



# (12)发明专利申请

(10)申请公布号 CN 109691124 A

(43)申请公布日 2019.04.26

(21)申请号 201780050009.5

(22)申请日 2017.06.19

(30)优先权数据

62/352,026 2016.06.20 US

(85)PCT国际申请进入国家阶段日

2019.02.14

(86)PCT国际申请的申请数据

PCT/IL2017/050678 2017.06.19

(87)PCT国际申请的公布数据

WO2017/221239 EN 2017.12.28

(71)申请人 皮克索洛特公司

地址 以色列特拉维夫市

(72)发明人 盖尔·奥兹 约阿夫·利伯曼

阿维·肖力

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262

代理人 陆建萍 杨明钊

(51)Int.Cl.

H04N 21/8549(2006.01)

H04N 21/2187(2006.01)

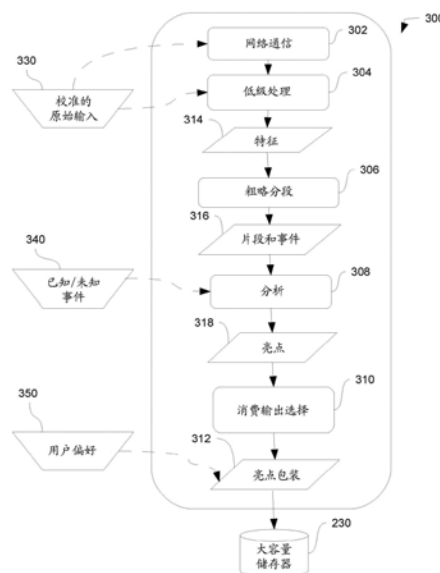
权利要求书2页 说明书6页 附图4页

## (54)发明名称

用于自动生成视频亮点的方法和系统

## (57)摘要

提供了用于从运动场的一个或多个视频流自动生成亮点视频的方法和系统。从至少一个摄像机捕获视频流，从音频、校准的视频和实际事件时间获得校准的和原始的输入。然后从校准的原始输入提取特征，创建片段，识别特定事件，以及确定亮点并输出亮点以用于消费，同时考虑不同类型的包装。包装的类型可以基于用户偏好。校准的视频流可以被实时地、周期性地接收和处理。



1. 一种用于从运动场的一个或多个视频流自动生成亮点视频的方法,包括:  
接收由位于所述运动场附近的至少一个摄像机捕获的校准的原始输入,其中所述校准的原始输入包括音频、校准的视频和实际事件时间,并且其中所述音频和视频是同步的(404);  
通过应用低级处理提取特征以将所述校准的原始输入转换成特征向量(406);  
从所述特征向量创建片段,并通过应用粗略分段来识别在片段的每一个片段中的特定事件(408);  
通过应用分析算法来确定所述事件中的每个事件是否是亮点(410);以及  
输出用于消费的亮点(412)并确定待输出以用于消费的包装的类型(414)。
2. 如权利要求1所述的方法,其中,接收校准的原始输入在每个预定义的时间段实时执行。
3. 如权利要求1所述的方法,其中,所述包装的类型包括以下中的至少一个:
  - 压缩的比赛;
  - 重播;
  - 亮点影片;以及,
  - 个人剪辑。
4. 如权利要求1所述的方法,还包括基于用户偏好确定所述包装的类型。
5. 如权利要求1所述的方法,其中,确定所述事件中的每个事件是否是亮点还基于深度学习过程。
6. 如权利要求1所述的方法,其中,所述音频包括环境音频。
7. 如权利要求1所述的方法,其中,所述音频包括评论员音频。
8. 如权利要求1所述的方法,其中,所捕获的原始输入包括以下中的至少一个:
  - 特定参赛者;
  - 所述运动场的各个区域;
  - 记分板;
  - 比赛时钟;
  - 参赛者在所述运动场上的活动的运动图片;
  - 相对时间;以及,
  - 用户收集的统计数据。
9. 如权利要求1所述的方法,还包括将所述原始输入存储在储存介质中,并通过通信网络(302)提供访问。
10. 一种用于从运动场的一个或多个视频流自动生成亮点视频的系统,包括:  
家庭服务器,其包括一个或多个处理器(100);  
一个或多个非暂时性计算机可读介质,其存储一组或更多组指令,在由一个或多个处理器执行时,所述指令使所述一个或多个处理器执行:  
接收由位于所述运动场附近的至少一个摄像机捕获的校准的原始输入,其中所述原始输入包括音频、视频和实际事件时间,并且其中所述音频和视频是同步的(404,302);  
通过应用低级处理提取特征以将所述校准的原始输入转换成特征向量(406,304);  
从所述特征向量创建片段,并通过应用粗略分段来识别在所述片段的每个片段中的特

定事件(408,306)；

通过应用分析算法来确定所述事件中的每个事件是否是亮点(410,308)；以及，  
输出用于消费的亮点并确定待输出以用于消费的包装的类型(412,414,310)。

11. 如权利要求10所述的系统，其中，接收所述校准的原始输入在每个预定义的时间段实时执行。

12. 如权利要求10所述的系统，其中，所述包装的类型包括以下中的至少一个：

- 压缩的比赛；
- 重播；
- 亮点影片；以及，
- 个人剪辑。

13. 如权利要求10所述的系统，还包括基于用户偏好确定所述包装的类型。

14. 如权利要求10所述的系统，其中，确定所述事件中的每个事件是否是亮点还基于深度学习过程。

15. 如权利要求10所述的系统，其中，所述音频包括环境音频。

16. 如权利要求10所述的系统，其中，所述音频包括评论员音频。

17. 如权利要求10所述的系统，其中，所捕获的原始输入包括以下中的至少一个：

- 特定的参赛者；
- 所述运动场的各个区域；
- 记分板；
- 比赛时钟；
- 参赛者在所述运动场上的活动的运动图片；
- 相对时间；以及，
- 用户收集的统计数据。

18. 如权利要求10所述的系统，还包括：

网络通信；并且，

其中，所述一个或多个处理器还执行：

将所述校准的原始输入存储在储存介质中；以及，

通过所述通信网络提供访问。

## 用于自动生成视频亮点的方法和系统

### 发明领域

[0001] 本发明的实施例涉及从运动场中的体育事件的一个或多个视频流创建视频亮点。更具体地,本发明的实施例涉及自动生成视频亮点用于共享和存储。

### [0002] 发明背景

[0003] 随着体育事件和其他事件数量的持续增加,亮点视频变得越来越重要。由于大多数观众想要看到这些事件中的许多事件,但时间有限,因此事件重要部分的视频亮点变得越来越重要。

### [0004] 发明概述

[0005] 以下是提供对本发明的初步理解的简要概述。概述不一定标识关键要素,也不限制本发明的范围,而是仅作为对以下描述的介绍。

[0006] 本发明的一个方面提供了一种用于从运动场的一个或多个视频流自动生成亮点视频的方法,其接收由位于运动场附近的至少一个摄像机捕获的校准的原始输入,提取特征以将校准的原始输入转换成特征向量,创建片段并识别片段的每一个中的特定事件,确定事件中每一个是否是亮点,并输出用于消费的亮点。

[0007] 本发明的这些、另外的和/或其它方面和/或优点:在下面的详细描述中阐述;可能从该详细描述中可推;和/或通过本发明的实践是可以知悉的。

### [0008] 附图简述

[0009] 关于本发明的主题在说明书的结论部分被特别指出并被清楚地要求保护。然而,本发明关于操作的组织和方法以及其目的、特征和优点,在参照附图一起阅读时,通过参考以下详细描述可得到最好的理解,其中:

[0010] 图1是对于系统的示例性环境的图,其中所公开主题的实施例根据本发明的某些实施例的多个方面来实施;

[0011] 图2示出了计算设备计算机体系结构的示例框图。

[0012] 图3示出了用于从运动场中的体育事件的视频流生成亮点的示例计算机体系结构。

[0013] 图4示出了用于从运动场中的体育事件的视频流生成亮点的示例方法的流程图。

[0014] 将认识到,为了说明的简单和清楚,图中所示的元素不一定按比例绘制。例如,为了清楚起见,一些元素的尺寸可以相对于其他元素被放大。此外,在认为适当的情况下,参考数字可在多个图中重复以指示对应的或类似的元素。

### [0015] 本发明的详细描述

[0016] 在以下详细描述中,阐述了许多具体细节以便提供对本发明的透彻理解。然而,本领域技术人员将理解,可以在没有这些具体细节的情况下实践本发明。在其他情况下,没有详细描述公知的方法、过程和部件,以免模糊本发明。

[0017] 本发明涉及用于从运动场中的体育事件的视频流识别并生成亮点的方法、系统和计算机程序产品。

[0018] 本发明的实施例可包括或使用专用或通用计算机硬件,如,例如一个或多个处

理器和系统存储器,如下面更详细地讨论的。本发明范围内的实施例还包括物理和其他计算机可读介质,以用于承载或存储计算机可执行指令和/或数据结构。存储计算机可执行指令的计算机可读介质是计算机储存介质。

[0019] 计算机储存介质设备包括ROM、RAM、EEPROM、闪存、GPU存储器和其他类型的存储器。

[0020] “通信网络”被定义为一个或多个数据链路,其实现计算机系统和/或模块之间电子数据的传输。当经由网络或另一个通信连接(硬接线、无线或者硬接线或无线的任何组合)将信息传递或提供到计算机时,计算机适当地将连接视为传输介质。

[0021] 进一步地,在到达各种计算机系统后,可将计算机可执行指令或数据结构的形式程序代码资料(means)从传输介质自动地传输到计算机储存介质设备(反之亦然)。

[0022] 在本文档中,“网站”是万维网(www)文件的相关的集合,其包括名为主页的开始文件或“网页”,且通常包括另外的文件或“网页”。术语“网站”被总体使用为包括“网站”和“网页”。

[0023] 本领域技术人员将认识到,本发明可以在具有多种类型的计算机系统配置的网络计算环境中实施,如个人计算机(例如膝上型计算机、台式计算机、平板计算机)或任何类型的计算设备,包括可以容易地从一个位置转移到另一个位置的移动设备(例如智能电话、个人数字助理(PDA)、移动电话或蜂窝电话)。

[0024] 根据上面定义的“计算机”,服务器通常是远程计算机或远程计算机系统,或其中的计算机程序,其通过通信介质可访问,如通信网络或其它计算机网络,包括因特网。“服务器”为同一计算机或其他计算机中的其他计算机程序(及其用户)提供服务或执行功能。服务器还可以包括虚拟机、基于软件的计算机仿真。

[0025] 中央处理单元(CPU)由一个或多个处理器形成,例如,传统处理器,如服务器、计算机和其他计算机化设备中所使用的那些处理器。例如,处理器可以包括AMD和英特尔的X86处理器、英特尔的Xenon®和Pentium®处理器,以及它们的任意组合。

[0026] 图形处理单元(GPU)是单芯片处理器,其主要用于管理和提高视频和图形的性能。GPU执行所有的计算和处理,这些计算和处理被优化用于图形计算和处理。示例GPU来自美国加利福尼亚州的Santa Clara的NVIDIA,且包括来自GeForce™产品的那些GPU。现在参考图1,图1显示了对于系统100的示例性环境的图,其包括网络102,家庭服务器104(也称为主服务器)链接到网络102。家庭服务器104执行根据本发明的各种处理。网络102例如是诸如局域网(LAN)或广域网(WAN)的通信网络,包括如因特网的公共网络。如图1所示,网络102可以是网络和/或多个网络的组合,包括例如蜂窝网络。本文使用的“链接”包括直接或间接的硬接线或无线链接,并且使包括服务器、部件等在内的计算机彼此进行电子通信和/或数据通信。

[0027] 摄像机106被链接到网络102以将记录的视频发送到家庭服务器104用于根据本发明进行处理,摄像机106代表位于运动场108附近的一个或多个摄像机,用于自动拍摄表示要针对其制作视频亮点的事件的体育事件。根据另一实施例,由摄像机106捕获的视频被放置在大容量储存器110(例如光盘、数字虚拟盘等)上,并通过网络102提供给管理员,以供家庭服务器104处理。在另一个实施例中,视频被实时处理。家庭服务器104具有包括一个或多个部件、引擎、模块等的体系结构,用于提供许多额外的服务器功能和操作。出于解释

的目的,家庭服务器104可以具有统一资源定位符(URL),例如www.hs.com。

[0028] 现在参考图2,其示出了家庭服务器104的示例框图。家庭服务器104包括计算机化的处理单元202、一个或多个存储器设备204、一个或更多个接口210、一个或更多个输入/输出设备226,该计算机化的处理单元202由具有一个或更多个处理器的中央处理单元(CPU)204和具有一个或更多个处理器的图形处理单元(GPU)206形成。处理器202还可以包括各种类型的计算机储存介质,如缓冲存储器。中央处理单元202使得CPU 204和GPU 206例如电气连接,包括彼此之间以及与存储器设备214、接口220、大容量储存设备230、I/O设备240和显示设备250进行电子通信和/或数据通信。

[0029] 中央处理单元204由包括微处理器的一个或更多个处理器形成,用于执行家庭服务器104的进一步描述的功能和操作。

[0030] 图形处理单元(GPU)206允许与CPU 204并行计算,使得计算机化的处理单元202可以与CPU 204和GPU 206中的一个或两个一起工作。

[0031] 存储器设备210与CPU 204和GPU 206相关联,并且包括各种计算机储存介质,如易失性存储器 and/或非易失性存储器。

[0032] 大容量储存设备230包括各种计算机储存介质,如磁盘、光盘。如图1所描绘,特别大容量的储存设备是硬盘驱动器232。大容量储存设备230包括可移除储存器234。

[0033] I/O设备240包括各种接口,如用户接口222、网络接口224和外围接口226,这些接口允许家庭服务器104与其他系统、设备和用户交互。

[0034] 更具体地,如图3中的计算机体系结构300所示。家庭服务器104包括网络通信模块302(可选的)、低级处理模块304、粗略分段(rough segmentation)模块306、分析模块308和消费输出选择310。

[0035] 网络通信模块302处理去往和来自家庭服务器104的所有通信,并且是可选的。

[0036] 低级处理模块304提取特征314。它将校准的原始输入330转换成特征向量(例如,输入图像中的轮廓和边缘)。例如,低级处理模块可以检测正在拍摄的篮球场上的篮球比赛的篮球周围的运动。

[0037] 粗略分段模块306用于预理由低级处理模块304所提取的特征314。模块306根据预定标准将各种阈值应用于所提取的特征314,以确定是否继续关注事件。结合前面提到的与篮球比赛相关的示例,监测篮球周围的运动的检测到的特征向量。如果篮球周围的运动没有持续,则该事件将被视为非事件并被丢弃。如果篮球周围的运动持续,则这是接下来要到分析模块308进行分析的事件316。

[0038] 分析模块308用于确定该事件是否是值得作为亮点的事件318,用于输出到消费输出选择模块310。根据一些实施例,分析模块308通过包括机器学习或“深度学习”的过程来执行其分析。在机器学习中,通过用已知事件和已知非事件340的大量样本进行编程,模块被编程或被教导什么是事件。例如,机器学习模块可以通过体验大量的“进球”和“非进球”的样本来决定该事件是否是足球比赛中的进球或非进球。类似地,分析篮筐周围视频的噪音,看看它是否是扣篮。通过其机器学习,这个模块现在确定该粗略分段是扣篮,从而现在将通过消费输出选择来处理该分段。

[0039] 消费输出选择模块310对从分析模块308接收的作为输入的亮点318进行分类和标记。根据一些实施例,标记可以是例如参与者(个人参赛者或团队)的例如亮点时间、真实时

间和比赛时间、比赛中的点、场地/球场位置、比赛特定的亮点类型(例如上篮、扣篮,以及事件的样本概率)。此外,模块310确定如何使用亮点,例如包装类型。根据实施例,用途是作为亮点影片中的亮点、压缩的比赛、重播、商业时间中断、参赛者特定的亮点等。模块310的输出是亮点包装312,并且可以存储在大容量存储器230或存储在云(未示出)中。它可以用来诸如例如经由社交媒体站点、网站、视频共享站点或其他地方与他人共享。

[0040] 根据一些实施例,可以提示用户选择期望的输出亮点和包装类型(用户偏好340)。

[0041] 图4示出了用于从运动场的视频流生成亮点视频的示例方法400的流程图。

[0042] 图4的方法始于开始块402。根据一些实施例,该方法是实时完成的,这意味着在每个预定义的时间连续接收原始输入404并遍历所有块。

[0043] 在块402处,摄像机被校准,使得运动场/球场上的特定点(位置)转换成来自摄像机的图像,音频也与视频同步。这是必要的准备阶段。

[0044] 该方法然后移动到块404,在块404中,家庭服务器104接收校准的原始输入,其形式为校准的视频,通常带有事件的音频以及实际事件时间。音频通常与视频以及其他事情相协调,例如记分板变化(在体育赛事的情况下)。音频包括环境音频,例如,由事件本身所产生的噪声,包括来自运动场的由参赛者所产生的噪声和群众噪声、广播员的评论等等。该视频包括参赛者、摄像机集中的运动场的各个区域、群众的视频(例如,坐和站以及在其间移动)。事件中也有事情发生,例如,记分板发生变化和比赛期间的相对时间,如时间段,如半场、四分之一场和其他分割(例如,整个比赛时间的开始、中间、结束)。例如,输入还包括群众标记(如进一步详细描述)、记分板视频、比赛时钟视频、终端用户收集的统计数据(例如来自社交媒体)。

[0045] 该方法移动到块406,在块406中,家庭服务器104执行低级处理406。在这种低级处理中,校准的原始输入被分成多个特征,例如,图像和视觉特征、音频、以及记分板发生变化、相对的比赛时间(例如比赛中的时间段)。根据一些实施例,低级处理406如下:

[0046] • 从摄像机110捕获的视频,校准的视频:校准的视频的处理是这样的,即利用关于摄像机视野中每个点的知识来处理视频,从而知道每个点如何转换成摄像机图像。例如,在知道了运动场的某些区域、通过球衣号码或其他标识符知道了某些参赛者(例如明星或关键表现参赛者)、知道了各个参赛者在场地上如何移动、识别了场地上的重要位置(如球网、中圈)的情况下来处理视频。校准的视频还包括处理特定于该比赛的独特的参赛者动作,例如射门、在篮筐附近起跳以将球灌进网中。此外,还处理全局人为因素(global artifact),诸如,例如啦啦队或其他团体表演等娱乐活动时的比赛休息中断,非比赛事件,如竞技、在竞技场上飞行的室内飞艇(indoor blimp)、t恤和热狗以及向群众分发的其他物品。

[0047] • 环境音频:环境音频的处理包括分析环境音频的音频水平(幅度和频率)及其上升和下降,例如欢呼和掌声形式的群众声音、中断时的声音(如暂停(timeout))、比赛停止、群众唱歌、指示进球和比赛“结束”时所播放的音乐、激励群众为球队加油并制造声势的音乐、场上哨声、包括球噪声和叫喊声的参赛者噪声、裁判哨声等。

[0048] • 评论员/广播员音频,这里称为评论员音频:评论员音频的处理包括分析评论员的语音水平,因为高音调和快速讲话通常指示重要事件。分析评论员音频的语音波动也是如此。此外,可以针对具有特定单词的语言来分析评论员音频,这些单词指示比赛中的重要

事件,如“得分”、“进球”、“投篮得分 (basket)”等。此外,还可以对语言识别进行编程,以发现评论员特定的短语,例如单词“进球”的拉长的发音是对于一个特定足球评论员的为人所熟知的短语。

[0049] • 记分板:记分板是从分数的变化、比赛时间、比赛时间事件(如时间段的结束、比赛停止的时间、暂停等)来观察的。分析记分板以确定比赛的相对时间,从而基于比赛时间查看发生了什么事。记分板可用于确定比赛中的趋势,例如一个队在短时间内快速得分,这可能指示亮点。

[0050] • 群众标记:群众标记包括接收与比赛中的某些事件相关联的输入。例如,群众中的成员可能会评论在特定时间得分的进球,这些群众评论可以经由社交网络进行,例如 Facebook®、Instagram®和Twitter,这些评论例如通过查看一天中的实际时间并将其与社交媒体帖子的实际时间进行比较而被关联到比赛中的各种时间点。

[0051] • 参赛者追踪:在参赛者追踪中,家庭服务器104寻找特定参赛者及其特定移动,例如当他们加速、减速或移动到场上的不同位置时。例如,如果一名球员在场地的进攻侧,则在相当长的时间里,他可能会进球,或者协助进球,这是对于亮点的主题。此外,参赛者替换和替换的时间可能指示重要事件。例如,在篮球比赛中,一个发挥很好的参赛者通常会在比赛结束时被替换,所以观众可以给那个参赛者欢呼、掌声等。

[0052] 该方法然后移动到块408,在块408中,家庭服务器104执行粗略分段。基于视频的低级处理406创建片段,并且识别每个片段中的特定事件。例如,如果某个时间段内有带有环境噪声和评论员噪声的视频,这可能是值得识别为片段并值得在该片段中识别特定事件的一个片段。如果这个片段超过预定阈值,则该片段作为可能的亮点片段被保留。

[0053] 该方法然后移动到块410,在块410中,根据实施例,由诸如深度学习过程的机器学习(ML)过程执行分析,其中,在这个阶段处,从最初校准的原始输入数据提取新的特征向量。从这个新的特征向量可以做出某个片段是否是亮点的确定。机器学习过程可以在深度学习网络上执行,深度学习网络使用离线过程来学习不同的片段被定义为亮点(或不被定义为亮点)。如本文使用的“深度学习”是基于一组算法的机器学习的一个分支,这些算法试图通过使用具有复杂结构或以其他方式由多个非线性变换组成的多个处理层来对数据中的高层抽象建模。

[0054] 以上提到的深度学习是例如一个两步过程。第一步是学习步骤,其中亮点被表征为特定事件(通常通过人工输入),每个特定事件都被学习。在从已学习事件的大规模样本学习了某些事件之后,家庭服务器104最终被训练为识别特定的事件。根据实施例,信号特征、图像特征、音频特征和视频特征被放入特征桶,并且特征桶进入深度学习机制。输入的用于学习的事件越多,对于亮点分类错误的概率就越小。此外,亮点可以被加权和过滤以作为最终步骤来确定它们是否是亮点。接下来,深度学习结果被应用于片段,以确定它是否是亮点,以及它是什么类型的亮点,例如一个进球、一场精彩的比赛等等。

[0055] 一旦深度学习完成,该方法移动到块412,在块412中,由深度学习过程确定为亮点的事件被输出以用于消费。输出的亮点包括相关联的分类决策和元数据,元数据指示输出亮点的亮点类型。

[0056] 该方法然后移动到块414,在块414中选择包装类型。根据一个实施例,它基于预定义的默认类型。根据另一个实施例,它基于用户偏好,例如,由系统提示用户选择期望的输



出亮点,以及对于亮点的包装类型。包装可以是例如压缩的比赛,其中考虑了比赛的所有动作,并且创建了讲述比赛“故事”的简短版本(通常在5到10分钟之间);特定重播,其中系统识别到不存在有效(active)比赛(例如暂停)并引入亮点作为重播;亮点影片,其聚焦于激动人心的时刻;个人剪辑,例如,聚焦于特定参赛者的剪辑,带有一些比赛亮点或一系列动作,这些亮点或动作被播放商业广告的时间所包围。

[0057] 根据一个实施例,该方法移动到其中方法结束的块416。根据另一个实施例,在实时处理视频的情况下,该方法移动到块404以处理下一个校准的原始输入。

[0058] 本发明的实施例的方法和/或系统的实现可以涉及手动地、自动地或其组合来执行或完成所选择的任务。此外,根据本发明的方法和/或系统的实施例的实际仪器和设备,可以使用操作系统通过硬件、通过软件或通过固件或通过其组合来实现若干任务。

[0059] 例如,根据本发明的实施例的用于执行所选任务的硬件可以被实现为芯片或电路。作为软件,根据本发明的实施例的所选任务可以被实现为由计算机使用任何合适的操作系统执行的多个软件指令。

[0060] 在本发明的示例性实施例中,根据本文描述的方法和/或系统的示例性实施例的一个或多个任务由诸如用于执行多个指令的计算平台的数据处理器执行。可选地,数据处理器包括用于存储指令和/或数据的易失性存储器和/或非易失性存储器,例如,非易失性储存介质,如用于存储指令和/或数据的磁性硬盘和/或可移除介质。还可选地提供显示器和/或用户输入设备(诸如键盘或鼠标)。

[0061] 虽然本文已经图示和描述了本发明的某些特征,但是本领域普通技术人员将想到许多修改、替换、改变和等同物。因此,应当理解,所附权利要求旨在覆盖落入本发明的真实精神内的所有这样的修改和改变。

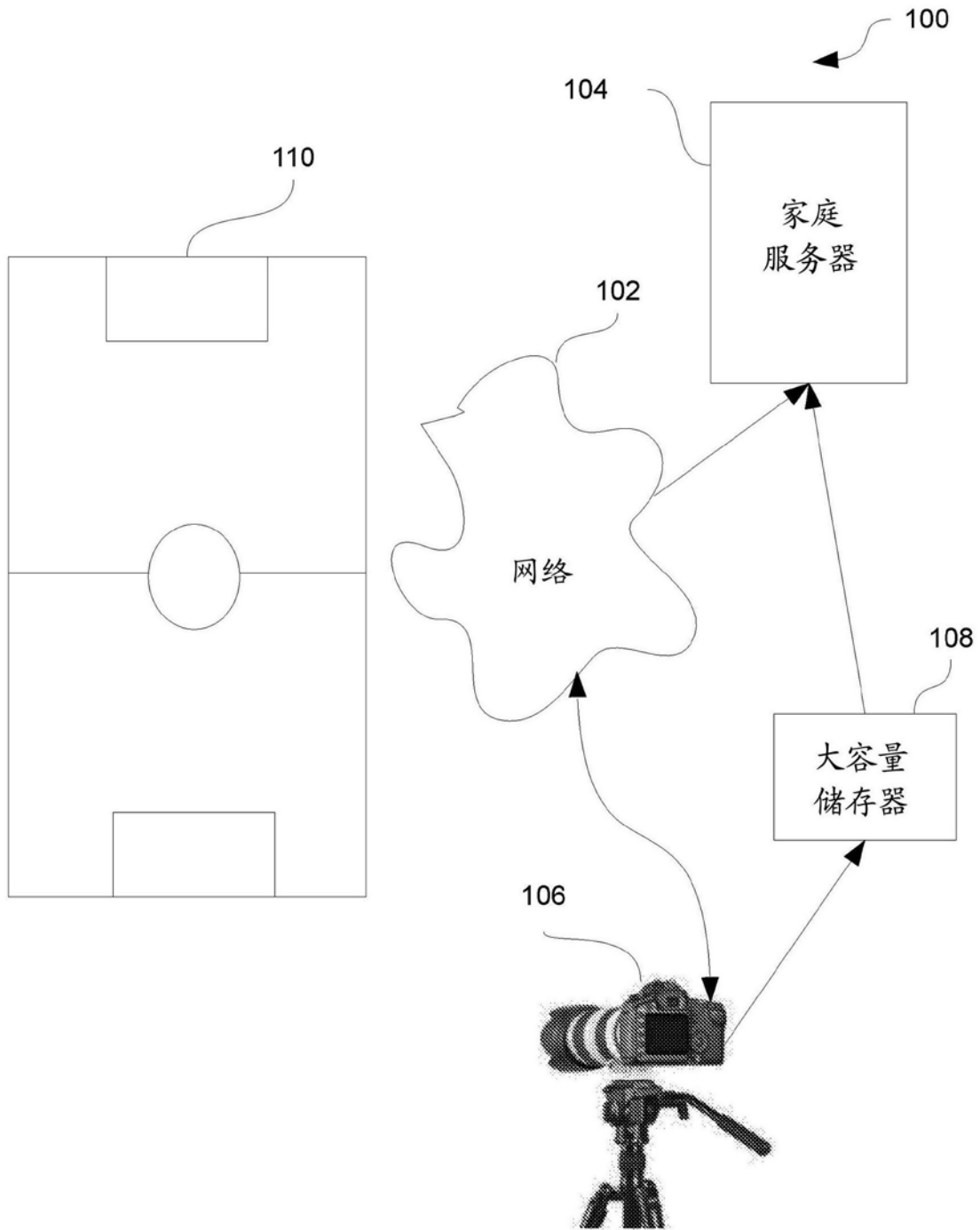


图1

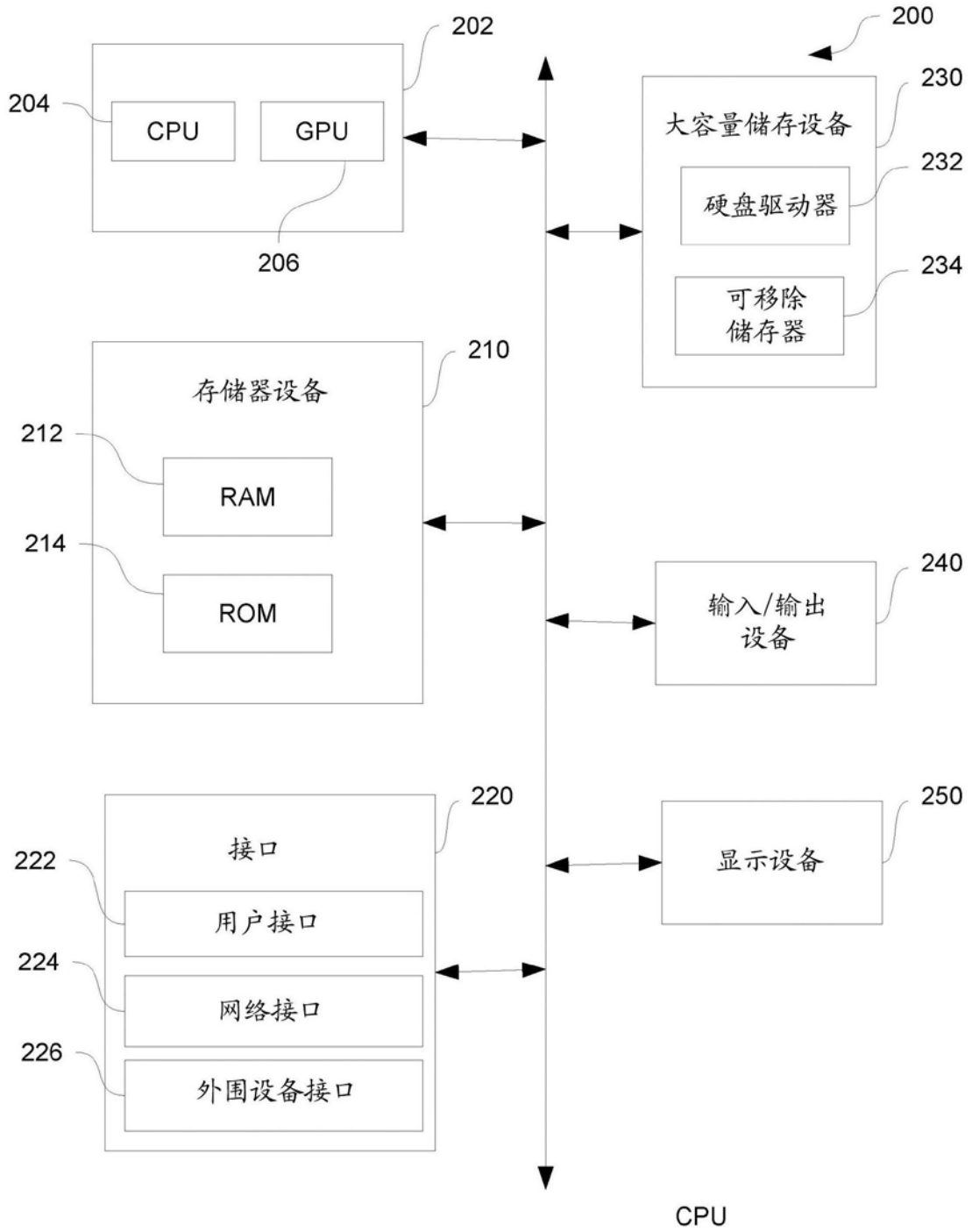


图2

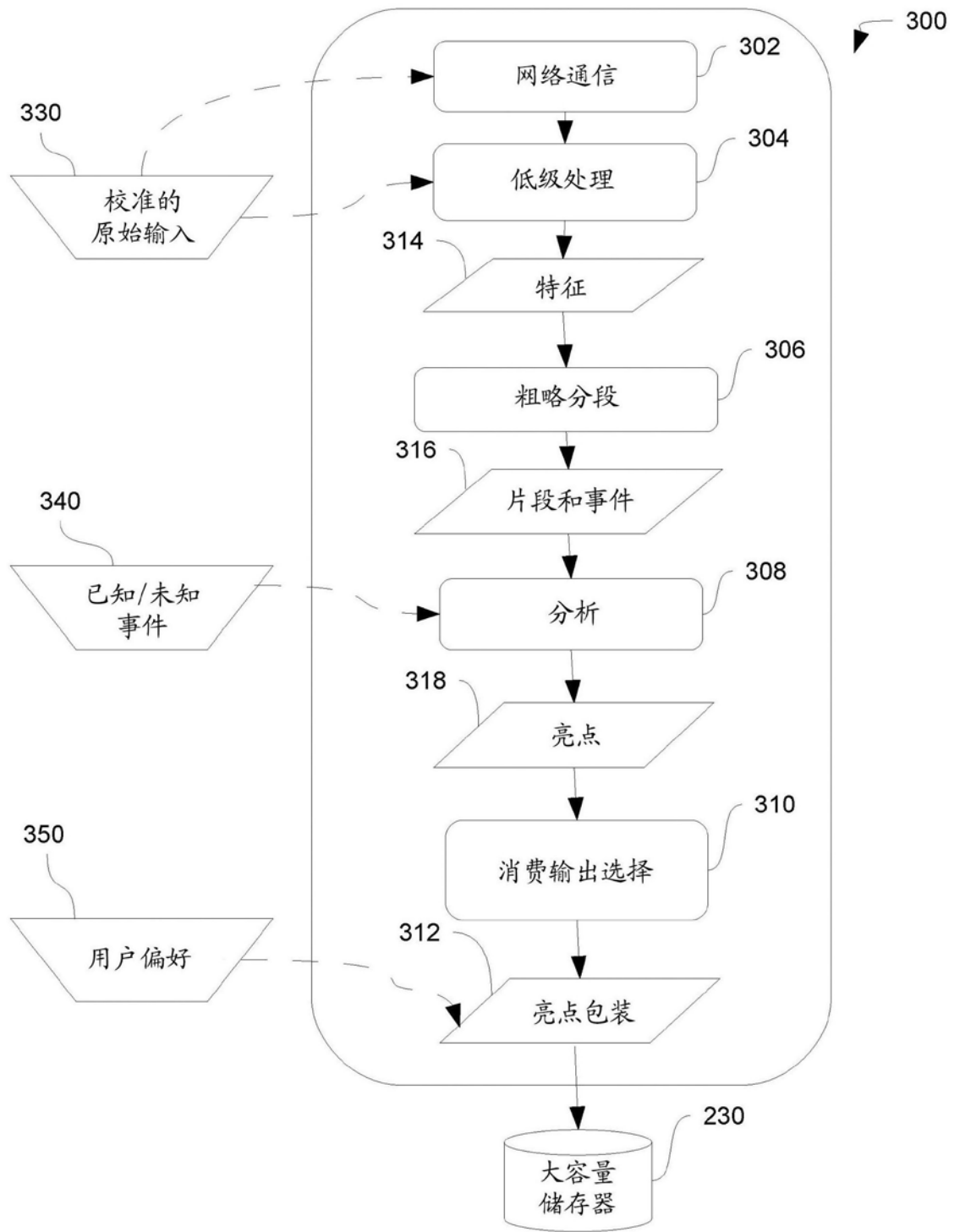


图3

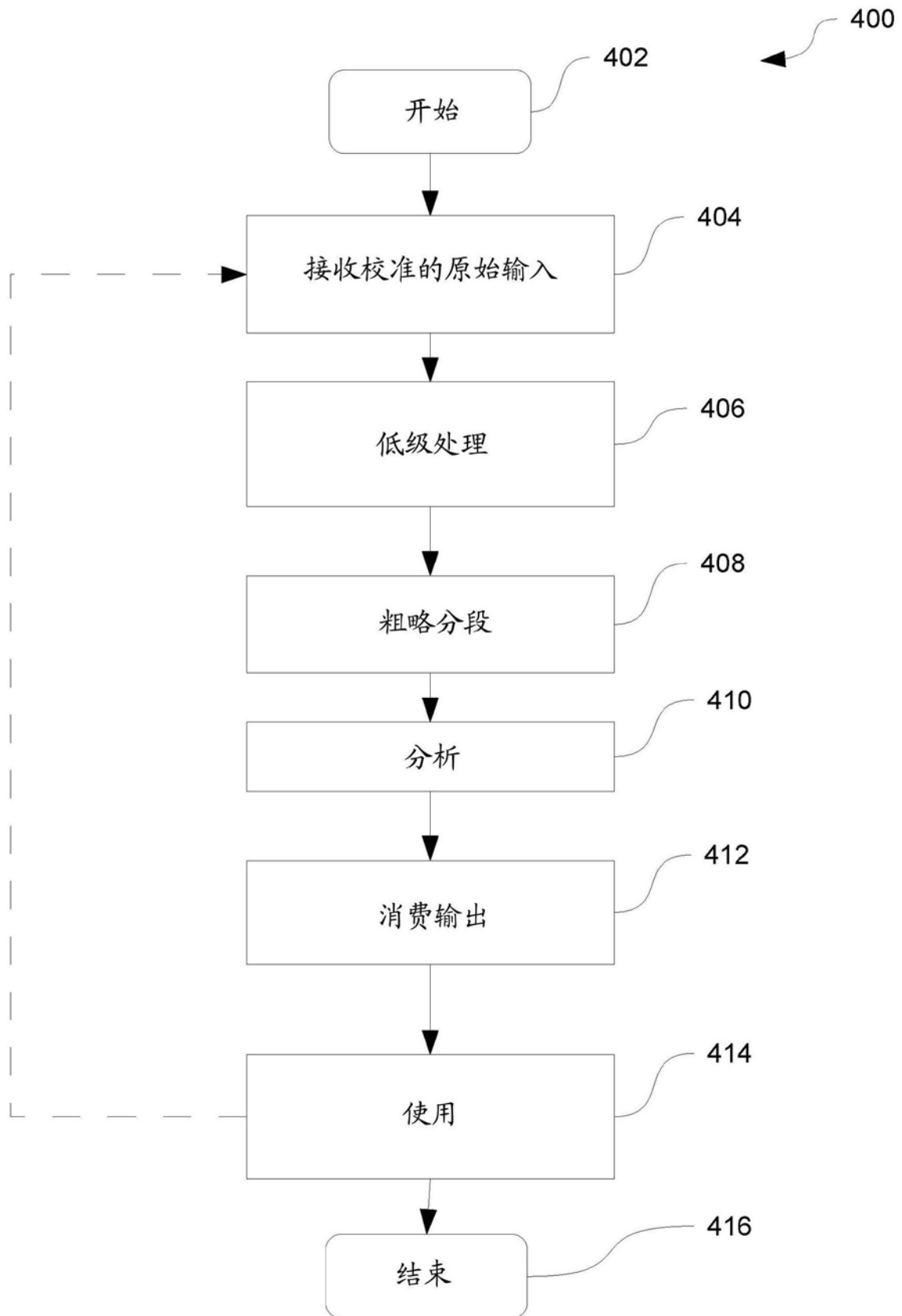


图4