(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 109906455 A
(43)申请公布日 2019.06.18

(21)申请号 201780057452.5

(22)申请日 2017.09.07

(30)优先权数据
62/384,855 2016.09.08 US

(85)PCT国际申请进入国家阶段日
2019.03.22

(86)PCT国际申请的申请数据
PCT/SG2017/050449 2017.09.07

(87)PCT国际申请的公布数据
WO2018/048355 EN 2018.03.15

(71)申请人 AIQ私人股份有限公司
地址 新加坡爱贝施@亨德森,201亨德森路,#02-09

(72)发明人 斯蒂芬·莫里斯·摩尔
拉里·帕特里克·默里
拉扎林格帕·山姆格曼尼

(74)专利代理机构 深圳市世纪恒程知识产权代理事务所 44287
代理人 胡海国

(51)Int.Cl.
G06K 9/00(2006.01)
G06F 16/783(2019.01)

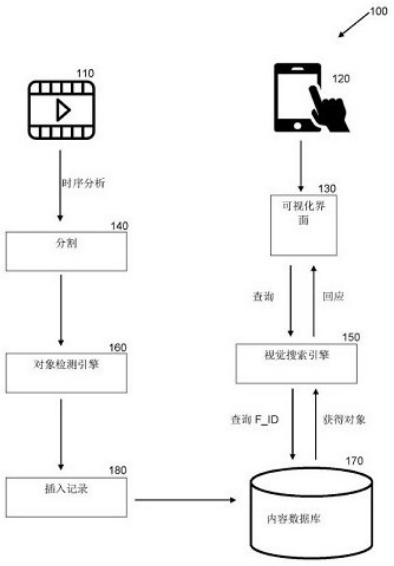
权利要求书2页 说明书13页 附图6页

(54)发明名称

视觉搜索查询中的对象检测

(57)摘要

本发明包括用已知对象填充数据库的系统和方法。该数据库可以用离线数据扩增(例如,网络爬虫)来填充,或者通过调整已知对象和元数据集群与定义的内容来填充。观众可以从实时或离线媒体查询图像。观众查询中的对象与数据库中的相似对象或推荐产品链接。



CN 109906455 A

1.一种检测视频中的对象并将该对象与一个或多个产品相匹配的方法,包括以下步骤:

a)获取视频;

b)基于所描述的设置和/或事件,通过比较连续帧的内容的相似性和差异性以分割视频;

c)编译相同或相似设置和/或事件的片段;

d)分析一个或多个片段以检测一个或多个对象;

e)将所述一个或多个对象与产品进行比较;

f)识别与所述一个或多个对象相关联的产品;

g)向一个或者多个观众通知所述产品。

2.根据权利要求1所述的方法,其中,使用一个卷积神经网络(CNN)来识别与所述一个或多个对象相关联的产品。

3.根据权利要求1所述的方法,其中,分析一个或多个片段以检测一个或多个对象的步骤包括将帧和/或帧的部分与数据库中定义的内容进行比较。

4.根据权利要求3所述的方法,其中,利用网络爬虫将定义的内容填充到所述的数据库。

5.根据权利要求3所述的方法,其中,通过调整已知对象和元数据集群以将定义的内容填充到所述的数据库。

6.根据权利要求1所述的方法,其中,第二屏幕内容增强用于实况或流视频。

7.根据权利要求1所述的方法,其中,向一个或者多个观众通知所述产品的步骤包括显示广告。

8.根据权利要求1所述的方法,其中,向一个或者多个观众通知所述产品的步骤包括提供超链接至网站或者视频。

9.一种检测屏幕截图中的一个或多个对象并将所述的一个或多个对象与宣传资料相匹配的方法,包括以下步骤:

a)接收来自观众的数字屏幕截图形式的询问;

b)识别所述屏幕截图中的一个或多个对象;

c)将所述一个或多个对象与产品进行比较;

d)匹配与所述一个或多个对象相关联的产品;

e)将与匹配产品相关的宣传资料推送给观众。

10.根据权利要求9所述的方法,其中,在匹配与所述一个或多个对象相关联的产品的步骤中使用卷积神经网络(CNN)。

11.根据权利要求9所述的方法,其中,识别屏幕截图中的一个或多个对象的步骤包括将屏幕截图和/或屏幕截图的部分与数据库中的定义内容进行比较。

12.根据权利要求11所述的方法,其中,利用网络爬虫将定义的内容填充到所述的数据库。

13.根据权利要求11所述的方法,其中,通过调整已知对象和元数据集群以将定义的内容填充到所述的数据库。

14.根据权利要求9所述的方法,其中,第二屏幕内容增强用于实况或流视频。

15.根据权利要求9所述的方法，其中，将与匹配产品相关的宣传资料推送给观众的步骤包括：显示广告和/或提供超链接至网站或视频。

16.一种系统，用于在视频中的对象与产品数据库中的产品之间产生关联，包括：

计算机化网络和系统，通过用户界面应用程序本地或远程地公开给用户或用户组；

用于在本地或服务器上检测和存储媒体内容的模块；

用于传输媒体内容至远程或基于服务器的处理器，以获取元数据和/或视觉特征的模块；

用于传输媒体内容至远程或基于服务器的处理器，以提取元数据和/或视觉特征的模块；

用于接收一个或多个用户的输入的装置，所述输入是包括视觉特征的数字图像；

用于识别所述视觉特征，并将所述视觉特征与数据库中的对象和/或相关产品组相关联的模块；和

分发与所述对象和/或相关产品组相关的信息给用户和/或用户组的网络服务。

17.根据权利要求16所述的系统，其中卷积神经网络(CNN)被用于分析视觉特征和元数据，以将视觉特征与对象和/或相关产品组相关联。

18.根据权利要求16所述的系统，其中与所述对象和/或相关产品组相关的信息包括广告。

19.根据权利要求16所述的系统，其中与所述对象和/或相关产品组有关的信息包括因特网可访问的超链接或内容。

视觉搜索查询中的对象检测

技术领域

[0001] 本发明涉及用于互联网营销的计算机技术，更具体地说，涉及一种联网的计算机化应用，用于将视频分成片段，识别片段中的对象，并将产品与对象相匹配。

背景技术

[0002] 电子商务是网上买卖的交易。电子商务已经成为全球企业的一个重要工具，不仅是向客户销售，也让客户参与进来。2012年，全球电子商务销售额超过1万亿美元。

[0003] 互联网营销指的是通过电子商务的方式利用网络和电子邮件以推动销售的广告和营销工作。它包括电子邮件营销、搜索引擎营销(SEM)、社交媒体营销、许多类型的展示广告(例如横幅广告)和移动广告。元数据是网络营销的重要组成部分。

[0004] 因为网上购物者无法亲自查看商品，他们通常用关键字等标准进行搜索。商家经常收集关于搜索和交易的元数据，这使他们能够分析销售趋势，制定营销计划并做出预测。同样的该元数据使企业能提供更个性化的购物体验，包括购买历史、多个发货地点的地址簿和产品推荐。

[0005] 如今，大多数网页都嵌入了元数据。网站搜索引擎建立了大量索引，这些索引使用页面文本及其附带的元数据以向用户提供相关的搜索结果。元数据可用于定向广告。根据其正在推广的产品，广告商可以使用复杂的方法利用特定的特征来锁定最容易接受的受众。

[0006] 例如，一个人可以使用网络浏览器搜索飞往新西兰的航班。元数据以来自他/她访问的网站的“cookies”形式通过用户的浏览器存储在计算机上。Cookies在互联网服务器和浏览器之间来回发送，从而识别用户和/或追踪他/她的活动。此后，该人可接收到例如与新西兰旅行相关的横幅广告，例如酒店、租车、旅游和航班信息。

[0007] 此外，基于人群统计数据，元数据可用来识别和锁定用户。商家可得知某一产品对某一人群有吸引力，并可直接向该人群推销。例如，投资证券的横幅广告可能对青少年无效。将该广告的对象锁定为年纪大一些的人群，尤其是那些寻求退休的人群，会更有效。如果元数据表明他/她不在目标人群中，用户将不会收到广告。

[0008] 随着市场的出现，展现出越来越多的可视化搜索用例，受众统计分析的丰富对消费者和公司都极有利。组织可使用这种模型来提供高度个性化的产品，评估使用模式(例如基于季节性)，甚至涵盖未来产品方向。

[0009] 基于视觉的搜索很常见，尤其是随着智能手机和平板电脑的普及。例如，上述例子中的用户可能会搜寻与新西兰徒步旅行相关的图片。因为他/她点击图像，浏览器无法基于键入浏览器的关键字记录元数据。同样，他/她可能会观看与新西兰景点相关的视频。使用传统技术，这些图像和视频都不会为定向营销提供元数据。

[0010] 然而，增加视频流和下载也带来了新的机会。例如，一位名人可能会出现在社交媒体上的视频中，提着一个特定的手提包。如果手提包可以被识别并可供购买，那么这个手提包的销量就会激增。

[0011] 鉴于各种架构的互联网视频分布平台的激增和普及，对视频内容的视觉搜索可能挖掘出由内容创作者、消费者和商业伙伴组成的数百万用户群。如果能够识别出视频的特定片段，有关方就能够用额外的内容扩充和/或拼合这些片段。这可以采取对这些片段进行信息丰富的形式。商业合作伙伴可能希望锁定相关部分作为分销产品的渠道。

[0012] 此外，营销人员正在寻求更好的方法来吸引习惯于跳过广告和使用点播媒体的年轻受众。产品放置和品牌娱乐提供了“全方位”的可能性，以更有效地吸引年轻和/或精通技术的消费者。

[0013] 这给广告商带来了一个问题。观众目前无法传递他/她对在节目或场景中所观看到的对象或产品的兴趣。例如，观众可以看到一个提着手提包的名人。然而，在哪里购买手提包可能并不清楚。观众不会搜索该手提包并会逐渐失去兴趣。人们已经尝试将视频链接和/或打印至网站。

[0014] 使用当前技术，销售商或广告商可以在印刷品或视频广告中添加QR码(快速反应码)。观众可以使用智能手机扫描QR码，这将引导他/她浏览网站和/或网页内容。然而，这需要将一个明显的代码块放置于该观众跟前。此外，必须为观众感兴趣的每个对象设置单独的QR码。就一个视频来说，QR码必须在其整个播放时间内都显示。

[0015] 因此，需要一种实现观看者表达其对某对象的兴趣和/或获得与图像和/或视频中的对象相关的附加信息的方法。该系统应使得用户可获得关于目标物的细节和更深入的信息，而无需进行关键字搜索或扫描QR码。它应该能够与印刷媒体(例如杂志广告)以及视频媒体(例如电视)一起使用。

[0016] 介绍

[0017] 本发明包括一种检测视频中的对象并将该对象与一个或多个产品相匹配的方法，该方法包括以下步骤：(a)获得视频，(b)基于所描述的设置和/或事件，通过比较连续帧的内容的相似性和差异性以分割视频；(c)编译相同或相似设置和/或事件的片段；(d)分析视频的一个或多个帧，以检测来自相同或相似设置和/或事件的每个片段的一个或多个对象，(e)将一个或多个对象与数据库中的产品进行比较，(f)识别与一个或多个对象相关联的产品，以及(g)向一个或多个观众通知所述产品。

[0018] 该方法可以使用卷积神经网络(CNN)来识别与一个或多个对象相关联的产品。数据库可以用离线数据扩增(网络爬虫)来填充，和/或通过调整已知对象和元数据集群与定义的内容来填充。第二屏幕内容增强可用于实况或流视频。通知一个或多个观众或者所述产品的步骤可以包括显示广告和/或通过提供超链接至网站或视频来通知一个或多个观众或产品。

[0019] 本发明还包括一种检测屏幕截图中的一个或多个对象并将该一个或多个对象与宣传材料匹配的方法，该方法包括以下步骤：接收来自观众的数字图像或者屏幕截图形式的询问；(b)识别屏幕截图中的一个或多个物品，(c)将该一个或多个物品与数据库中的产品进行比较，(d)识别与该一个或多个物品相关联的产品，以及(e)向观众推送与所识别的产品相关的宣传材料。

[0020] 数据库可以用离线数据扩增(网络爬虫)来填充，和/或通过调整已知对象和元数据集群与定义的内容以将产品填充至数据库。第二屏幕内容增强可用于实况或流视频。向观众推送宣传材料的步骤可以包括显示广告和/或提供超链接至网站或视频。

[0021] 本发明还包括一种系统，用于在视频中的对象与产品数据库中的产品之间产生关联，并分布关于产品的信息，该系统包括(a)计算机化网络和系统，该网络和系统将通过诸如移动、浏览器或任何类似的计算机化系统的用户界面应用程序本地或远程公开给用户或用户组，(b)用于在本地或服务器上检测和存储媒体内容的模块，(c)用于传输媒体内容至远程或基于服务器的处理器，以获取元数据和/或视觉特征的模块，(d)用于传输媒体内容至远程或基于服务器的处理器，以提取元数据和/或视觉特征的模块，(e)用于接收一个或多个观众的输入的装置，所述输入是包括视觉特征的数字图像形式，(f)用于识别所述视觉特征，并将所述视觉特征与对象和/或相关产品组相关联的模块，和(d)分发与所述对象和/或相关产品组相关的信息给用户和/或用户组的网络服务。

[0022] 一个卷积神经网络(CNN)可以用来分析视觉特征和元数据，以将视觉特征与对象和/或相关产品组相关联。数据库可以利用离线数据扩增(网络爬虫)来填充，和/或通过调整已知对象和元数据集群与定义的内容以将已知的对象填充数据库。相关产品的相关信息包括广告和/或通过互联网访问内容的超链接。

发明内容

[0023] 本发明的第一方面是提供一种系统，用于在用户视觉查询和从对象数据库语料库中检测到的对象之间生成关系。

[0024] 本发明的第二方面是一种计算机化网络和系统，该网络和系统将通过诸如移动、浏览器或任何类似的计算机化系统的用户界面应用程序本地或远程公开给用户或用户组，

[0025] 本发明的第三方面是一种用于在本地或服务器上检测和存储媒体内容的模块。

[0026] 本发明的第四方面是一种模块，用于传输媒体内容至远程或基于服务器的处理器，以输入和提取元数据和/或视觉特征。

[0027] 本发明的第五方面是用于分析视觉特征和元数据以与特定对象和/或对象组相关联的计算机模型。

[0028] 本发明的第六方面是本地或服务器端托管的模块，用于将检测到的对象链接到相关对象的组。

[0029] 本发明的第七方面是本地或服务器端托管的模块，用于调整已知对象和元数据集群与预定义内容。

[0030] 本发明的第八方面是向用户和/或用户组分发内容的网络服务。

附图说明

[0031] 图1描绘了本发明一个实施例的整个流程图。

[0032] 图2描绘了视频场景分割。

[0033] 图3描绘了视觉搜索概述。

[0034] 图4描绘了离线产品数据扩充。

[0035] 图5描绘了离线预摄取内容的产品推荐框架。

[0036] 图6描绘了用户查询和产品推荐。

具体实施方式

[0037] 定义

[0038] 本说明书中对“一个实施例/方面”或“实施例/方面”的引用意味着结合该实施例/方面描述的特定特征、结构或特性被包括在公开的至少一个实施例/方面中。说明书中的各处使用短语“在一个实施例/方面”或“在另一个实施例/方面”不一定都指相同的实施例/方面，也不一定是与其他实施例/方面互斥的单独的或替代的实施例/方面。此外，描述了可以由一些实施例/方面而不是其他实施例/方面展示的各种特征。类似地，描述了各种要求，这些要求可能是针对一些实施例/方面的要求，但不是针对其他实施例/方面的要求。实施例和方面在某些情况下可以互换使用。

[0039] 在本说明书中使用的术语在本领域、在本公开的上下文中以及在使用每个术语的特定上下文中通常具有它们的普通含义。用于描述本公开的某些术语将在下文，或说明书中的其他地方讨论，以便就本公开的描述向实践者提供额外的指导。为方便起见，某些术语可能会突出显示，例如使用斜体和/或引号。突出显示的使用对术语的范围和意义没有影响；一个术语的范围和含义，在相同的上下文中，不管它是否被突出显示，是相同的。可以理解，同样的事物可以用多于一种的方式表达。

[0040] 因此，可替换的用语和同义词可用于本文讨论的任何一个或多个术语。对于是否在本文中详细阐述或讨论术语也没有任何特殊意义。提供了某些术语的同义词。对一个或多个同义词的列举并不排除使用其他同义词。本说明书中任何地方实施例的使用，包括这里讨论的任何术语的例子，仅仅是说明性的，并不旨在进一步限制本公开或任何示例性术语的范围和含义。同样，本公开不限于本说明书中给出的各种实施例。

[0041] 在不进一步限制本公开的范围的情形下，下面给出根据本公开的实施例的仪器、装置、方法及其相关结果的示例。注意，为了方便读者，在示例中可使用标题或子标题，这些使用绝不应该限制本公开的范围。除非另有定义，本文使用的所有技术和科学术语与本公开内容所属领域的普通技术人员的通常理解具有相同含义。一旦发生冲突，本文件，包括定义，将会受到控制。

[0042] 术语“手机应用程式”或“手机应用程式”是指为实现特定目的而设计的，尤其是下载到移动设备上时独立程序或软件。

[0043] 术语“单词包”或“BoW模型”是指通过将图像特征视为单词来进行图像分类。在文档分类中，单词包是单词出现次数的稀疏向量，即词汇上的稀疏直方图。在计算机视觉中，视觉单词包是局部图像特征词汇的出现计数向量。

[0044] 术语“cookie”、“互联网cookie”或“HTTP cookie”是指由用户的网络浏览器从网站发送并存储在用户计算机上的一小块数据。Cookies在互联网服务器和浏览器之间来回发送，这使得用户可以被识别或跟踪他/她的进展。Cookies提供了关于消费者访问哪些页面、查看每个页面花费的时间、点击的链接、搜索和互动的详细信息。从这些信息中，cookie发行者收集了对用户浏览趋势和兴趣的理解，从而生成了一个简档。分析简档，广告商能够基于具有相似反馈信息的用户来创建定义受众群，从而建档。

[0045] 术语“集群”或“集群分析”是指对一组对象进行分组的任务，使得同一组(称为集群)中的对象比其他组(集群)中的对象更为相似(在某种意义上)。这是探索性数据挖掘的一项主要任务，也是统计数据分析的一项常用技术，应用于许多领域，包括机器学习、模式

识别、图像分析、信息检索、生物信息学、数据压缩和计算机图形学等。

[0046] 术语“数据增加”是指增加数据点的数量。就图像而言，这可能意味着增加数据集中的图像数量。就传统的行/列格式数据而言，这意味着增加行或对象的数量。

[0047] 术语“深度学习”是指对于包含一个以上的隐藏层的人工神经网络在学习任务中的应用。深度学习是基于学习数据表示的广义上的机器学习方法的一部分，而不是赋任务于特定算法。

[0048] 在模式识别和机器学习中，术语“特征向量”指的是表示某个对象的数字特征的n维向量。机器学习中的许多算法需要对象的数字表达，因为这种表达便于处理和统计分析。当表达图像时，特征值可能对应于图像的像素，当表达文本时，可能是术语出现频率。

[0049] 术语“不平衡数据集”是指分类问题的一种特殊情况，在这种情况下，类别之间的分布不一致。通常，它们由两类组成:多数(消极)类和少数(积极)类。为了使数据成为可用的形式，可能需要进行类平衡。

[0050] 术语“倒排索引”、“张贴文件”或“倒排文件”是一种索引数据结构，用于存储从内容，如单词或数字，到其在数据库文件，或文档或一组文档中的位置的映射(与从文档映射到内容的前向索引相反)。倒排索引以在将文档添加到数据库中时会增加处理量为代价，其目的是允许快速全文搜索。

[0051] 术语“k-最近邻”或“k-NN”是指最近邻分类对象，其中距离度量(“最近”)和邻居数量均可改变。该对象使用预测方法对新的观测进行分类。该对象包含用于训练的数据，因此可以计算重新替换预测。

[0052] 链分析

[0053] 术语“模块”指的是一个独立的单元，例如电子元件和相关线路的组件或一段计算机软件，它本身执行确定的任务，并且可以与其他这样的单元链接以形成更大的系统。

[0054] 术语“多层感知神经网络”或“MLP”是指在输入和输出层之间具有的一层或多层前馈神经网络。前馈意味着数据在一个方向上从输入层流向输出层(正向)。MLPs广泛用于模式分类、识别、预测和近似。多层感知器可以解决不可线性分离的问题。

[0055] 术语“元数据”是指描述其他数据的数据。它提供关于某项内容的信息。图像可以包含描述图像大小、颜色深度、图像分辨率以及图像创建时间的元数据。文本文档的元数据可包含关于文档长度、作者、文档书写时间以及文档的简短摘要的信息。

[0056] 术语“元标签”是指包含在网页上的元数据。描述和关键字元标签通常用于描述网页的内容。大多数搜索引擎在向搜索索引中添加页面时都会使用这些数据。

[0057] 术语“QR码”或“快速反应码”指的是包含关于其所附物品的信息的矩阵条形码(或二维条形码)。QR码包括排列在白色背景上的正方形网格中的黑色正方形，可以由像照相机这样的成像设备读取，并使用里德-所罗门误差校正进行处理，直到该QR码被正确理解。然后从图像的水平和垂直分量中存在的图案中提取所需的数据。

[0058] 术语“合成数据”是指适用于给定情况的不是通过直接测量获得的任何生产数据。

[0059] 术语“支持向量机”或“SVM”是指具有相关学习算法的监督学习模型，该模型分析用于分类和回归分析的数据。给定一组训练示例，每个示例都被标记为属于两个类别中的一个或另一个类别，SVM训练算法构建一个模型，将新示例分配给一个或另一个类别，使其成为非概率二元线性分类器。

[0060] 术语“定向广告”指的是一种广告形式，在线广告商可以使用精确的方法，根据广告商推销的产品或人员，以某些特征来瞄准最容易接受的受众。这些特征可以是关注以种族、经济状况、性别、年龄、教育水平、收入水平和就业人口统计学特征，也可以是关注于消费者价值观、个性、态度、观点、生活方式和兴趣的心理学特征。它们也可是行为变量，例如浏览器历史、购买历史和其他近期活动。

[0061] 在图像检索系统中使用的术语“视觉单词”或“视觉单词簇”是指图像的小部分，这一小部分携带与特征相关的某种信息(例如颜色、形状或质地)，或者携带像素中发生的变化，例如过滤、低级特征描述符(SIFT、SURF，...等)。

[0062] 术语“白化变换”或“球化变换”是指线性变换，其将具有已知协方差矩阵的随机变量向量变换成一组协方差是同一性矩阵的新变量，这意味着它们是不相关的，并且都具有方差1。这种变换被称为“白化”，因为它将输入向量改变为白噪声向量。

[0063] 这里使用的其他技术术语在本领域中具有它们所使用的普通含义，例如各种技术词典。

[0064] 优选实施例的描述

[0065] 在这些非限制性示例中讨论的特定值和配置可以变化，并且被引用仅仅是为了说明至少一个实施例，而不是为了限制其范围。

[0066] 视觉搜索(相对于传统的基于文本的搜索)对于人群分析的主要好处之一是可以获得关于查询的更多信息，可以确定该查询。例如，用户可以在搜索引擎(或电子商务网站)中搜索棕色鞋子。然后，用户可以选择购买或询问一种非常特殊的棕色鞋子(休闲鞋、鞋带等)。

[0067] 仅访问文本搜索查询，如果没有任何进一步的信息，就不可能提取关于搜索对象的更大粒度。然而，通过视觉搜索用例，查询本身可以告诉我们更多关于用户查询的性质。

[0068] 为了提取关于视觉搜索查询的元数据，可以使用高级分类算法，包括但不限于深度学习、监督学习和非监督学习。因此，从输入图像中，可以获得描述性元数据的列表(例如鞋、棕色、鞋带、布洛克、上下文、生产地、材料以及任何提供了图像中内容状态清晰度的信息)。

[0069] 在这里描述的本发明的一个实施例中，可以从帧中提取组成图像的对象列表，该列表链接到被分析为对应于语义上不同的“主题”的帧序列。

[0070] 图1描绘了本发明的整个过程流。来自视频110的内容被收集和编译以建立内容数据库170。视频的观看者可以通过查询120，如向系统提交来自视频的截屏的，来访问内容数据库170。

[0071] 处理视频文件110以自动确定并提取一帧或语义相似的帧组中的对象的元数据和属性。时序分析(如下所述)包括视频的分割140。在对象检测引擎160中，可以针对对象分析关键帧。创建插入记录180以说明对象的时间位置。识别的对象和识别信息被添加到内容数据库170。

[0072] 用户可以通过使用视觉接口130拍摄视频图像来查询帧120。该系统包括可以访问内容数据库170的视觉搜索引擎150。

[0073] 时序分析

[0074] 涉及时序分析200的步骤如图2所示。2.将视频110引入到框架中，并且分析时间帧

序列以分割媒体内容。这种分割的目的是识别、隔离和标记帧序列，使得每个片段对应于单个事件或主题。此后，可以分析片段中的对象。

[0075] 按顺序遍历210视频110以检测超出相似性阈值的帧对或序列。也就是说，帧比较220可以指示帧组成的显著变化，这意味着场景或主题的变化。

[0076] 描述相同事件/场景的帧不会超出相似性阈值。在这种情况下，评估250下一帧。描绘不同事件/场景的帧通常会超出相似性阈值，在这种情况下，识别240该片段为事件/场景。这个过程可以重复，使得视频110中的每个帧都包含在一个片段中。此后，对象检测引擎160可以分析视频片段中的对象。

[0077] 当处理表现出时间多态性的对象时(即在帧间改变其形状)，分割变得更加重要。一旦完成识别视频中的场景或帧的排序，就有可能链接关键帧的对象，该对象可能改成成了未知的形状。另一个实施例是用特定物体的所有已知形变的例子来训练物体检测模型。

[0078] 元数据可以通过片段标识符链接到帧本身的可视内容，并且数据可以被摄取到并行视觉搜索数据库中。当用户查询摄取帧的图像时，它会被发送到对应的服务器。片段标识符可用于识别与帧相关联的帧序列。该信息用于检索对象列表和任何链接的扩充内容，为了将上述对象传递回用户120。请注意，查询框架本身不会针对对象进行分析。相反，将它与一个片段标识符匹配，该片段标识符链接到该片段的一个预先分析过的对象语料库。

[0079] 对象检测

[0080] 可以分析通过时序分析识别的关键帧，以识别和确定各个对象。在一个实施例中，可以使用由多层组成的深度卷积神经网络(CNN)来执行这个任务。

[0081] 图3描绘了如何利用基于“视觉单词包”方法的视觉搜索引擎从视觉查询图像中搜索片段标识符。搜索引擎使用图像数据库385。训练图像310(即具有已知特征的图像)与定义的对象一起提交。图像用于生成特征320。训练图像365可以被插入到数据库375中。

[0082] 为了训练CNN模型，有必要为每个要摄取的对象引入大量的图像示例。为具有多个属性的多个对象及其位置组织大量的图像数据。该数据进一步经过白化转换、数据扩充和类平衡。这个对象的数据库被用作训练深层卷积网络的输入。

[0083] 可以训练模型，以便一起学习对象及其属性。此外，该模型提供了隐藏层的值，这是各种抽象图像的实值向量描述。模型的推断可以提供对象标签、置信度得分、隐藏层向量和属性。

[0084] 例如，对象可以是服装领域中的一项(如包、牛仔裤等)以及它们的属性(如颜色、图案、长度等)中。此外，在多个区域上对图像的推断是物体的位置，可以用来确定图像在每个关键帧之前或之后的帧中最可能的位置。

[0085] 一旦该模型被评估为达到成功训练的状态(根据测试集的最小误差度量进行评估)，它就会被传播到框架内的实时摄取模块。在视频片段的时序分析之后，分析每个关键帧以获得图像中存在的对象的位置。在该帧中检测到的每个对象都可以在该段内的两个方向上短暂跟踪。对于在片段中找到的每个位置，可推断标签、属性和内层，并通过置信度度量进行时间加权平均。这最终为每个片段和/或帧生成对象特征向量。隐藏层生成的向量描述通常包含高维度。可以收集大量图像，以便为特定的分布获取压缩技术。在这里描述的框架的示例性实施例中，深度自动编码器提供了具有最低的搜索精度损失的最佳压缩。

[0086] 视觉搜索

[0087] 当用户希望与特定视频交互时，该/她可以通过用他们的移动设备相机从通常在用户的移动设备上操作的计算机程序中捕捉帧来表示他们的兴趣，该计算机程序被称为“应用程序”该应用程序可以上传查询图像以供进一步处理。

[0088] 在视觉查询图像从用户设备发送到服务器后，它被用来搜索已知的图像数据库以识别可能的匹配。然后，排名靠前的结果被用于进一步识别与问题中的帧最想关联的片段。此后，片段标识符可用于检索与特定片段链接的对象。在查询响应被聚合之后，所有扩充的内容被转回用户的移动设备。

[0089] 例如，观看者可以提交他/她在视频中注意到的手提包的查询。手提包可以根据几个标准与数据库中的图像匹配，包括形状、图案、品牌、形状、尺寸、品牌和其他细节。系统可以返回许多已排序的匹配图像。

[0090] 这个过程如图3所示。其中用户查询图像120。视觉单词簇被分配340。从视觉单词簇中，填充一个反排文件列表。查询350并存储355反排文件列表。过滤最佳候选图像360，最佳候选图像可以被添加到图像数据库385。可以对最佳匹配380进行空间验证，并且最佳匹配可以返回给观众390。

[0091] 离线数据扩充

[0092] 未定义的图像，例如网络爬虫获得的图像，也可以用于填充内容数据库。图4描绘了离线产品数据扩充400。在示例性框架中，离线数据扩充用于填充对象和相关元数据的数据库170。

[0093] 爬虫过程可用于从各种在线来源(例如来自电子商务平台的产品列表或来自社交媒体网络的图像)检索图像及其注释元数据420。这些爬行图像和元数据属性通过数据清理阶段430，数据清理阶段430将原始爬行数据转换成适于插入数据库440的记录。

[0094] 数据库插入格式可以将每个爬行图像记录链接到对象标识符。由此实现将从片段中检测到的对象链接到产品数据库中的对象，从而提供一个界面，为从视觉查询图像中识别的对象或对象组提供扩充数据。

[0095] 用例

[0096] 产品推荐

[0097] 本发明可用于为现有视觉媒体内容提供第二屏幕内容扩充服务。

[0098] 例如，流行的电视节目(或电影)可以被摄入到平台中，以分析可通过各媒体访问的产品组合。观众可以被提示或意识到在节目播出期间或之后可能发生的互动机制(或者在线视频传送的情况下是流式的)。

[0099] 任何用户对摄取视频中的帧的视觉查询都可以随后通过对象检测框架来丰富。这为内容创建者以及消费者提供了一个与所提供的产品/服务交互的独特平台。在框架内检测到的物体的范围可以包括框架内无生命的物体(服装、家具、旅行机会等)，或者扩展到检测到的实体或与片段内的实体相对应的实体(即演员、剧组人员等)。

[0100] 例如，广告商或零售商可以使用产品放置来在视频或电影中推广手提包。观众提交对包括手提包的场景的查询。匹配的产品(即手提包)的促销广告可以在用户的设备上播放。还可以向观众提供额外信息，包括购买说明。

[0101] 图5描述了处理这里描述的用例的示例性框架。视频110可通过离线摄取用于扩充内容数据库。视频经历镜头变化检测510。视频540的片段由关键帧560、对象定位590和跟踪

630识别。下一步是对象识别、属性标记和特征提取660，随后是时间平均690。

[0102] 视频摄取还可以包括帧采样550和图像处理特征570。数据库内容可以包括框架610、产品670和对象710。产品摄取(网络爬行、聚集和扩充)发生在640。

[0103] 用户可以查询帧120。通过将图像与数据库中相似的580排序来匹配图像。可以生成620段号以及视觉搜索排名650。匹配的产品680可以响应于用户的查询而被转发给用户。广告商和/或营销商希望推广的产品可以获得更高的视觉搜索排名650。

[0104] 直播电视第二屏

[0105] 除了离线、预先存在的内容之外，还可以扩展这个框架来处理实时视频流。这种情况下的挑战是确保传播到对象检测平台的每帧在同一帧的任何查询之前完成摄取机制。

[0106] 为了说明实时视频，时序分析模块可以被交易内容数据库取代，该数据库保留过去“N”分钟(或必要时数小时)的临时历史记录。在这个修改后的框架中，来自实时视频流的每一帧都被整合到对象检测和视觉搜索数据库中，并有一个“生存时间”机制来确保数据在可配置的延迟后过期。以这种方式，数据库(和计算集群)的大小被限制在能够在卷中维持低延迟操作的高性能状态。

[0107] 这种框架可用于为第二屏幕内容提供:

[0108] -现场体育赛事，展示运动员职业和/或比赛统计数据

[0109] -实时新闻广播，显示公告中检测到的对象/位置的信息图形

[0110] -电话营销广播，显示检测到的产品的价格比较。

[0111] 本发明在网上购物中的应用

[0112] 如图6所示，本发明使得视频110的观众可以通过查询视频的屏幕截图，以获得他/她注意到的关于产品的附加信息。在该示例600中，观众在电视上观看戏剧。观众注意到一名演员穿着一件特定的衬衫。观众可以拍摄屏幕截图以提交到系统120中。对于电视观看，他/她可以使用应用程序拍摄屏幕照片。如果观众正在将视频流式传输到手机、平板电脑或电脑上，他/她可以拍摄一张提交的屏幕截图。静止照片的图像(例如杂志广告)也可以提交。

[0113] 系统检测屏幕截图中的对象。在这种情况下，它检测到一件深色短袖衬衫。符合这一标准的商用产品将展示给观众610。他/她可以提交额外的标准，使查询更加具体。例如，对象搜索可以进一步缩小以仅包括特定设计(例如v领)或来自特定设计者的衬衫。

[0114] 此后，观众可以通过参与供应商在线购买产品620。这允许供应商通过产品放置来销售产品，而不需要额外的商业时间和/或广告。

[0115] 前面的描述仅公开了本发明的示例性实施例。落入本发明范围内的上述公开的设备和方法的修改对于本领域普通技术人员来说是显而易见的。因此，尽管已经结合本发明的示例性实施例公开了本发明，但是应当理解，其它实施例也可以落入由所附权利要求限定的本发明的精神和范围内。

[0116] 操作环境:

[0117] 该系统通常由中央服务器组成，该服务器通过数据网络连接到用户计算机。中央服务器可以由连接到一个或多个大容量存储设备的一个或多个计算机组成。中央服务器的精确架构并不限制要求保护的发明。此外，用户的计算机可以是笔记本或台式个人计算机。它可以是手机、智能手机或其他手持设备，包括平板电脑。用户计算机的精确形式因素并不

限制要求保护的发明。适用于本发明的众所周知的计算系统、环境和/或配置的示例包括但不限于个人计算机、服务器计算机、手持、膝上型或移动计算机或通信设备，例如手机和PDA、多处理器系统、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人计算机、小型计算机、大型计算机、包括上述系统或设备中的任何一种的分布式计算环境等。用户计算机的精确形式因素并不限制所要求保护的发明。在一个实施例中，省略了用户的计算机，而是提供了与中央服务器一起工作的单独计算功能。在这种情况下，用户将从另一台计算机登录服务器，并通过用户环境访问系统。

[0118] 用户环境可以容纳在中央服务器中或者可操作地连接到中央服务器。此外，用户可以通过因特网从中央服务器接收数据并向中央服务器发送数据，由此用户使用因特网网页浏览器访问账户，并且浏览器显示可操作地连接到中央服务器的交互式网页。中央服务器响应于从浏览器发送的数据和命令来发送和接收数据，该数据响应于用户对浏览器用户界面的操作。本发明的一些步骤可以在用户的计算机上执行，并将临时结果传送给服务器。这些临时结果可以在服务器上处理，最终结果会传回给用户。

[0119] 这里描述的方法可以在计算机系统上执行，该计算机系统通常包括中央处理单元((CPU)，该中央处理单元可操作地连接到存储设备、数据输入和输出电路((I/O)以及计算机数据网络通信电路。由计算机处理器执行的计算机代码可以获取由数据通信电路接收的数据，并将其存储在存储设备中。此外，CPU可以从I/O电路获取数据，并将其存储在存储设备中。而且，该CPU可以从存储设备获取数据，并通过I/O电路或数据通信电路将其输出。存储在存储器中的数据可以进一步从存储器中调出，由CPU以本文描述的方式进一步处理或修改，并在同一个存储器设备或不同的存储器设备中复原，该存储器设备可操作地连接到CPU，包括通过数据网络电路。该存储设备可以是任何类型的数据存储电路或磁存储器或光学设备，包括硬盘、光盘或固态存储器。I/Odevices可包括显示屏、扬声器、麦克风和可移动鼠标，其向计算机指示光标在显示器上的相对位置以及一个或多个可被驱动以指示命令的按钮。

[0120] 计算机可以在显示屏上显示用户界面的外观，该显示屏可操作地连接至I/O电路。作为计算机生成数据的结果，各种形状、文本和其他图形形式被显示在屏幕上，该数据使由像素构成的显示屏产生用户激活的浏览器用户界面。本发明的一些步骤可以在用户的计算机上执行，并将临时结果传送给服务器。可以在服务器上处理这些临时结果，最终结果会传回给用户。

[0121] 计算机可以在显示屏上显示用户界面的外观，该显示屏可操作地连接至I/O电路。作为计算机生成数据的结果，各种形状、文本和其他图形形式显示在屏幕上，该数据使由像素构成的显示屏产生呈现各种颜色和阴影。用户界面还显示在本领域中称为光标的图形对象。该对象在显示器上的位置向用户指示对屏幕上另一对象的选择。用户可以通过由I/O电路连接到计算机的另一个设备来移动光标。该设备检测用户的某些物理运动，例如，手在平面上的位置或手指在平面上的位置。这种设备在本领域中可以称为鼠标或触控板。在一些实施例中，显示屏本身可以通过感测显示屏表面上的一个或多个手指的存在和位置而用作触控板。当光标位于如按钮或开关的图形对象上时，用户可以通过接合鼠标或触控板或计算机设备上的物理开关，或者轻触触控板或触敏显示器来启动按钮或开关。当计算机检测到物理开关已被接通(或者触控板或触敏屏幕被敲击)，它会获取光标在屏幕上的明显位置

(或者以触敏屏幕为例,检测到的手指的位置),并执行与该位置相关联的处理。作为一个示例,并非旨在限制所公开的发明的范围,可以在屏幕上显示外观为二维框的图形对象,其中包含单词“enter”。如果光标(或敏感屏幕的手指位置)处于图形对象(例如显示框)的边界内时,计算机检测到开关已被接通,计算机将执行与“输入”命令相关联的过程。如此,屏幕上的图形对象创建了一个用户界面,允许用户控制计算机上的操作过程。

[0122] 本发明也可以完全在一个或多个服务器上执行。服务器可以是由具有大容量存储设备和网络连接的中央处理单元组成的计算机。此外,服务器可以包括与数据网络或其他数据传输连接连接在一起的多个这样的计算机,或者,网络上的具有网络访问存储的多个计算机,以提供这样的功能的方式作为一个组。普通技术人员会认识到,在一个服务器上完成的功能可以在多个服务器上分割和完成,这些服务器通过适当的进程间通信由计算机网络可操作地连接。此外,网站的访问可以通过处理安全或公共页面的互联网浏览器,或者通过在通过计算机网络连接到服务器的本地计算机上运行的客户端程序。数据消息和数据上传或下载可以使用典型的协议在互联网上传递,包括TCP/IP、HTTP、TCP、UDP、SMTP、RPC、FTP或允许在两台远程计算机上运行的进程通过数字网络通信交换信息的其他类型的数据通信协议。因此,数据消息可以是从计算机发送或由计算机接收的数据包,包含目标网络地址、目标进程或应用标识符以及数据值,目标应用程序可以在位于目标网络地址的目的计算机上解析这些数据值,以便目标应用程序提取和使用相关数据值。中央服务器的精确结构并不限制所要求保护的发明。此外,数据网络可以以几个级别运行,使得用户的计算机通过防火墙连接到一个服务器,该服务器发送通信至另一个执行所本发明方法的服务器上。

[0123] 用户计算机可以运行从远程服务器接收数据文件的程序,该数据文件被传递一个程序,该程序解释数据文件中的数据并命令显示设备呈现特定的文本、图像、视频、音频和其他对象的程序。当操作鼠标按钮时,该程序可以检测光标的相对位置,并且当按钮被按下时,该程序可以基于显示器上指示的相对位置上的位置来解释要执行的命令。该数据文件可以是HTML文档,该程序可以是网络浏览器程序,该命令可以是超链接,使浏览器向另一个远程数据网络地址位置请求新的HTML文档。HTML还可以有引用,引至被调用和被执行的其他代码模块,例如Flash或其他本机代码。

[0124] 相关领域的技术人员将理解,本发明可以用其他通信、数据处理或计算机系统配置来实现,包括:无线设备、互联网设备、手持设备(包括个人数字助理(PDAs))、可穿戴计算机、各种蜂窝或移动电话、多处理器系统、基于微处理器或可编程的消费电子产品、机顶盒、网络个人电脑、微型计算机、大型机等。实际上,术语“计算机”、“服务器”等在这里可互换使用,并且可以指任何上述设备和系统。

[0125] 在一些情况下,特别是当用户计算机是用于通过网络访问数据的移动计算设备时,网络可以是任何类型的蜂窝、基于IP或可转换的电信网络,包括但不限于全球移动通信系统(GSM)、时分多址(TDMA)、码分多址(CDMA)、正交频分多址(OFDM)、通用分组无线服务(GPRS)、增强型数据GSM环境(EDGE)、高级移动电话系统(AMPS)、全球微波接入互操作性(WiMAX)、通用移动电信系统(UMTS)、演进数据优化(EVDO)、长期演进(LTE),超级移动宽带(UMB),网络协议语音服务系统(VoIP),或者非授权移动接入(UMA)。

[0126] 互联网是一个计算机网络,它允许操作个人计算机的客户与远程计算机服务器进行交互,并通过网络查看从服务器传送到个人计算机的作为数据文件的内容。在一种协议

中，服务器使用被称为浏览器的本地程序来提供客户个人计算机上的网页。浏览器从服务器接收一个或多个数据文件，显示在客户的个人电脑屏幕上。浏览器从特定地址查找这些数据文件，该地址由一个名为统一资源定位器(URL)的字母数字字符串表示。然而，该网页可能包含从各种URL或IP地址下载的组件。网站是相关URL的集合，通常都共享相同的根地址或在某个实体的控制下。在一个实施例中，模拟空间的不同区域具有不同的区域。也就是说，模拟空间可以是一个单一的数据结构，但是不同的URL引用了数据结构中不同的位置。这使得有可能模拟一个大的区域，并让参与者开始在他们的虚拟社区内使用它。

[0127] 实现本文之前描述的全部或部分功能的计算机程序逻辑可以以各种形式实现，包括但不限于源代码形式、计算机可执行形式和各种中间形式(例如，由汇编器、编译器、链接器或定位器生成的形式)。源代码可以包括一系列计算机程序指令，用于各种操作系统或操作环境，该指令用各种编程语言(例如，目标代码、汇编语言、或诸如C、C-HF、C#、动作脚本、PHP、EcmaScript、JavaScript、JAVA或5HTML之类的高级语言)中的任一种来实现的。源代码可以定义并使用各种数据结构和通信消息。源代码可以是计算机可执行形式(例如，经由解释器)，或者源代码可以被转换(例如，经由翻译器、汇编器或编译器)为计算机可执行形式。

[0128] 一般性地描述计算机执行的计算机可执行指令(例如程序模块)的上下文描述了本发明。通常，程序模块包括程序、计算机程序、对象、组件、数据结构等，执行特定任务或实现特定抽象数据类型。计算机程序和数据可以以任何形式(例如源代码形式，计算机可执行形式，或中间形式)永久地或暂时地固定在有形存储介质中，例如半导体存储设备(例如RAM、ROM、PROM、EEPROM或闪存可编程RAM)、磁存储设备(例如磁盘或固定硬盘)、光存储设备(例如CD-ROM或DVD)、PC卡(例如PCMCIA卡)，或其他存储设备。计算机程序和数据可以任何形式固定在信号中，该信号可以使用各种通信技术中的任何一种传输到计算机，包括但不限于模拟技术、数字技术、光学技术、无线技术、网络技术和互联网技术。计算机程序和数据可以以任何形式作为可移动存储介质分发，附带打印或电子文件(例如，压缩打包软件或磁带)，预加载计算机系统(例如，在系统ROM或硬盘上)，或者从服务器或电子公告板分发至通信系统(例如，互联网或万维网)。应该认识到，如果需要，本发明的任何软件可以ROM(只读存储器)形式实现。如果需要，通常可以使用常规技术在硬件中执行软件。

[0129] 本发明也可以在分布式计算环境中实施，其中任务由通过通信网络连接的远程处理设备执行。在分布式计算环境中，程序模块可以位于本地和远程计算机存储介质中，包括存储器存储设备。普通技术人员将认识到，本发明可以在通过数据网络连接的一个或多个计算机处理器上执行，所述数据网络包括例如互联网。在另一个实施例中，该过程的不同步骤可以由地理上分离但通过数据网络连接的一个或多个计算机和存储设备执行，使得它们一起操作以执行该过程的各步骤。在一个实施例中，用户的计算机可以运行一个应用程序，该应用程序使得用户的计算机通过数据网络将一个或多个数据包的流传输到第二计算机上，第二计算机在这里指的是服务器。该服务器又可以连接到存储数据库的一个或多个大容量数据存储设备。该服务器可以执行一个程序，该程序接收发送的数据包并对发送的数据包进行解释，以便提取数据库查询信息。然后，该服务器可以通过访问大容量存储设备来获得查询的期望结果，从而执行本发明的其余步骤。或者，服务器可以将查询信息发送到连接到大容量存储设备的另一台计算机，该计算机可以执行本发明以获得期望的结果。然后，该结果可以由一个或多个数据包的另一个流传输回用户计算机，该数据包适当地寻址到用

户计算机。在一个实施例中，关系数据库可以容纳在一个或多个可操作地连接的服务器中，该服务器可操作地连接至计算机存储器(例如磁盘驱动器)。在另一个实施例中，关系数据库的初始化可以在服务器组上准备，并且与用户计算机的交互发生在整个过程中的不同位置。

[0130] 应当注意，此处使用流程图来演示本发明的各个方面，并且不应当被解释为将本发明限制于任何特定的逻辑流程或逻辑实现。所描述的逻辑可以被划分成不同的逻辑块(例如，程序、模块、函数或子程序)，而不改变总体结果或以其他方式偏离本发明的真实范围。通常，逻辑元件可以被添加、修改、省略、以不同的顺序执行，或者使用不同的逻辑结构(例如逻辑门、循环原语、条件逻辑和其他逻辑结构)来实现，而不改变总体结果或者以其他方式脱离本发明的真实范围。

[0131] 本发明的所述实施例旨在举例说明，对于本领域技术人员来说，许多变化和修改是显而易见的。所有这样的变化和修改都在所附权利要求所限定的本发明的范围内。虽然本文已详细描述和说明了本发明，但应清楚地理解，这仅仅是作为说明和示例，而不是作为限制。应当理解，为了清楚起见，在单独实施例的上下文中描述的本发明的各种特征也可以在单个实施例中组合提供。

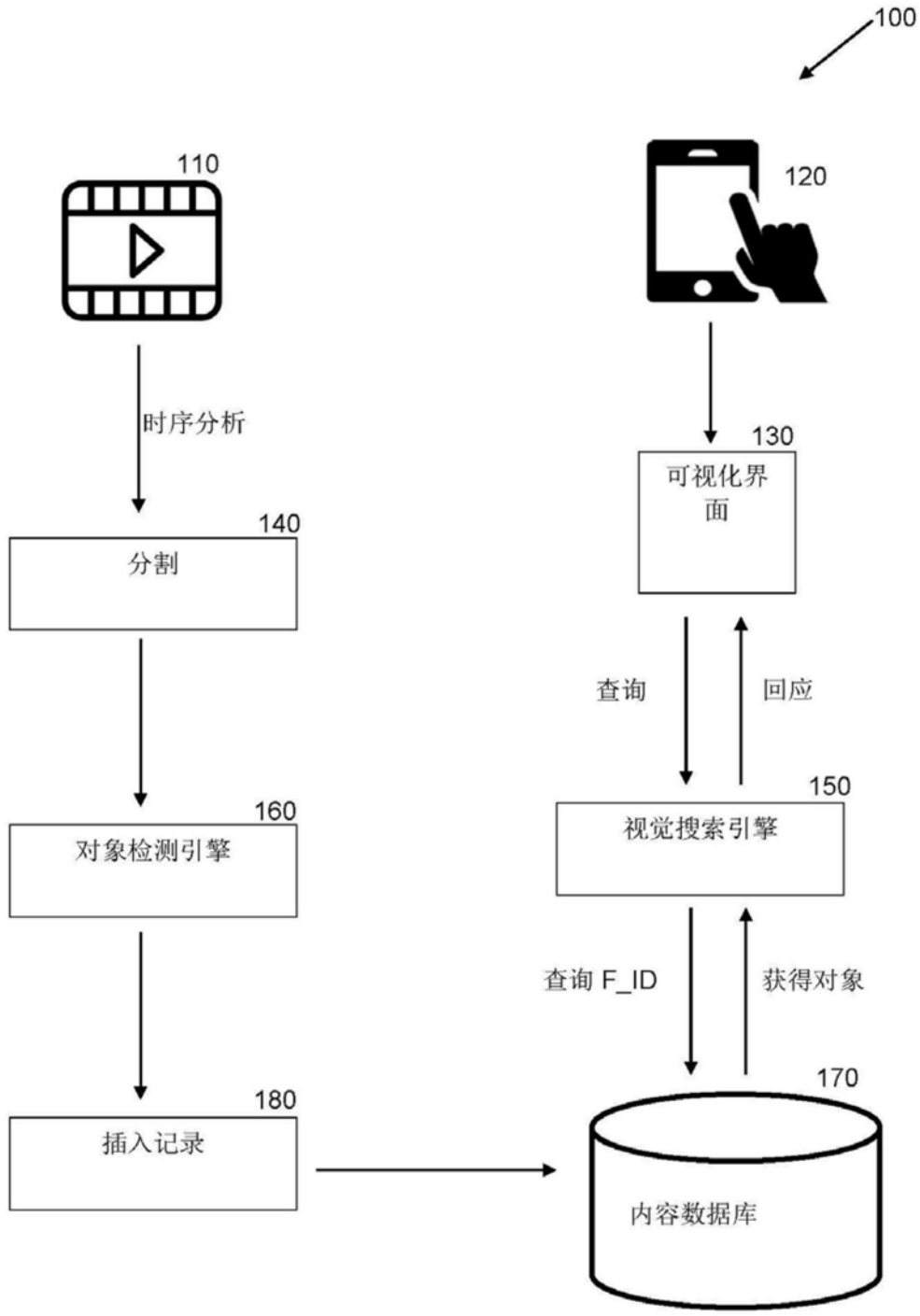

100
110
120
时序分析
130
可视化界面
140
分割
查询
回应
150
视觉搜索引擎
160
对象检测引擎
查询 F_ID
获得对象
170
180
插入记录
内容数据库

图1

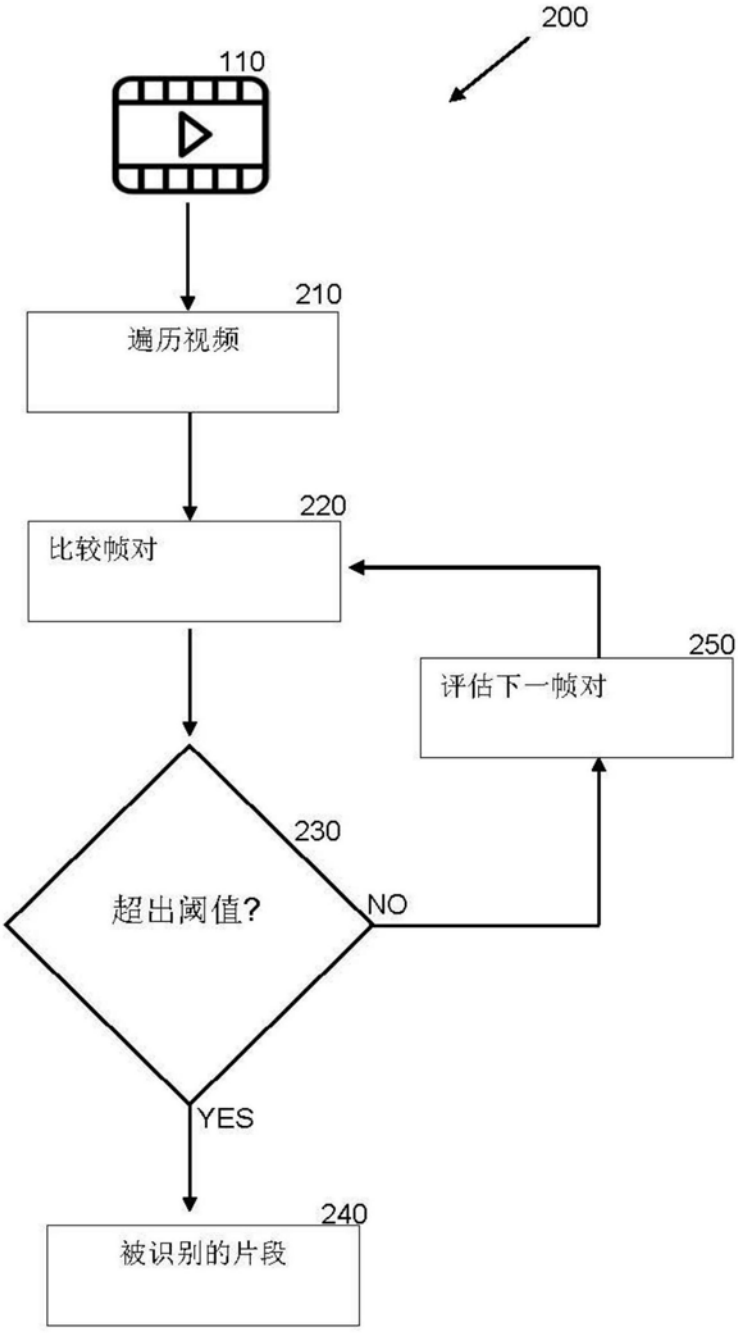

200
110
210
遍历视频
220
比较帧对
250
评估下一帧对
230
超出阈值?
NO
YES
240
被识别的片段

图2

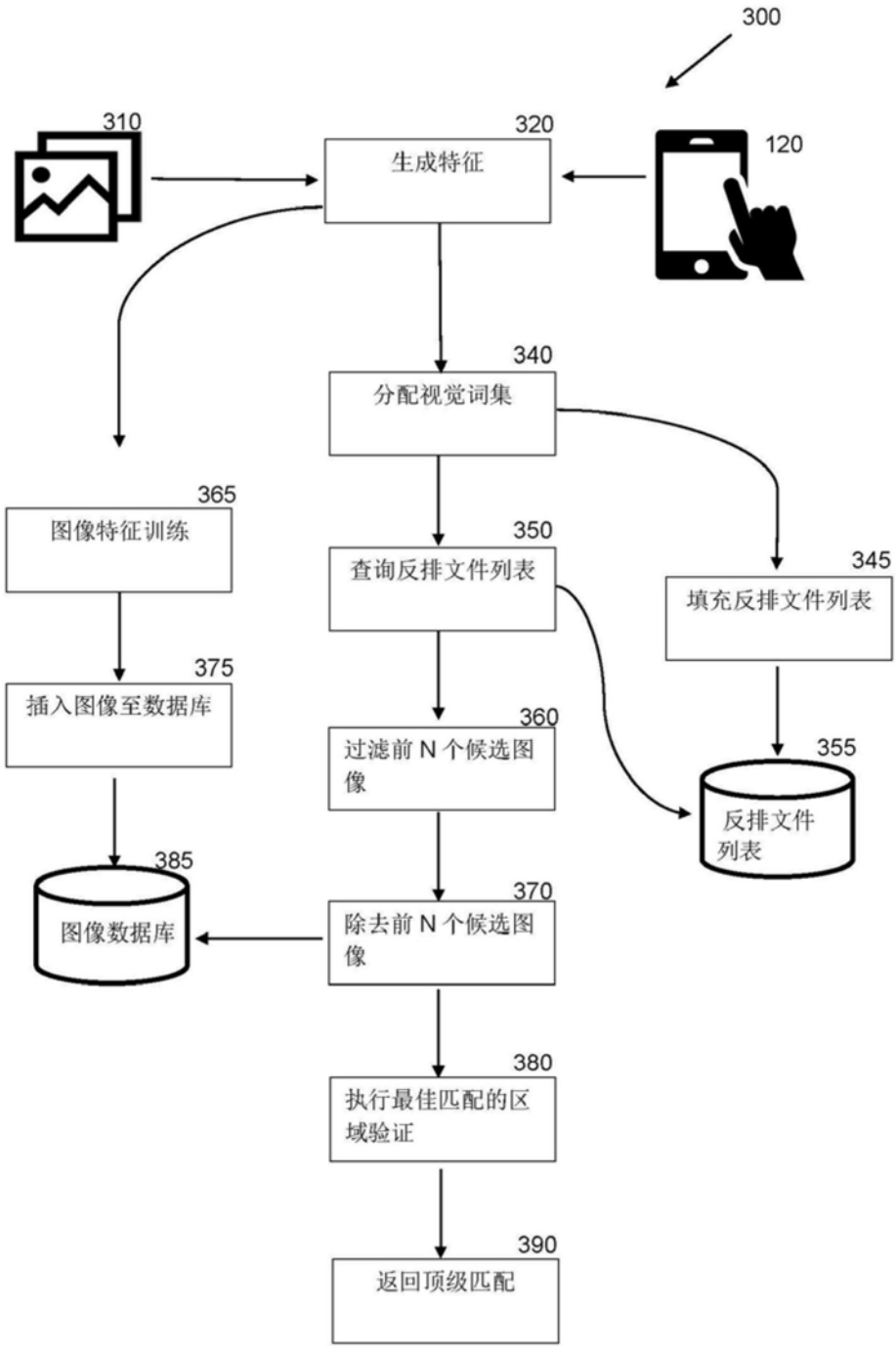

300
310
320
生成特征
120
340
分配视觉词集
365
图像特征训练
350
查询反排文件列表
345
填充反排文件列表
375
插入图像至数据库
360
过滤前 N 个候选图像
355
反排文件列表
385
图像数据库
370
除去前 N 个候选图像
380
执行最佳匹配的区域验证
390
返回顶级匹配

图3

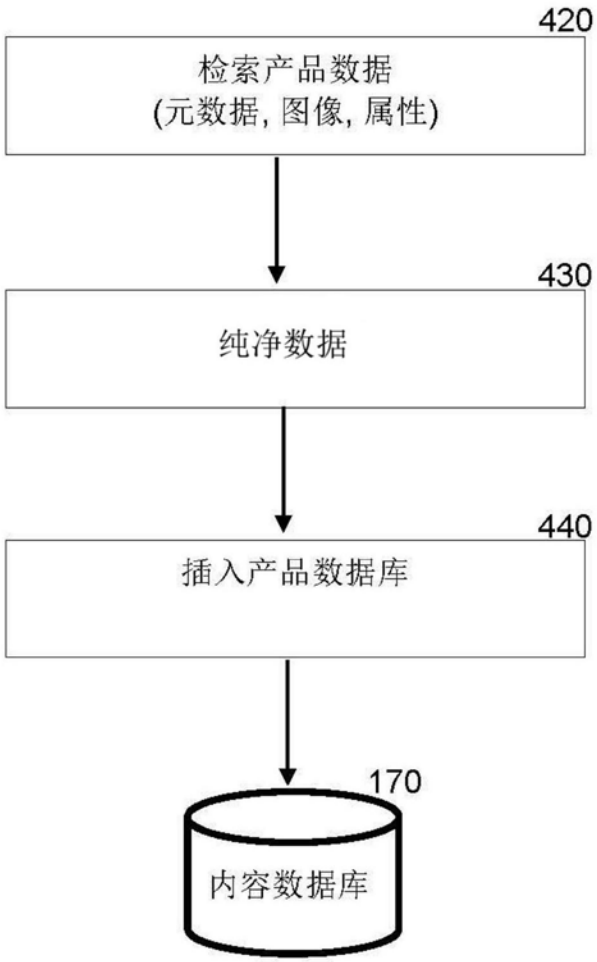


图4

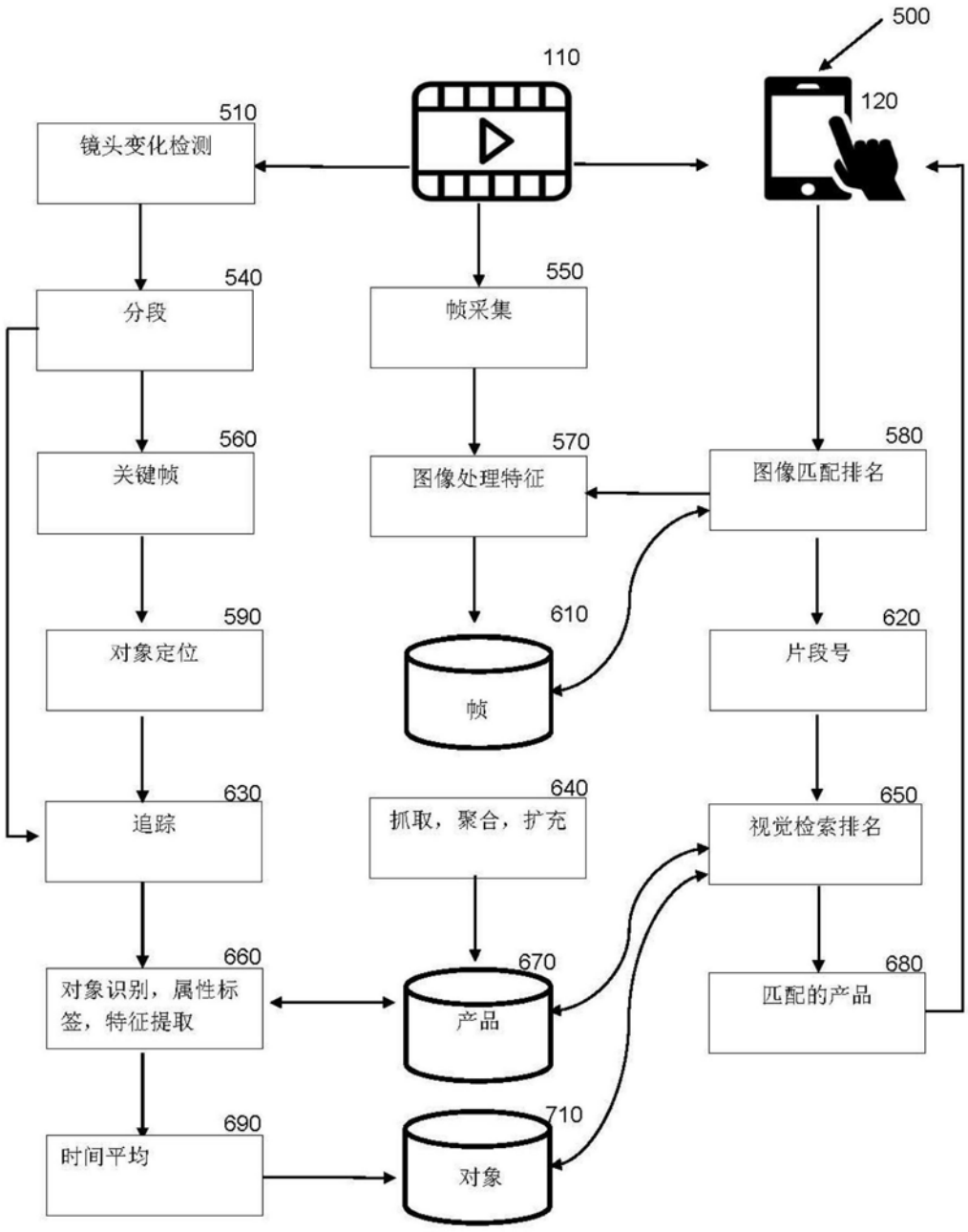

500
110
120
510
镜头变化检测
540
分段
550
帧采集
560
关键帧
570
图像处理特征
580
图像匹配排名
590
对象定位
610
帧
620
片段号
630
追踪
640
抓取，聚合，扩充
650
视觉检索排名
660
对象识别，属性标签，特征提取
670
产品
680
匹配的产品
690
时间平均
710
对象

图5

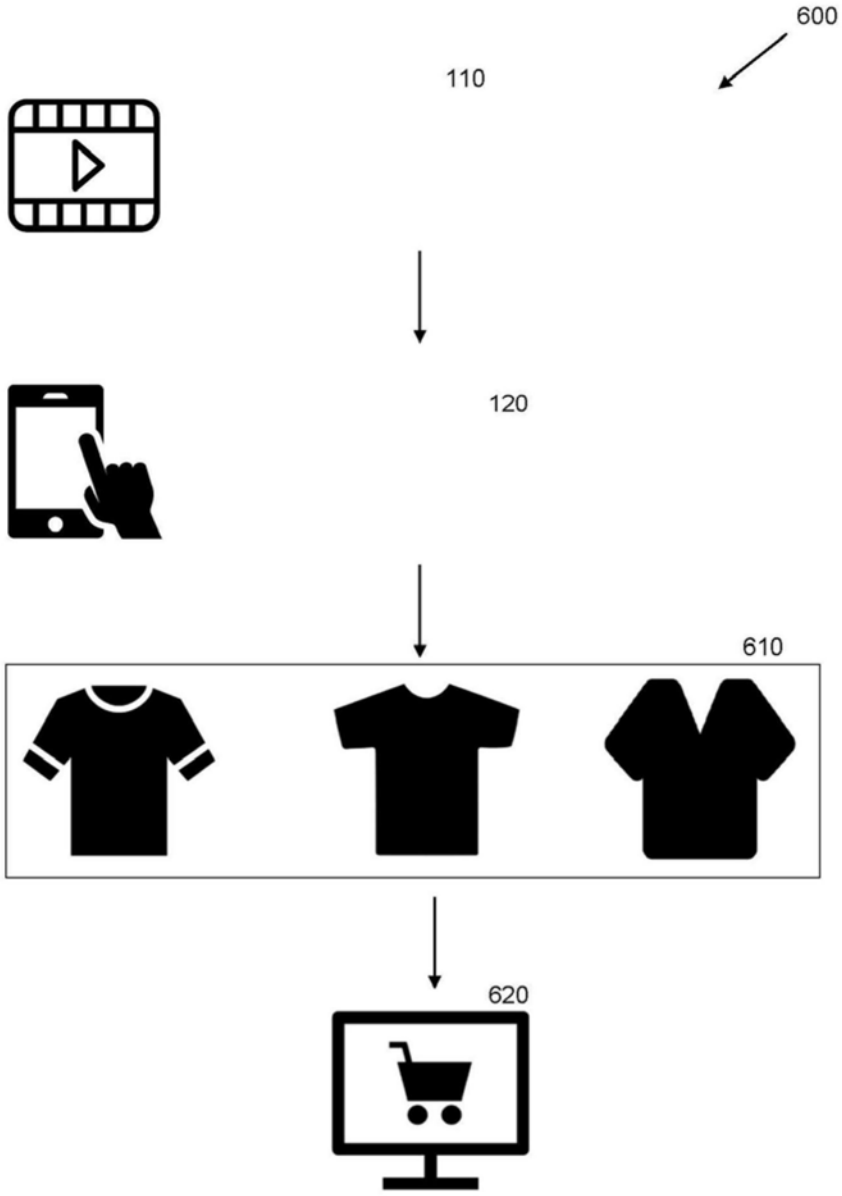


图6