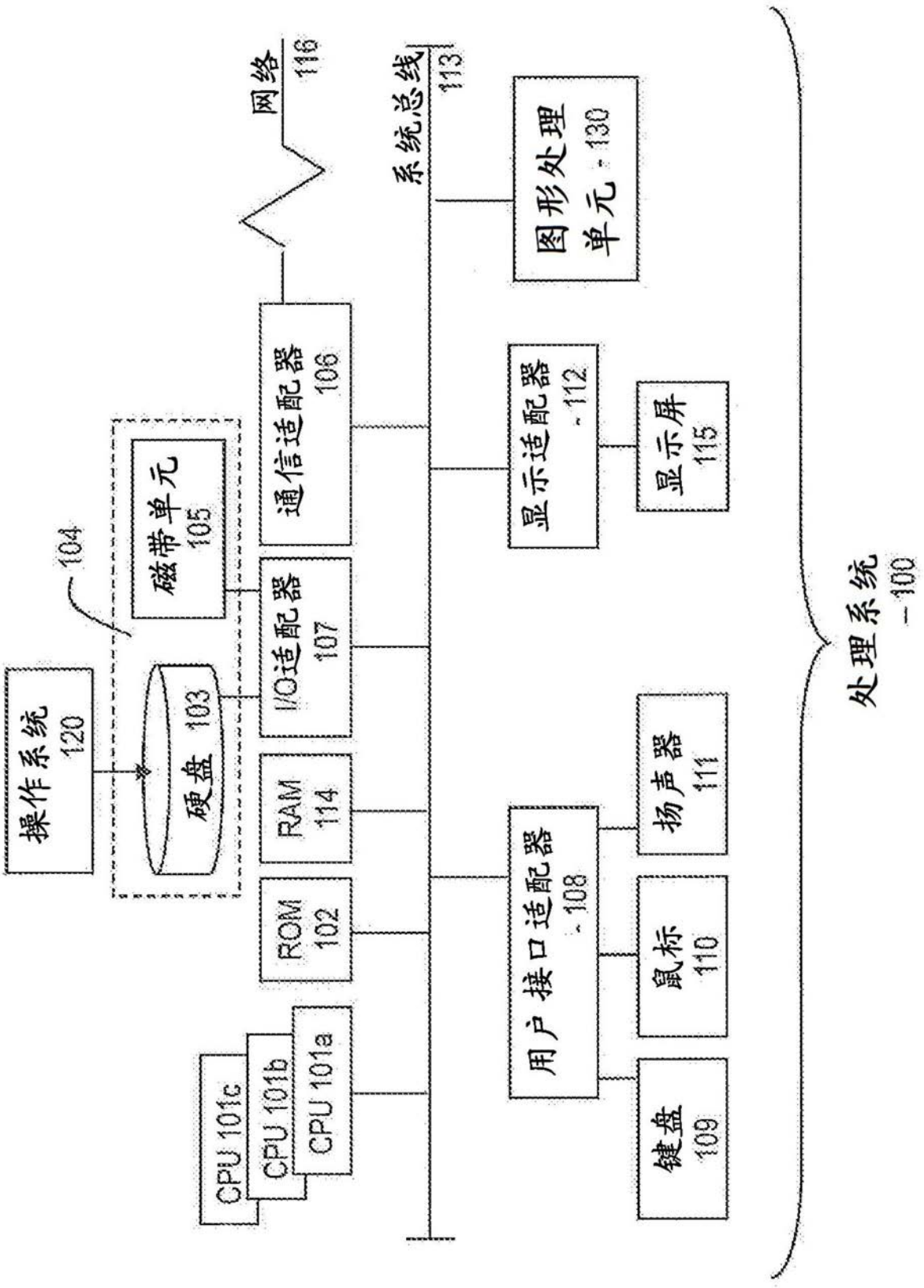

处理系统
100
操作系统
120
104
硬盘 103
磁带单元
105
网络
116
通信适配器
106
系统总线
113
图形处理
单元 130
I/O适配器
107
显示适配器
112
显示屏
115
RAM
114
ROM
102
CPU 101c
CPU 101b
CPU 101a
用户接口适配器
108
扬声器
111
鼠标
110
键盘
109

图3

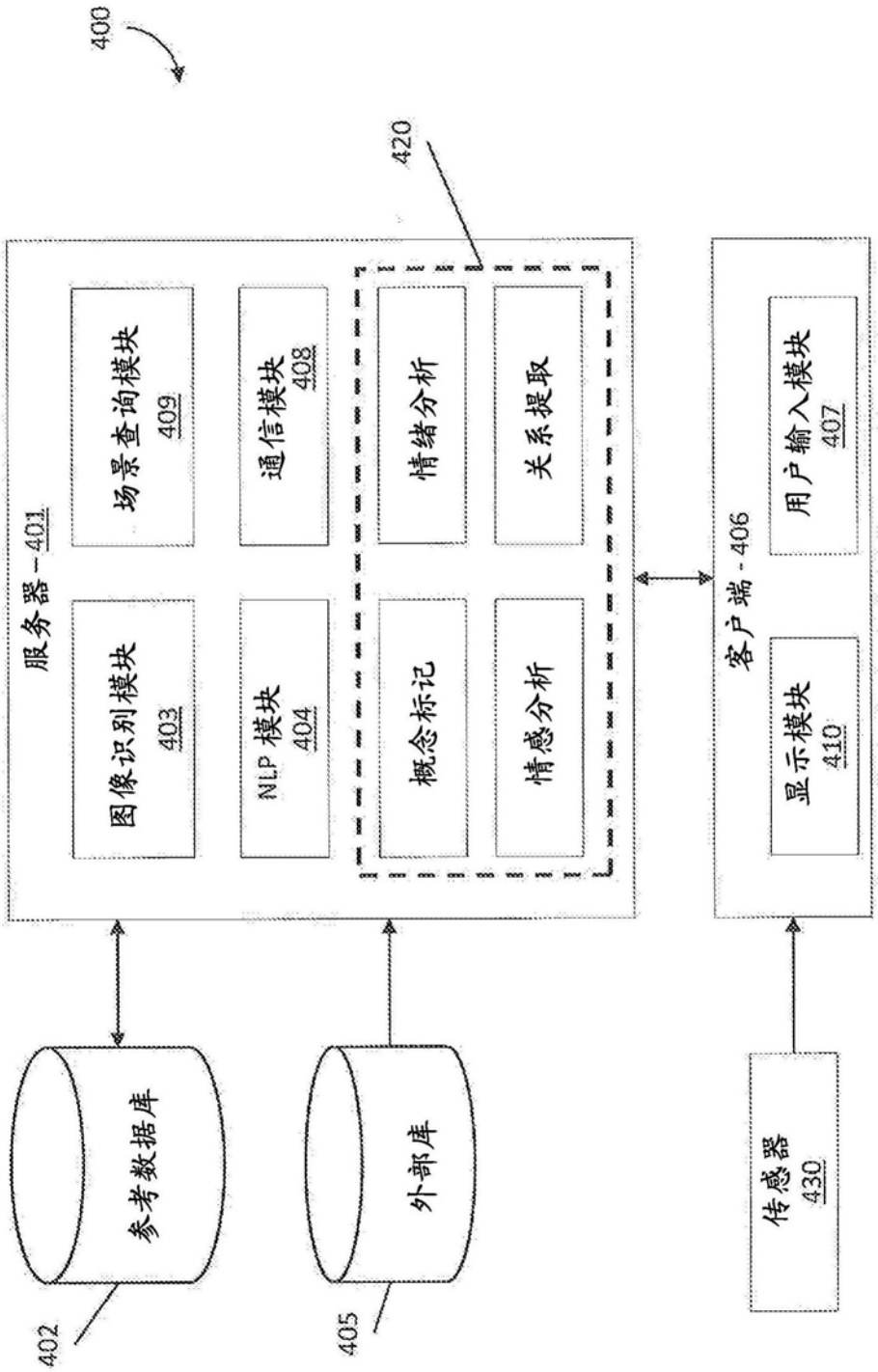

400
420
服务器 401
场景查询模块 409
通信模块 408
情绪分析
关系提取
图像识别模块 403
NLP 模块 404
概念标记
情感分析
客户端 406
用户输入模块 407
显示模块 410
参考数据库
外部库
传感器 430
402
405

图4

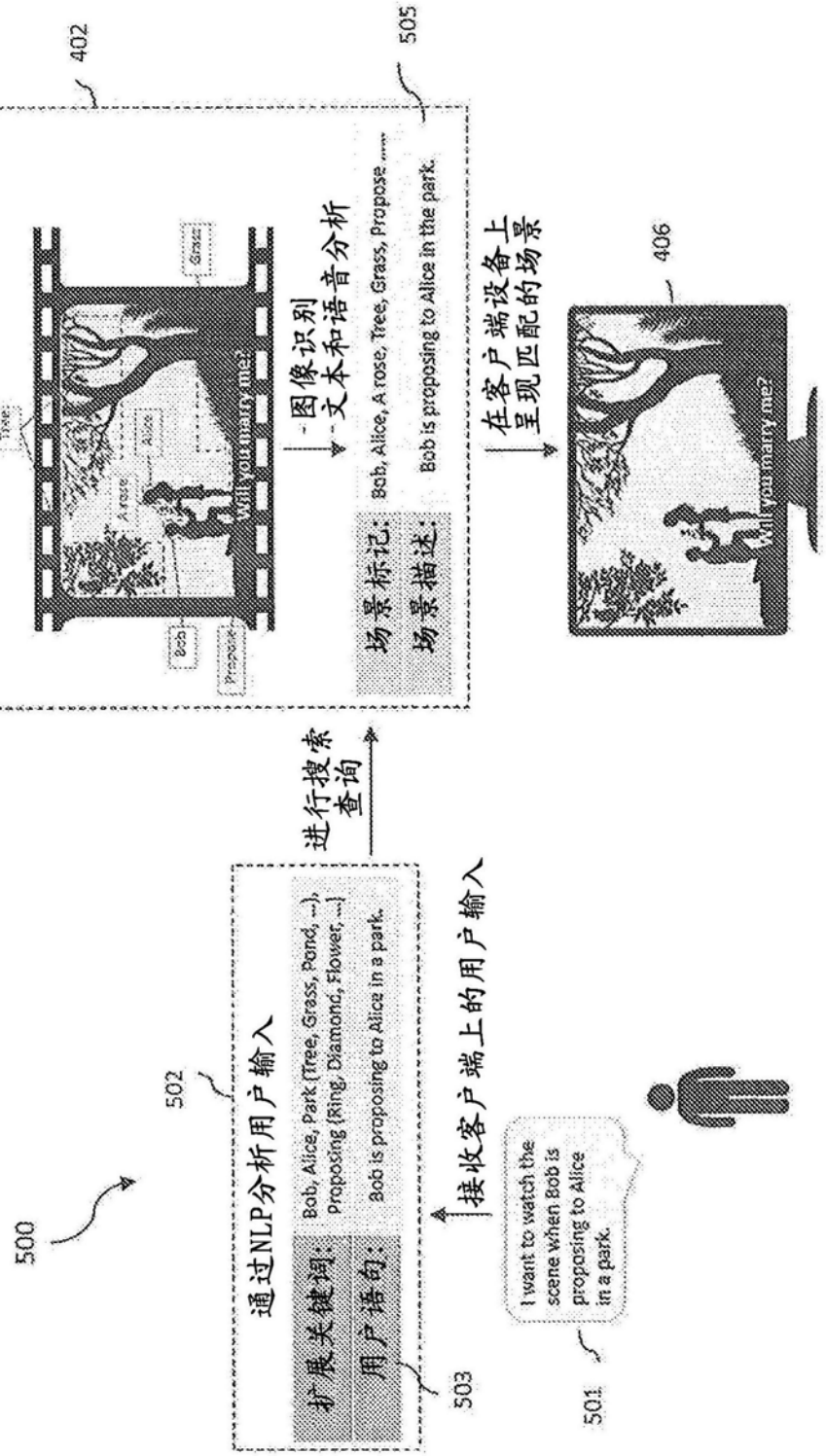

500
502
通过NLP分析用户输入
扩展关键词:
Bob, Alice, Park (Tree, Grass, Pond, ...), proposing (Ring, Diamond, Flower, ...)
用户语句:
Bob is proposing to Alice in a park.
503
进行搜索查询
接收客户端上的用户输入
I want to watch the scene when Bob is proposing to Alice in a park.
501
视频场景参考数据库
402
Will you marry me?
图像识别
文本和语音分析
505
场景标记:
Bob, Alice, A rose, Tree, Grass, Propose .....
场景描述:
Bob is proposing to Alice in the park.
在客户端设备上呈现匹配的场景
406
Will you marry me?

图5

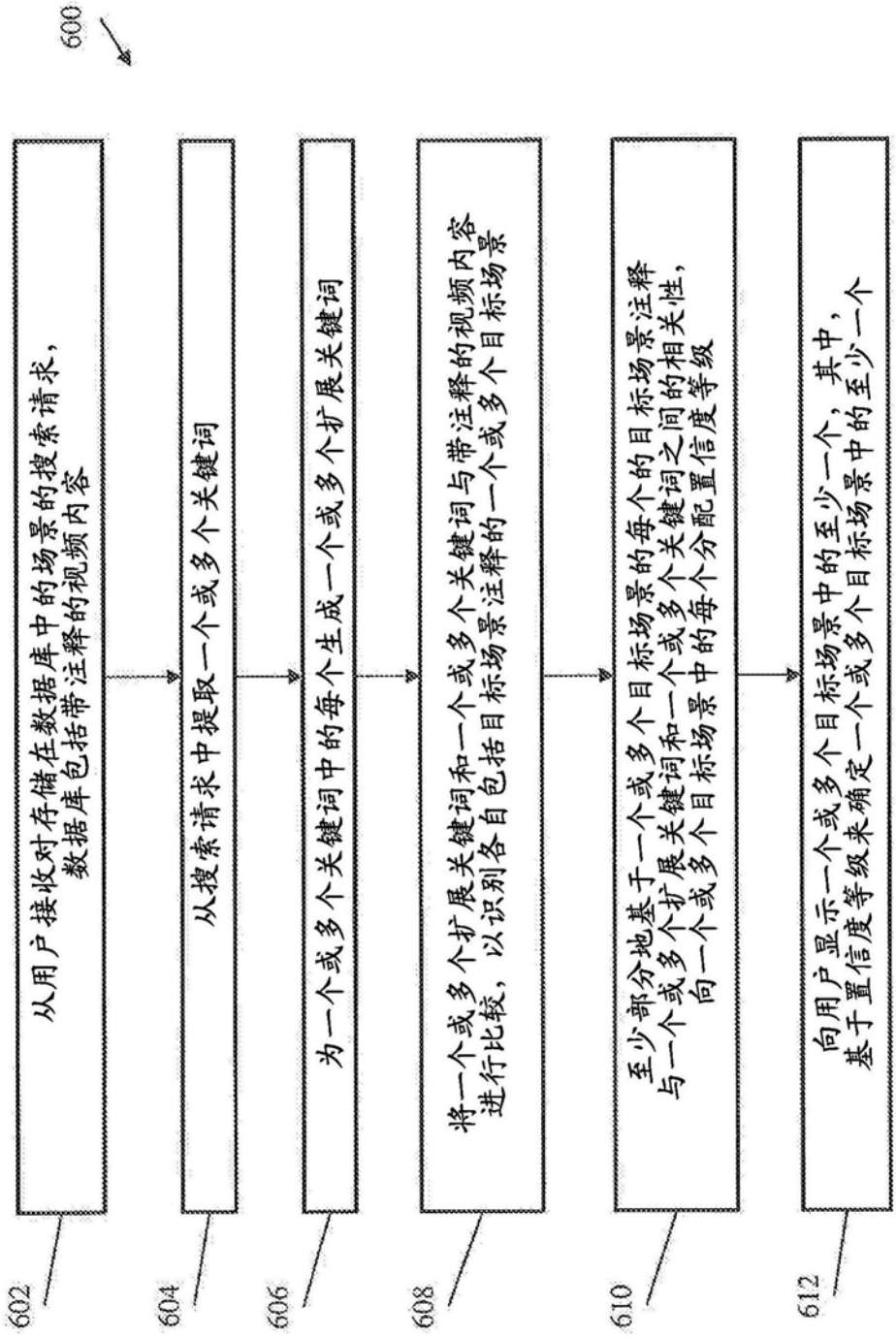

600
从用户接收对存储在数据库中的场景的搜索请求，数据库包括带注释的视频内容
602
从搜索请求中提取一个或多个关键词
604
为一个或多个关键词中的每个生成一个或多个扩展关键词
606
将一个或多个扩展关键词和一个或多个关键词与带注释的视频内容进行比较，以识别各自包括目标场景注释的一个或多个目标场景
608
至少部分地基于一个或多个目标场景的每个的目标场景注释与一个或多个扩展关键词和一个或多个关键词之间的相关性，向一个或多个目标场景中的每个分配置信度等级
610
向用户显示一个或多个目标场景中的至少一个，其中，基于置信度等级来确定一个或多个目标场景中的至少一个
612

图6

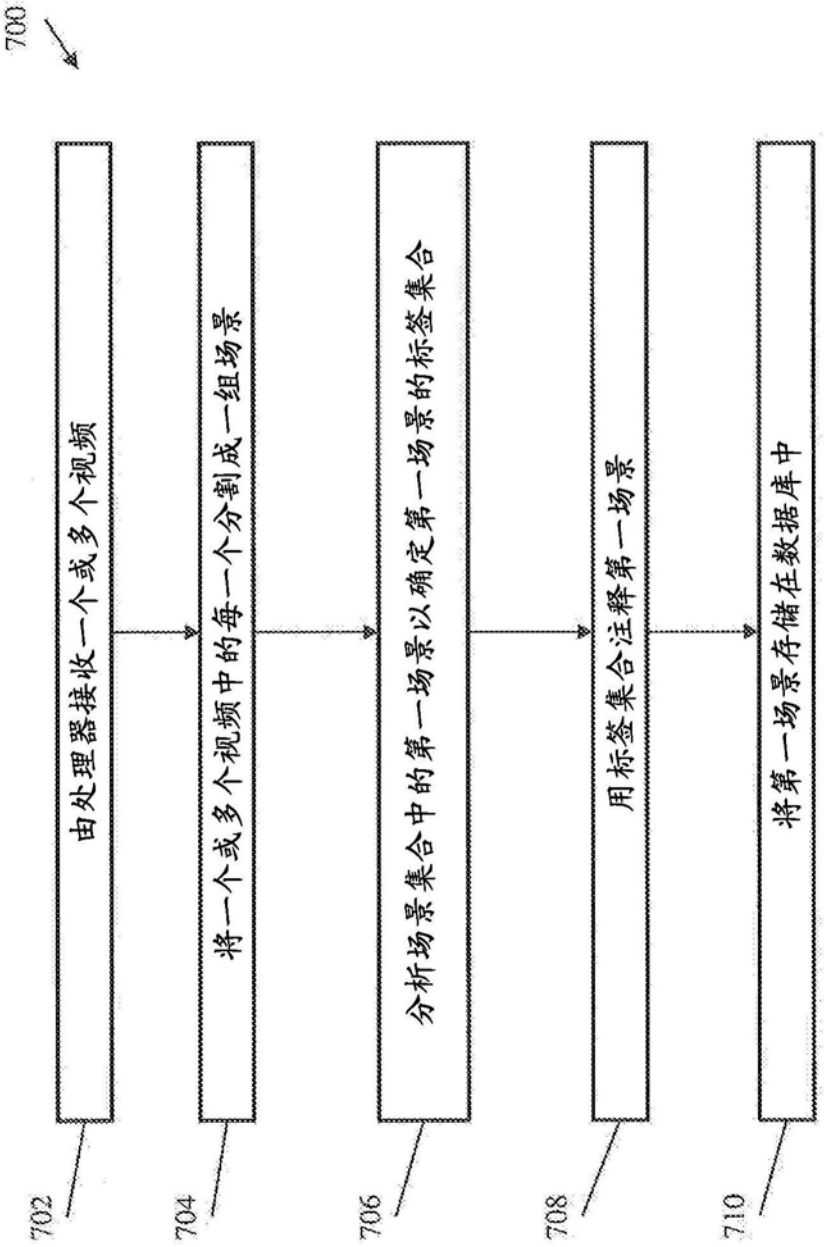

700
由处理器接收一个或多个视频
将一个或多个视频中的每一个分割成一组场景
分析场景集合中的第一场景以确定第一场景的标签集合
用标签集合注释第一场景
将第一场景存储在数据库中
702
704
706
708
710

图7