



(12) 发明专利申请

(10) 申请公布号 CN 103620682 A

(43) 申请公布日 2014. 03. 05

(21) 申请号 201280023474. 7

(74) 专利代理机构 中科专利商标代理有限责任

(22) 申请日 2012. 05. 17

公司 11021

(30) 优先权数据

代理人 潘剑颖

13/110, 085 2011. 05. 18 US

(51) Int. Cl.

(85) PCT国际申请进入国家阶段日

G11B 27/24 (2006. 01)

2013. 11. 15

G11B 27/00 (2006. 01)

(86) PCT国际申请的申请数据

PCT/US2012/038233 2012. 05. 17

(87) PCT国际申请的公布数据

W02012/158859 EN 2012. 11. 22

(71) 申请人 高智 83 基金会有限责任公司

地址 美国内华达州

(72) 发明人 基思·斯托尔·卡恩

布鲁斯·哈罗德·皮尔曼

亚伦·托马斯·狄佛

约翰·R·麦科伊 弗兰克·拉扎维

罗伯特·格特辛格

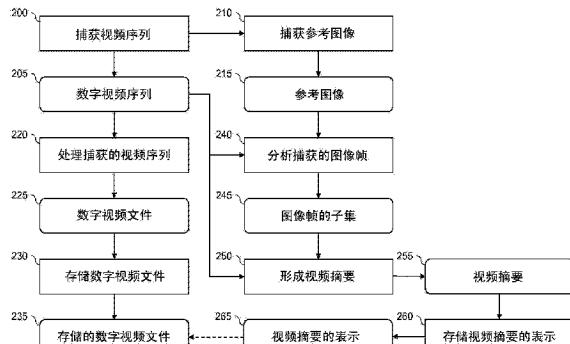
权利要求书2页 说明书13页 附图6页

(54) 发明名称

包括感兴趣的特征的视频摘要

(57) 摘要

一种使用包括以下各项的方法来提供视频摘要的数字摄像机系统：指定参考数据，其中，所述参考数据指示感兴趣的特征；使用图像传感器来捕获场景的视频序列，该视频序列包括图像帧的时间序列；使用视频处理路径来处理捕获的视频序列，以形成数字视频文件；在捕获视频序列期间，使用特征识别算法分析捕获的图像帧，以识别包含感兴趣的特征的图像帧的子集；形成包括与捕获的视频序列中的所有图像帧相比更少的图像帧的视频摘要，其中，视频摘要包括所识别的包含感兴趣的特征的图像帧的子集的至少一部分；将数字视频文件存储在存储设备中；以及将视频摘要的表示存储在存储设备中。



1. 一种用于捕获视频序列并且提供相关联的视频摘要的数字摄像机系统，包括：
图像传感器，用于捕获数字图像；
光学系统，用于在所述图像传感器上形成场景的图像；
数据处理系统；
存储设备，用于存储捕获的视频序列；以及
程序存储器，以通信方式连接到所述数据处理系统并且存储指令，所述指令被配置为使所述数据处理系统执行用于形成视频摘要的方法，其中，所述方法包括：
指定参考数据，其中，所述参考数据指示感兴趣的特征；
使用所述图像传感器捕获所述场景的视频序列，所述视频序列包括图像帧的时间序列；
使用视频处理路径处理所捕获的视频序列，以形成数字视频文件；
使用特征识别算法分析所捕获的图像帧，以识别包含所述感兴趣的特征的所述图像帧的子集；
形成包括与所捕获的视频序列中的所有图像帧相比更少的图像帧的所述视频摘要，其中，所述视频摘要包括所识别的包含所述感兴趣的特征的图像帧的子集的至少一部分；
将所述数字视频文件存储在所述存储设备中；以及
将所述视频摘要的表示存储在所述存储设备中。
2. 根据权利要求 1 所述的数字摄像机系统，其中，所述参考数据包括包含所述感兴趣的特征的参考图像。
3. 根据权利要求 2 所述的数字摄像机系统，其中，所述参考图像包括特定的个人或特定的对象。
4. 根据权利要求 1 所述的数字摄像机系统，其中，所述参考数据是感兴趣的特征的文本描述。
5. 根据权利要求 4 所述的数字摄像机系统，其中，所述文本描述包括个人的名字、对象、位置或事件。
6. 根据权利要求 1 所述的数字摄像机系统，其中，所述参考数据包括指定期望的图像帧特征的信息。
7. 根据权利要求 6 所述的数字摄像机系统，其中，所述期望的图像帧特征是照相机运动特征、照相机缩放处理特征、或者对象运动特征。
8. 根据权利要求 1 所述的数字摄像机系统，其中，所述视频摘要的所述表示是摘要数字视频文件。
9. 根据权利要求 8 所述的数字摄像机系统，其中，所述摘要数字视频文件是使用与用于形成所述数字视频文件的视频处理路径不同的视频处理路径，从所捕获的视频序列的所述图像帧中形成的。
10. 根据权利要求 8 所述的数字摄像机系统，其中，所述摘要数字视频文件至少部分地与所述数字视频文件同时形成。
11. 根据权利要求 1 所述的数字摄像机系统，其中，所述视频摘要的所述表示包括指定所述数字视频文件中要包含在所述视频摘要中的图像帧的集合的帧标识元数据。
12. 根据权利要求 11 所述的数字摄像机系统，其中，所述帧标识元数据被存储在所述

数字视频文件中。

13. 根据权利要求 1 所述的数字摄像机系统, 其中, 除了所识别的包含所述感兴趣的特征的图像帧的子集以外, 所述视频摘要还包括额外的图像帧。

14. 根据权利要求 13 所述的数字摄像机系统, 其中, 所述额外的图像帧包括紧邻所识别的包含所述感兴趣的特征的图像帧的子集中的图像帧之前或之后的图像帧。

15. 根据权利要求 14 所述的数字摄像机系统, 其中, 所述视频序列的所述图像帧被压缩以存储在所述数字视频文件中, 并且选择紧邻所识别的包含所述感兴趣的特征的图像帧的子集中的图像帧之前或之后的所述图像帧, 以使得所述视频摘要包括能够在无需对压缩的图像帧进行解码的情况下从所述数字视频文件中提取的图像帧组。

16. 根据权利要求 13 所述的数字摄像机系统, 其中, 所述额外的图像帧包括所捕获的视频序列中被确定为重要部分的其它部分。

17. 根据权利要求 16 所述的数字摄像机系统, 其中, 所捕获的视频序列的所述重要部分包括使用关键帧提取算法识别的关键图像帧。

18. 根据权利要求 1 所述的数字摄像机系统, 其中, 所述分析所捕获的图像帧的步骤是在捕获所述视频序列期间执行的。

19. 根据权利要求 1 所述的数字摄像机系统, 其中, 所述视频摘要是从所识别的包含所述感兴趣的特征的图像帧的子集中选择的静态图像的集合。

20. 一种用于形成视频序列的视频摘要的方法, 包括 :

接收包括图像帧的时间序列的视频序列 ;

指定参考数据, 其中, 所述参考数据指示感兴趣的特征 ;

使用数据处理器以利用特征识别算法自动地分析所述图像帧, 以识别包含所述感兴趣的特征的所述图像帧的子集 ;

形成包括与所述视频序列中的所有图像帧相比更少的图像帧的所述视频摘要, 其中, 所述视频摘要包括所识别的包含所述感兴趣的特征的图像帧的子集的至少一部分 ; 以及

将所述视频摘要的表示存储在处理器可访问的存储设备中。

包括感兴趣的特征的视频摘要

技术领域

[0001] 本发明涉及数字视频处理领域,更具体地,涉及用于形成数字视频摘要的系统和方法。

背景技术

[0002] 很多数字捕获设备能够捕获视频以及静态图像。然而,管理数字视频内容可能是一项困难的任务。通常在视觉上使用视频的第一帧的缩略图图像来表示视频。这可能不能提供对视频的内容的深刻理解。确定特定的事件是否包含在给定的视频中通常需要查看整个视频。对于漫长的视频而言,用户可能更希望能够在不必查看整个视频的情况下获得视频的快速摘要。

[0003] 从共享的角度来看,数字视频也可能带来实际的问题。很多数字捕获设备以 30 或 60 帧每秒、以高达 1920×1080 个像素的空间分辨率来录制视频。即使当被压缩时,产生的数据的量可能使得共享即使相当短的视频也是不切实际的。

[0004] 视频编辑软件可以用于手动地将视频概括为更容易地共享的更短的版本。然而,手动视频编辑可能是漫长、费力的过程,并且很多用户对手动编辑不感兴趣。还存在自动视频概括算法。这些解决方案以捕获的视频作为输入开始,并且分析视频以确定视频摘要。Ratakonda 的题为“Method for hierarchical summarization and browsing of digital video”的美国专利 5,995,095 公开了一种用于基于视频序列的关键帧生成分层摘要的方法。Li 等的题为“Scalable video summarization and navigation system and method”的美国专利 7,035,435 描述了一种用于向视频中的每一个场景、镜头和帧指派重要性值并且使用该重要性值来确定视频摘要的关键帧的方法。Edwards 等的题为“Automatic editing of a visual recording to eliminate content of unacceptably low quality and/or very little or no interest”的美国专利 7,483,618 公开了一种用于确定从视频中消除了质量较低或不太感兴趣的内容的视频摘要的方法。

[0005] 然而,自动视频概括算法非常复杂,这是因为必须对视频进行解码以执行对于确定视频摘要而言所需的分析。因此,不能在数字捕获设备上立即查看与刚捕获的视频相对应的视频摘要。这种缺点使得难以促进快速浏览和共享捕获的视频。

[0006] 当创建视频摘要时,通常期望在摘要中具有特定的特征。视频摘要被创建为包含其中存在特征的视频内容的一些或全部。这些特征的示例可以包括人、宠物、事件、位置、活动或对象。手动地创建这种定制的视频摘要可能是冗长乏味的过程。使用桌面软件生成这种定制的视频摘要阻碍了快速浏览和共享视频摘要的能力。

[0007] 因此,将期望提供用于在数字捕获设备中计算视频摘要的系统和方法。具体地说,将期望提供允许在完成视频捕获以后以最小的延迟在数字捕获设备上生成视频摘要的解决方案。此外,将期望提供包含用户指定的特征的视频摘要。

发明内容

[0008] 本发明表示一种用于捕获视频序列并且提供相关联的视频摘要的数字摄像机系统,包括:

- [0009] 图像传感器,用于捕获数字图像;
 - [0010] 光学系统,用于在所述图像传感器上形成场景的图像;
 - [0011] 数据处理系统;
 - [0012] 存储设备,用于存储捕获的视频序列;以及
 - [0013] 程序存储器,以通信方式连接到所述数据处理系统并且存储指令,所述指令被配置为使所述数据处理系统执行用于形成视频摘要的方法,其中,所述方法包括:
 - [0014] 指定参考数据,其中,所述参考数据指示感兴趣的特征;
 - [0015] 使用所述图像传感器捕获所述场景的视频序列,所述视频序列包括图像帧的时间序列;
 - [0016] 使用视频处理路径处理所捕获的视频序列,以形成数字视频文件;
 - [0017] 使用特征识别算法分析所捕获的图像帧,以识别包含感兴趣的特征的所述图像帧的子集;
 - [0018] 形成包括与所捕获的视频序列中的所有图像帧相比更少的图像帧的所述视频摘要,其中,所述视频摘要包括所识别的包含感兴趣的特征的图像帧的子集的至少一部分;
 - [0019] 将所述数字视频文件存储在所述存储设备中;以及
 - [0020] 将所述视频摘要的表示存储在所述存储设备中。
- [0021] 本发明的优点在于,它在捕获视频帧时分析视频帧以确定包含感兴趣的特征的视频帧的子集,从而消除了当创建视频摘要时对解压缩视频数据的需要。
- [0022] 本发明的额外优点在于,它将视频摘要的表示存储在存储设备中,而无需对存储的数字视频序列进行解压缩。这允许在完成视频捕获以后以最小的延迟在数字捕获设备上生成和查看视频摘要。
- [0023] 在一些实施例中,在数字视频文件中使用元数据对视频摘要进行编码,而无需将视频摘要编码为分离的文件。它的优点在于,可以使用理解视频摘要的元数据的“智能”视频播放器方便地查看视频摘要,同时视频摘要的元数据对于传统的播放器是透明的。

附图说明

- [0024] 图 1 是示出了用于形成视频摘要的系统的组件的高层级示意图;
- [0025] 图 2 是用于形成视频摘要的方法的流程图;
- [0026] 图 3 是示出了使用两个不同的视频处理路径来处理数字视频序列的流程图;
- [0027] 图 4 是示出了根据备选实施例的使用两个不同的视频处理路径来处理数字视频序列的流程图;
- [0028] 图 5 是根据第二实施例的用于创建视频摘要的方法的流程图;以及
- [0029] 图 6 是根据第三实施例的用于创建视频摘要的方法的流程图。

具体实施方式

[0030] 在下面的描述中,将按照通常以软件程序实现的方式来描述本发明的优选实施例。本领域技术人员将容易认识到,还可以以硬件构造这种软件的等价物。因为图像处理

算法和系统是公知的,因此本描述将特别针对形成根据本发明的系统和方法的一部分或者与根据本发明的系统和方法更直接协作的算法和系统。在本文中未具体示出或描述的此类算法和系统的其它方面以及用于产生和以其它方式处理与之相关的图像信号的硬件或软件可以从本领域公知的这些系统、算法、组件和元件中选择。考虑到下面的材料中的根据本发明所描述的系统,用于实现本发明的未在本文中具体示出、建议或描述的软件是传统的,并且落入这些领域的普通技术之内。

[0031] 此外,如本文所使用的,用于执行本发明的方法的计算机程序可以存储在计算机可读存储介质中,该计算机可读存储介质可以包括例如:磁存储介质,例如,磁盘(例如,硬盘驱动器或软盘)或磁带;光学存储介质,例如,光盘,光学磁带、或机器可读条形码;固态电子存储设备,例如,随机存取存储器(RAM)或只读存储器(ROM);或者用于存储计算机程序的任何其它物理设备或介质,该计算机程序具有用于控制一个或多个计算机以实施根据本发明的方法的指令。

[0032] 本发明包括本文所描述的实施例的组合。对“特定实施例”等的提及是指存在于本发明的至少一个实施例中的特征。对“实施例”或“特定实施例”等的单独提及不一定是指相同的实施例;然而,除非指示为是互斥的或者本领域技术人员很清楚是互斥的,否则这些实施例并不互斥。使用单数或复数来提及“方法”或“多个方法”等并不是限制性的。应当注意的是,除非上下文另外明确指示或者要求,否则在本公开中以非排他性的意义使用“或者”一词。

[0033] 因为采用成像设备和有关的电路来进行信号捕获和处理并且进行显示的数字照相机是公知的,因此本描述将特别针对形成根据本发明的方法和装置的一部分或者与根据本发明的方法和装置更直接协作的元件。本文未具体示出或描述的元件是从本领域中公知的元件中选择的。以软件提供了要描述的实施例的特定方面。考虑到下面的材料中的根据本发明示出和描述的系统,用于实现本发明的未在本文中具体示出、描述或建议的软件是传统的,并且落入这些领域的普通技术中。

[0034] 本领域技术人员将熟悉关于数字照相机的以下描述。显而易见,存在关于该实施例的可行并且被选择以减小成本、添加特征、或者改善照相机的性能的很多变形。

[0035] 图1描绘了根据本发明包括能够捕获视频图像的数字照相机10的数字摄影系统的框图。优选地,数字照相机10是便携式电池操作设备,其足够小以便用户在捕获和浏览图像时容易握住。数字照相机10产生作为数字图像文件使用存储设备30存储的数字图像。本文使用的短语“数字图像”或“数字图像文件”是指任何数字图像文件,例如,数字静态图像或数字视频文件。

[0036] 在一些实施例中,数字照相机10捕获运动视频图像和静态图像。在其它实施例中,数字照相机10是仅捕获运动视频图像的数字摄像机。数字照相机10还可以包括其它功能,包括但不限于:数字音乐播放器(例如,MP3播放器)、录音设备、移动电话、GPS接收机、或者可编程数字助理(PDA)的功能。

[0037] 数字照相机10包括具有可调整孔径和可调整快门6的镜头4。在优选的实施例中,镜头4是变焦镜头,并且由变焦和聚焦电机驱动器8来控制。镜头4将来自场景(未示出)的光聚焦到诸如单片彩色CCD或CMOS图像传感器等的图像传感器14上。镜头4是一种用于在图像传感器14上形成场景的图像的光学系统。在其它实施例中,光学系统可以使

用具有可变或固定焦点的固定焦距的镜头。

[0038] 图像传感器 14 的输出被模拟信号处理器 (ASP) 和模数 (A/D) 转换器 16 转换为数字形式，并且被临时存储在缓冲存储器 18 中。存储在缓冲存储器 18 中的图像数据接下来由处理器 20 使用存储在固件存储器 28 中的嵌入式软件程序（例如，固件）来处理。在一些实施例中，使用只读存储器 (ROM) 将软件程序永久性地存储在固件存储器 28 中。在其它实施例中，可以使用例如闪存 EPROM 存储器来修改固件存储器 28。在这些实施例中，外部设备可以使用有线接口 38 或无线调制解调器 50 来更新存储在固件存储器 28 中的软件程序。在这些实施例中，固件存储器 28 还可以用于存储图像传感器校准数据、用户设置选择和在照相机被关闭时必须保存的其它数据。在一些实施例中，处理器 20 包括程序存储器（未示出），并且存储在固件存储器 28 中的软件程序在由处理器 20 执行之前被复制到程序存储器中。

[0039] 将理解的是，可以使用单个可编程处理器或者使用多个可编程处理器（包括一个或多个数字信号处理器 (DSP) 设备）来提供处理器 20 的功能。备选地，可以通过定制电路（例如，通过专门设计用于数字照相机的一个或多个定制集成电路 (IC)）或者通过可编程处理器与定制电路的组合来提供处理器 20。将理解的是，可以使用公共数据总线来构成处理器 20 与图 1 中所示的各个组件中的一些或全部之间的连接器。例如，在一些实施例中，可以使用公共数据总线来构成处理器 20、缓冲存储器 18、存储设备 30、以及固件存储器 28 之间的连接。

[0040] 然后，使用存储设备 30 存储经处理的图像。将理解的是，存储设备 30 可以是本领域技术人员已知的任何形式的存储器，包括但不限于：可拆卸闪存卡、内部闪存芯片、磁存储器、或者光学存储器。在一些实施例中，存储设备 30 可以包括内部闪存芯片和针对诸如安全数字 (SD) 卡等的可拆卸闪存卡的标准接口。备选地，可以使用不同的存储卡格式，例如，微型 SD 卡、紧凑型闪存 (CF) 卡、多媒体卡 (MMC)、xD 卡或记忆棒。

[0041] 图像传感器 14 受到定时发生器 12 的控制，其中，定时发生器 12 产生各种时钟信号以选择行和像素并且同步 ASP 和 A/D 转换器 16 的操作。图像传感器 14 可以具有例如 12.4 兆像素 (4088×3040 个像素) 以提供具有近似 4000×3000 个像素的静态图像文件。为了提供彩色图像，通常使用滤色阵列来覆盖图像传感器，滤色阵列提供具有像素阵列的图像传感器，其中，像素阵列包括不同颜色的像素。可以以很多不同的模式来布置不同颜色的像素。举一个例子，可以使用公知的 Bayer 滤色阵列来布置不同颜色的像素，如在 Bayer 的题为“Color imaging array”的共同受让人的美国专利 3,971,065 中所描述的。举第二个例子，可以如 Compton 和 Hamilton 的题为“Image sensor with improved light sensitivity”的共同受让人的美国专利申请公开 2007/0024931 中所述地布置不同颜色的像素。这些示例不是限制性的，并且可以使用很多其它颜色模式。

[0042] 将理解的是，图像传感器 14、定时发生器 12 以及 ASP 和 A/D 转换器 16 可以是单独制造的集成电路，或者它们可以如对于 CMOS 图像传感器通常所做的那样被制造为单个集成电路。在一些实施例中，该单个集成电路可以执行图 1 中所示的其它功能中的一些，包括由处理器 20 提供的功能中的一些。

[0043] 当定时发生器 12 在第一模式中驱动图像传感器 14 时，图像传感器 14 用于提供较低分辨率的传感器图像数据的运动序列，在捕获视频图像时并且还在预览要捕获的静态图

像时使用所述较低分辨率的传感器图像数据的运动序列以编排图像。可以将该预览模式的传感器图像数据作为例如具有 280×720 个像素的 HD 分辨率图像数据来提供, 或者作为例如具有 640×480 个像素 (或者使用与图像传感器的分辨率相比, 具有明显更少的列和行的数据的其它分辨率) 的 VGA 分辨率图像数据来提供。

[0044] 可以通过组合具有相同颜色的相邻像素的值或者通过消除像素值中的一些或者通过组合一些颜色的像素值同时消除其它颜色的像素值, 来提供预览模式的传感器图像数据。可以如 Parulski 等的题为 “Electronic camera for mitiatmg capture of still images while previewing motion images” 的共同受让人的美国专利 6,292,218 所述地处理预览模式的图像数据。

[0045] 当定时发生器 12 在第二模式中驱动图像传感器 14 时, 图像传感器 14 还用于提供高分辨率的静态图像数据。该最终模式的传感器图像数据作为高分辨率输出图像数据来提供, 对于具有高照明水平的场景, 高分辨率输出图像数据包括图像传感器的所有像素, 并且可以例如是具有 4000×3000 个像素的 12 兆像素的最终图像数据。在较低的照明水平, 可以通过“合并读出 (binning)”图像传感器上的某一数量的颜色相似的像素来提供最终的传感器图像数据, 以增加传感器的信号电平并且因此增加 “ISO 速度”。

[0046] 通过处理器 20 提供的控制信号来控制变焦和聚焦电机驱动器 8, 以提供适合的焦距设置并且将场景聚焦到图像传感器 14 上。通过控制可调整孔径和可调整快门 6 的光圈数和曝光时间、经由定时发生器 12 对图像传感器 14 的曝光时段、以及 ASP 和 A/D 转换器 16 的增益 (即, ISO 速度) 设置, 来控制图像传感器 14 的曝光水平。处理器 20 还控制可以照亮场景的闪光灯 2。

[0047] 可以通过使用 “直通镜头 (through-the-lens)” 自动聚焦在第一模式下聚焦数字照相机 10 的镜头 4, 如 Parulski 等在题为 “Electronic Camera with Rapid Automatic Focus of an Image upon a Progressive Scan Image Sensor” 的共同受让人的美国专利 5,668,597 中所描述的。这一点是通过使用变焦和聚焦电机驱动器 8 来将镜头 4 的焦点位置调整到从近焦位置到无穷远焦点位置之间的多个位置来完成的, 而处理器 20 确定提供了图像传感器 14 捕获的图像的中心部分的最高锐度值的最近焦点位置。然后, 与最近焦点位置相对应的焦点距离可以用于多种目的, 例如, 自动地设置适合的场景模式, 并且可以作为元数据与其它镜头和照相机设置一起存储在图像文件中。

[0048] 处理器 20 产生菜单和低分辨率彩色图像, 其被临时存储在显示器存储器 36 中并且被显示在图像显示器 32 上。图像显示器 32 通常是有源矩阵彩色液晶显示器 (LCD), 但是可以使用诸如有机发光二极管 (OLED) 显示器等的其它类型的显示器。视频接口 44 从数字照相机 10 向诸如平板 HDTV 显示器等的视频显示器 46 提供视频输出信号。在预览模式或视频模式中, 由处理器 20 处理来自缓冲存储器 18 的数字图像数据以形成通常作为彩色图像显示在图像显示器 32 上的一系列运动预览图像。在浏览模式中, 使用来自存储在存储设备 30 中的数字图像文件的图像数据来产生显示在图像显示器 32 上的图像。

[0049] 响应于用户控制 34 提供的用户输入来控制显示在图像显示器 32 上的图形用户界面。用户控制 34 用于选择各种照相机模式, 例如, 视频捕获模式、静态捕获模式和浏览模式, 并且用于发起对静态图像的捕获和对运动图像的录制。在一些实施例中, 当用户部分地按压快门按钮 (其是用户控制 34 之一) 时, 发起上述第一模式 (即, 静态预览模式), 并且

当用户完全按压快门按钮时,发起第二模式(即,静态图像捕获模式)。用户控制34还用于打开照相机,控制镜头4,并且发起拍照过程。用户控制34通常包括按钮、摇臂开关、操纵杆或旋转拨号盘的某种组合。在一些实施例中,通过在图像显示器32上使用触摸屏覆盖来提供用户控制34中的一些。在其它实施例中,可以使用额外的状态显示器或图像显示器。

[0050] 可以使用用户控制34选择的照相机模式包括“定时器”模式。当选择“定时器”模式时,在用户完全按压快门按钮以后、在处理器20发起对静态图像的捕获之前,发生短暂延迟(例如,10秒)。

[0051] 连接到处理器20的音频编解码器22从麦克风24接收音频信号,并且向扬声器26提供音频信号。这些组件可以用于录制和回放音轨以及视频序列或静态图像。如果数字照相机10是诸如组合照相机和移动电话等的多功能设备,则麦克风24和扬声器26可以用于打电话。

[0052] 在一些实施例中,扬声器26可以作为用户界面的一部分被使用,以例如提供各种可听信号,这些可听信号指示用户控制已经被按压,或者特定的模式已经被选择。在一些实施例中,麦克风24、音频编解码器22和处理器20可以用于提供语音识别,使得用户可以通过使用语音命令而不是用户控制34来向处理器20提供用户输入。扬声器26也可以用于向用户通知呼入电话。这可以通过使用存储在固件存储器28中的标准铃声或者通过使用从无线网络58下载并且存储在存储设备30中的定制铃声来完成。此外,振动设备(未示出)可以用于提供关于呼入电话的静默(例如,不可听)通知。

[0053] 在一些实施例中,数字照相机10还包含加速度计27,加速度计27提供与照相机的任何运动有关的数据。优选地,加速度计27检测到针对三个正交方向中的每一个的线性加速度和旋转加速度(总共6个维度的输入)。

[0054] 处理器20还提供了对来自图像传感器14的图像数据的额外处理,以产生渲染的sRGB图像数据,该sRGB图像数据被压缩并且存储在存储设备30中的诸如公知的Exif-JPEG图像文件等的“完成的”图像文件中。

[0055] 数字照相机10可以经由有线接口38连接到接口/充电器48,接口/充电器48被连接到计算机40,计算机40可以是位于家中或办公室中的台式计算机或便携式计算机。有线接口38可以符合例如公知的USB2.0接口规范。接口/充电器48可以经由有线接口38向数字照相机10中的一组可再充电电池(未示出)供电。

[0056] 数字照相机10可以包括无线调制解调器50,无线调制解调器50在射频频带52上与无线网络58接口。无线调制解调器50可以使用各种无线接口协议,例如,公知的蓝牙无线接口或公知的802.11无线接口。计算机40可以经由互联网70向诸如柯达EasyShare画廊等的照片服务提供商72上传图像。其它设备(未示出)可以访问由照片服务提供商72存储的图像。

[0057] 在备选的实施例中,无线调制解调器50通过射频(例如,无线)链路与诸如3GSM网络等的与互联网70连接的移动电话网络(未示出)进行通信,以上传来自数字照相机10的数字图像文件。可以将这些数字图像文件提供给计算机40或照片服务提供商72。

[0058] 现在参照图2来描述本发明。首先,在捕获视频序列步骤200中使用诸如数字照相机10(图1)等的数字视频捕获设备以使用图像传感器14来捕获数字视频序列205,其中,数字视频序列205包括图像帧的时间序列。

[0059] 在捕获视频序列步骤 200 中, 执行捕获参考图像步骤 210 以使用图像传感器 14 捕获参考图像 215, 其中, 参考图像 215 包含特定人。参考图像 215 可以是来自捕获的数字视频序列 205 的图像帧, 并且可以通过使用用户控制 34 来选择参考图像 215。例如, 在视频捕获期间, 用户可以通过按压适合的用户控制按钮来请求捕获参考图像 215, 其中, 用户控制按钮将信号发送到处理器 20 以将缓冲存储器 18 中的当前视频帧指定为参考图像 215。参考图像 215 包含感兴趣的特定人。在优选的实施例中, 参考图像 215 包含足够的空间分辨率的特定人的正面, 以允许根据参考图像 215 来确定面部特征。在一些实施例中, 参考图像 215 仅包含单个人, 并且特定人被识别为参考图像 215 中唯一的人。在其它实施例中, 参考图像 215 可以包含多个人, 并且可以通过很多方式中的任意一种来选择特定人, 这些方式包括: 选择最大的人, 选择最接近参考图像 215 的中心的人, 选择具有最大可视面部的人, 或者使用适合的用户界面(例如, 使用户使用指向设备来选择面部)来手动地选择这个人。备选地, 参考图像 215 中的每一个人可以被指定为特定人。在一些实施例中, 可以通过将参考图像 215 中存在的面部与图像识别数据库中的已知面部进行比较并且选择已知的面部来选择特定人。在其它实施例中, 可以提供用户界面以使用户能够手动地指定包含在参考图像 215 中的特定人。

[0060] 处理捕获的视频序列步骤 220 使用视频处理路径来处理捕获的数字视频序列 205 以形成数字视频文件 225。该视频处理路径可以包括例如视频压缩步骤。例如在 MPEG 和 H.263 标准中规定的视频压缩算法对于本领域技术人员而言是公知的。

[0061] 在捕获数字视频序列 205 期间, 分析捕获的视频帧步骤 240 使用人识别算法来分析捕获的数字视频序列 205 中的图像帧, 以识别包含特定人的图像帧的子集 245。在优选的实施例中, 人识别算法可以是面部识别算法, 并且分析捕获的图像帧步骤 240 识别包含与参考图像中的特定人的面部相同的面部的图像帧。面部识别算法在本领域中是公知的。例如, Turk 等在文章“Eigenfaces for Recognition”(Journal of Cognitive Neuroscience, Vol. 3, pp. 71–86, 1991) 中描述了可以根据本发明使用的面部识别算法。备选地, 人识别算法可以是将图像帧中的人与参考图像 215 中的特定人进行匹配的任意算法。这种算法可以包括性别分类、高度估计和衣服分析的步骤, 并且可以从本领域技术人员公知的此类算法中选择这种算法。

[0062] 形成视频摘要步骤 250 形成包括与捕获的数字视频序列 205 中的所有图像帧相比更少的图像帧的视频摘要 255, 其中, 视频摘要 255 包括所识别的包含特定人的图像帧的子集 245 的至少一部分。在本发明的一个实施例中, 仅使用包含特定人的那些图像帧来形成视频摘要 255。

[0063] 在一些实施例中, 视频摘要 255 仅包括包含特定人的图像帧的子集 245 的一部分。例如, 视频摘要 255 可以被限制为包括与预定数量的图像帧相比更少的图像帧。在备选的实施例中, 视频摘要 255 可以包括来自图像帧的子集 245 中的每一组连续图像帧的单个图像帧。通过这种方式, 视频摘要 255 可以是由从所识别的包含特定人的图像帧的子集 245 中选择的静态图像集合构成的“幻灯片”。

[0064] 在另一实施例中, 除了所识别的包含特定人的图像帧的子集 245 以外, 视频摘要 255 还包括额外的图像帧。在本发明的一个实施例中, 额外的图像帧包括紧邻所识别的包含特定人的图像帧的子集 245 中的图像帧之前或之后的图像帧。这些帧可以被选择作为例如

过渡帧,以允许从视频摘要 255 的一部分向另一部分的一段过渡。这些帧还可以被选择为使得视频摘要 255 包括可以容易地从数字视频文件 225 中提取出的画面组。诸如 MPEG 等的视频压缩标准对视频序列进行编码,使得(在无需参考其它帧的情况下)独立地对一些帧进行编码,并且在无需参照画面组以外的任何帧的情况下对具有按时间顺序的帧的一些组或者画面组进行编码。因此,可以在无需对压缩的视频数据进行解码的情况下从压缩的数字视频文件 225 中提取出表示该画面组的压缩的视频数据。

[0065] 在本发明的另一实施例中,额外的图像帧包括捕获的数字视频序列 205 中被确定为重要部分的其它部分。可以通过执行关键帧提取或者视频摘要算法来识别数字视频序列 205 的这些其它重要部分。在 Deever 的于 2011 年 12 月 1 日公开的题为“Method for determining key video frames”的共同受让人的、未决的美国申请公开 No. 2011/0292288 以及 Deever 的题为“Video summary method and system”的共同受让人的、未决的美国申请公开 No. 2011/0293018 中描述了这些算法。

[0066] 在美国申请公开 No. 2011/0293018 中,公开了一种用于形成视频摘要的方法,在该方法中,当进行捕获时对图像帧进行分析以确定特征值。在无需对压缩的数字视频序列进行解压缩的情况下对这些特征值进行分析,以识别包括视频摘要的关键视频片段。

[0067] 在美国申请公开 No. 2011/0292288 中,公开了一种用于确定关键视频片段的方法,在该方法中,使用在捕获时确定的特征值来分析数字视频序列,以确定作为时间的函数的重要性值。重要性值用于形成数字视频序列的扭曲(warped)时间表示。扭曲时间表示向数字视频序列中的一些视频帧提供较大的时间权重并且向其它视频帧提供较小的时间权重。扭曲时间数字视频序列表示被划分为相同时间间隔的集合。通过分析与相应的时间间隔内的视频帧相关联的特征值来从每一个时间间隔选择关键视频帧。这种标准可以包括选择在完成放大过程以后立即出现的关键视频帧,或者选择在视频帧的中心区域中具有适当的局部运动水平的关键视频帧。

[0068] 在一些实施例中,可以通过选择在每一个关键视频帧周围的视频帧集合来形成关键视频片段。例如,可以通过选择在关键视频帧之前和之后的两秒的视频帧来形成关键视频片段,从而形成长度为四秒的关键视频片段。

[0069] 备选地,可以对关键视频帧进行排序,并且仅针对与最高排序的关键视频帧相对应的关键视频帧的子集形成关键视频片段。对关键视频帧进行排序可以包括分析数字视频序列以确定照相机定影模式,并且响应于照相机定影模式来对关键视频帧进行排序。在捕获时确定的与全局运动相对应的特征值可以提供全局运动轨迹,其指示照相机在整个数字视频序列期间的定影区域。与较高定影的区域(也即是说,针对整个视频捕获的相当大的比例,照相机保持定影的区域)相对应的视频帧被排序地较高。可以迭代地执行排序过程,在每一个步骤,选择次高排序的关键视频帧,并且在每一个步骤,该过程可以使表示未表示的定影区域的关键视频帧处于表示已经包含在排序过程中的定影区域的关键视频帧之上。一旦已经对关键视频帧进行了排序,就可以选择排序最高的关键视频帧以包含在关键视频片段中。由这些关键视频片段表示的额外图像帧可以与所识别的包含特定人的图像帧的子集进行组合以形成视频摘要。

[0070] 存储数字视频文件步骤 230 将数字视频文件 225 存储到存储设备 30,从而产生存储的数字视频文件 235。存储视频摘要的表示步骤 260 将视频摘要的表示 265 存储在存储

设备 30 中。在一个实施例中,将视频摘要的表示 265 存储在存储设备 30 中包括存储帧标识元数据,其提供关于数字视频序列 205 中与视频摘要 255 相对应的图像帧的指示。可以与存储的数字视频文件 235 相关联地存储帧标识元数据。例如,可以将视频摘要的表示 265 作为视频摘要元数据存储在存储的数字视频文件 235 中,其指示与包含在视频摘要中的片段相对应的一系列起始帧和结束帧。这允许除了存储帧标识元数据所需的较小的存储量以外,在无需使用任何额外的物理存储器的情况下存储视频摘要的表示 265。然后,可以使用理解视频摘要元数据的“智能”视频播放器播放视频摘要 255,同时视频摘要元数据对于传统的播放器是透明的。

[0071] 在另一实施例中,视频摘要的表示 265 是摘要数字视频文件。在该实施例中,视频摘要 255 作为与存储的数字视频文件 235 分离的数字视频文件被存储。在该情况下,视频摘要的表示 265 是可以独立于存储的数字视频文件 235 查看或共享的摘要数字视频文件。在优选的实施例中,存储的视频摘要的表示 265 是具有适合于使用标准视频播放器播放的格式的摘要数字视频文件。

[0072] 在一些实施例中,可以在完成对视频序列的捕获以后创建摘要数字视频文件。可以从存储的数字视频文件 235 中提取出视频摘要的期望帧。如果视频摘要的期望帧被选择为与单独编码的画面组相对应,则可以通过在无需对压缩的视频数据进行解码的情况下提取出与期望帧相对应的压缩数据来创建摘要数字视频文件。

[0073] 在一些实施例中,摘要数字视频文件是使用与用于形成数字视频文件 225 的视频处理路径不同的视频处理路径,从捕获的数字视频序列 205 的图像帧中形成的。例如,很多视频捕获设备可以同时捕获和编码两个视频流。在这些设备中,单个图像传感器 14 用于捕获每一个图像帧。然后,使用两个不同的处理路径来处理每一个图像帧。一个处理路径可以产生数字视频文件 225。第二处理路径可以产生用于存储视频摘要 255 的摘要数字视频文件。第二处理路径与第一处理路径的不同之处在于,它可以生成每一个图像帧的空间分辨率减小的版本,并且对较低分辨率的视频进行编码。很多视频捕获设备可以同时捕获 1080p 分辨率的视频和 QVGA 分辨率的视频。

[0074] 图 3 示出了具有两个处理路径的视频捕获过程。捕获的图像帧 310 去往第一视频处理路径 320 和第二视频处理路径 330。第一视频处理路径 320 包括第一视频编码器 340,其提供第一空间分辨率的第一编码图像帧 350。第二视频处理路径 330 包括可选的图像调整器 360,其产生具有第二空间分辨率的修改的图像帧 370。第二视频处理路径 330 还包括第二视频编码器 380,其对修改的图像帧 370 进行编码,从而提供第二编码图像帧 390。本领域技术人员将认识到,第一视频处理路径 320 和第二视频处理路径 330 也可以可选地包括其它视频处理步骤 300,例如,色度和色调缩放处理、噪声减小、锐化以及图像稳定化。虽然这些其它视频处理步骤 300 被示出为被首先应用,但是应当认识到,可以在第一视频处理路径 320 和第二视频处理路径 330 中的任意一点处应用这些视频处理步骤 300,或者这些视频处理步骤 300 甚至穿插在其它步骤之间。本领域技术人员还将认识到,图 3 示出了两个视频处理路径可能不同的一种可能的方式,并且落入本发明的范围内的其它差别是可能的。

[0075] 通过使用与用于形成数字视频文件 225(图 2)的视频处理路径不同的视频处理路径,可以与数字视频文件 225 同时创建摘要数字视频文件。图 4 示出了图 3 中所示的方法

的变形,其可以用于同时创建两个视频文件。与图3一样,每一个捕获的图像帧310通过包括第一视频编码器340的第一视频处理路径320以产生第一编码图像帧350。第一编码图像帧350被包含在数字视频文件225中。

[0076] 捕获的图像帧310还通过第二视频处理路径430。除了添加了图像包含特定人的测试460以外,第二视频处理路径430与图3中的第二视频处理路径330类似。图像包含特定人的测试460使用人识别算法来分析捕获的图像帧310,以确定它是否包含来自参考图像215的特定人。如果否,则使用丢弃图像步骤470来丢弃捕获的图像帧310。如果特定人存在于捕获的图像帧310中,则处理前进至图像调整器360以确定修改的图像帧370,并且前进至第二视频编码器380以提供第二编码图像帧390。第二编码图像帧390包含在摘要数字视频文件中。在备选的实施例中,图像调整器360可以在图像包含特定人的测试460之前较早地应用于第二视频处理路径中。在其它实施例中,图像调整器360可能根本不存在于第二视频处理路径430中。

[0077] 图4中所示的第一视频处理路径320和第二视频处理路径430用于生成两个分离的视频文件。第一视频文件(数字视频文件225)包含所有捕获的图像帧310并且以第一空间分辨率被编码。第二视频文件(摘要数字视频文件)仅包括包含感兴趣的特定人的那些捕获的图像帧310,并且以第二空间分辨率被编码。第二视频文件与第一视频文件同时形成。

[0078] 备选地,摘要数字视频文件可以仅部分地与数字视频文件同时形成。可以如图4所示地创建初始摘要数字视频文件。在捕获完成以后,可以使用从存储的数字视频文件235中提取的额外数据来增加摘要数字视频文件。在该情况下,摘要数字视频文件可以扩展为包含最初未通过第二视频处理路径编码的图像帧。与在捕获以后形成整个摘要数字视频文件相比,这允许更快速地生成摘要数字视频文件。

[0079] 图5示出了本发明的备选的实施例。在该情况下,不是在与捕获视频序列步骤200的同时捕获包含特定人的参考图像215,而是在独立的捕获参考图像步骤510中捕获包含特定人的参考图像215,其中,可以在捕获视频序列步骤200之前或之后执行该独立的捕获参考图像步骤510。参考图像215可以是在静态捕获模式中获得的单个图像,或者可以是来自现有的捕获的数字视频序列的图像帧。在一些实施例中,数字照相机可以存储与数字照相机的所有者感兴趣的人的集合相对应的人图像的集合(例如,可以针对每一个家庭成员存储人图像),并且可以使用适合的用户界面(例如,预定的人的名字的列表)从人图像的集合中选择参考图像215。可以在数字照相机10上捕获参考图像215,或者备选地,可以在分离的图像捕获设备上捕获参考图像215并且将参考图像215输入数字照相机10中。

[0080] 在一些实施例中,可以指定包含特定人的不同视图的多个参考图像215。分析捕获的图像帧步骤240可以使用多个参考图像215来帮助更可靠地确定图像帧是否包含特定人。

[0081] 在一些实施例中,分析捕获的图像帧步骤240与捕获视频序列步骤200同时发生,使得在生成数字视频文件225的同时形成视频摘要255。在其它实施例中,可以使用后处理操作来形成视频摘要255,其中,后处理操作是在数字视频文件225已经被捕获和存储以后执行的。在一些情况下,可以在数字照相机10上执行分析捕获的图像帧步骤240。备选地,还可以在已经将存储的数字视频文件235装载到诸如主机计算机等的某一其它设备上

以后执行分析捕获的图像帧步骤 240。图 5 中的剩余元素与图 2 的实施例中所示的元素相同。

[0082] 现在将参照图 6 来描述本发明的备选实施例。在该情况下,不是基于参考图像 215 分析捕获的图像帧,而是分析捕获的图像帧步骤 640 基于指示感兴趣的特征的参考数据的集合 615 来选择图像帧的子集 245,其中,参考数据的集合 615 是在指定参考数据步骤 610 中指定的。图 6 的方法中的剩余元素与图 2 的实施例中所示的元素相同。

[0083] 该参考数据 615 可以具有很多形式。在一些实施例中,参考数据 615 可以是感兴趣的特征的文本描述。例如,参考数据 615 可以是个人的名字、对象、位置或事件。在一些实施例中,参考数据 615 可以是感兴趣的特征的数学表示。例如,参考数据 615 可以是彩色柱状图、特征值、模板、或者任何其它特征向量。本领域技术人员将认识到,在本发明的范围内存在可以用于表示图像信息的多种方法。在一些实施例中,参考数据 615 可以与非图像信息相关联。例如,参考数据 615 可以是与以下各项相关联的信息:音频信息、全球定位系统 (GPS) 数据、自动聚焦数据、自动曝光数据、自动白平衡数据、缩放镜头数据、加速度计数据、陀螺仪数据或者红外线传感器数据。本领域技术人员将认识到,存在可以作为参考数据 615 提供的各种类型的信息,所述参考数据 615 由分析捕获的图像帧步骤 640 使用以识别图像帧的子集 245。

[0084] 在捕获数字视频序列 205 期间,分析捕获的图像帧步骤 640 使用特征识别算法分析数字视频序列 205 中的图像帧,以识别包含感兴趣的特征的图像帧的子集,其中,感兴趣的特征由参考数据 615 指定。例如,如果参考数据 615 指定的感兴趣的特征是先前已经指定了其参考面部图像的特定名字的人,则可以使用面部识别算法以通过参照参考面部图像分析图像帧来确定图像帧是否包含特定名字的人。在另一示例中,如果参考数据 615 指定的感兴趣的特征是事件标签(例如,“高尔夫挥杆动作”),则可以使用特征识别算法以确定图像帧是否与指定的事件相对应。在一些情况下,特征识别算法可能需要分析一系列图像帧以确定适合的事件标签(例如,检测哪些图像帧包含高尔夫挥杆动作的运动特征)。在一示例中,如果参考数据 615 指定的感兴趣的特征是对象标签(例如,“鲸鱼”),则可以使用特征识别算法以确定图像帧是否包含指定的对象。所有这些示例都是可以在捕获时识别感兴趣的特征以确定要包含在视频摘要中的图像帧的子集 245 的情形。

[0085] 在一些实施例中,参考数据 615 指定的感兴趣的特征可以与同图像帧相关联的一般特征相对应。例如,参考数据 615 指定的感兴趣的特征可以是具有低关联的照相机运动的图像帧。在该情况下,特征识别算法可以用于分析陀螺仪数据、加速度计数据、或者基于图像的运动估计数据以识别满足指定的照相机运动条件的图像帧。类似地,参考数据 615 指定的感兴趣的特征可以是经过照相机缩放过程以后的图像帧。在该情况下,特征识别算法可以用于分析缩放镜头数据或基于图像的缩放估计数据,以识别刚刚完成照相机缩放之后的图像帧。在另一示例中,参考数据 615 指定的感兴趣的特征可以是在场景中具有较大幅度的对象运动的图像帧。在该情况下,特征识别算法可以用于量化场景中的对象运动的量以识别满足对象运动条件的图像帧。本领域技术人员将认识到,这些仅是落入本发明的范围内的很多感兴趣的特征和特征识别算法的示例。

[0086] 计算机程序产品可以包括一个或多个存储介质,例如:磁存储介质,如磁盘(例如,软盘)或磁带;光存储介质,例如,光盘、光学磁带、或者机器可读条形码;固态电子存储

设备,例如,随机存取存储器 (RAM)、或只读存储器 (ROM);或者用于存储计算机程序的任何其它物理设备或介质,所述计算机程序具有用于控制一个或多个计算机以实施根据本发明的方法的指令。

- [0087] 部件列表
- [0088] 2 闪光灯
- [0089] 4 镜头
- [0090] 6 可调整孔径和可调整快门
- [0091] 8 对焦和聚焦电机驱动器
- [0092] 10 数字照相机
- [0093] 12 定时发生器
- [0094] 14 图像传感器
- [0095] 16ASP 和 A/D 转换器
- [0096] 18 缓冲存储器
- [0097] 20 处理器
- [0098] 22 音频编解码器
- [0099] 24 麦克风
- [0100] 25 压力传感器
- [0101] 26 扬声器
- [0102] 27 加速度计
- [0103] 28 固件存储器
- [0104] 30 存储设备
- [0105] 32 图像显示器
- [0106] 34 用户控制
- [0107] 36 显示器存储器
- [0108] 38 有线接口
- [0109] 40 计算机
- [0110] 42 俯仰传感器
- [0111] 44 视频接口
- [0112] 46 视频显示器
- [0113] 48 接口 / 充电器
- [0114] 50 无线调制解调器
- [0115] 52 射频频带
- [0116] 58 无线网络
- [0117] 70 互联网
- [0118] 72 照片服务提供商
- [0119] 200 捕获视频序列步骤
- [0120] 205 数字视频序列
- [0121] 210 捕获参考图像步骤
- [0122] 215 参考图像

- [0123] 220 处理捕获的视频序列步骤
- [0124] 225 数字视频文件
- [0125] 230 存储数字视频文件步骤
- [0126] 235 存储的数字视频文件
- [0127] 240 分析捕获图像帧步骤
- [0128] 245 图像帧的子集
- [0129] 250 形成视频摘要步骤
- [0130] 255 视频摘要
- [0131] 260 存储视频摘要的表示步骤
- [0132] 265 视频摘要的表示
- [0133] 300 其它视频处理步骤
- [0134] 310 捕获的图像帧
- [0135] 320 第一视频处理路径
- [0136] 330 第二视频处理路径
- [0137] 340 第一视频编码器
- [0138] 350 第一编码图像帧
- [0139] 360 图像调整器
- [0140] 370 修改的图像帧
- [0141] 380 第二视频编码器
- [0142] 390 第二编码图像帧
- [0143] 430 第二视频处理路径
- [0144] 460 图像包含特定人的测试
- [0145] 470 丢弃图像步骤
- [0146] 510 捕获参考图像步骤
- [0147] 610 指定参考数据步骤
- [0148] 615 参考数据
- [0149] 640 分析捕获的图像帧步骤

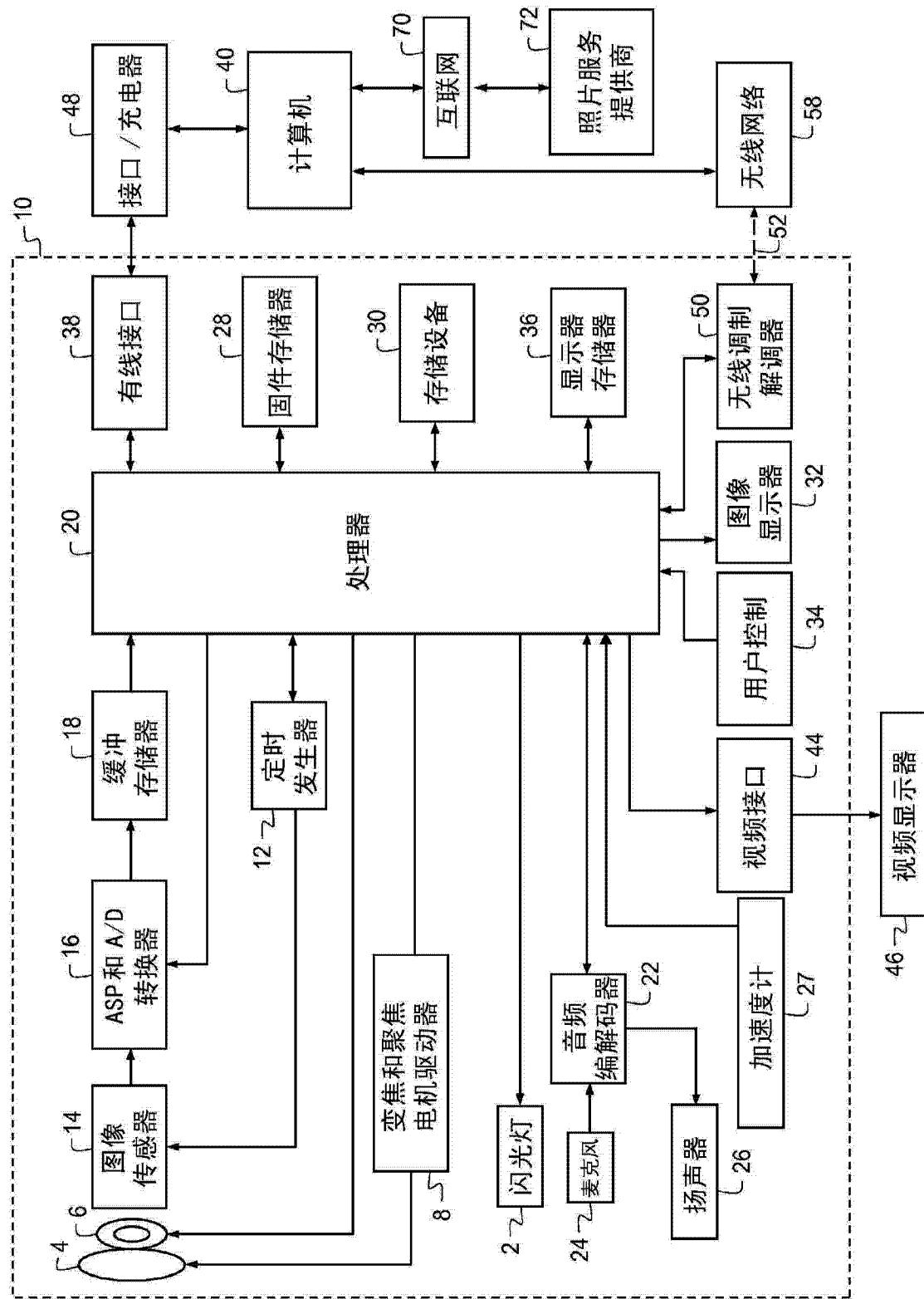


图 1

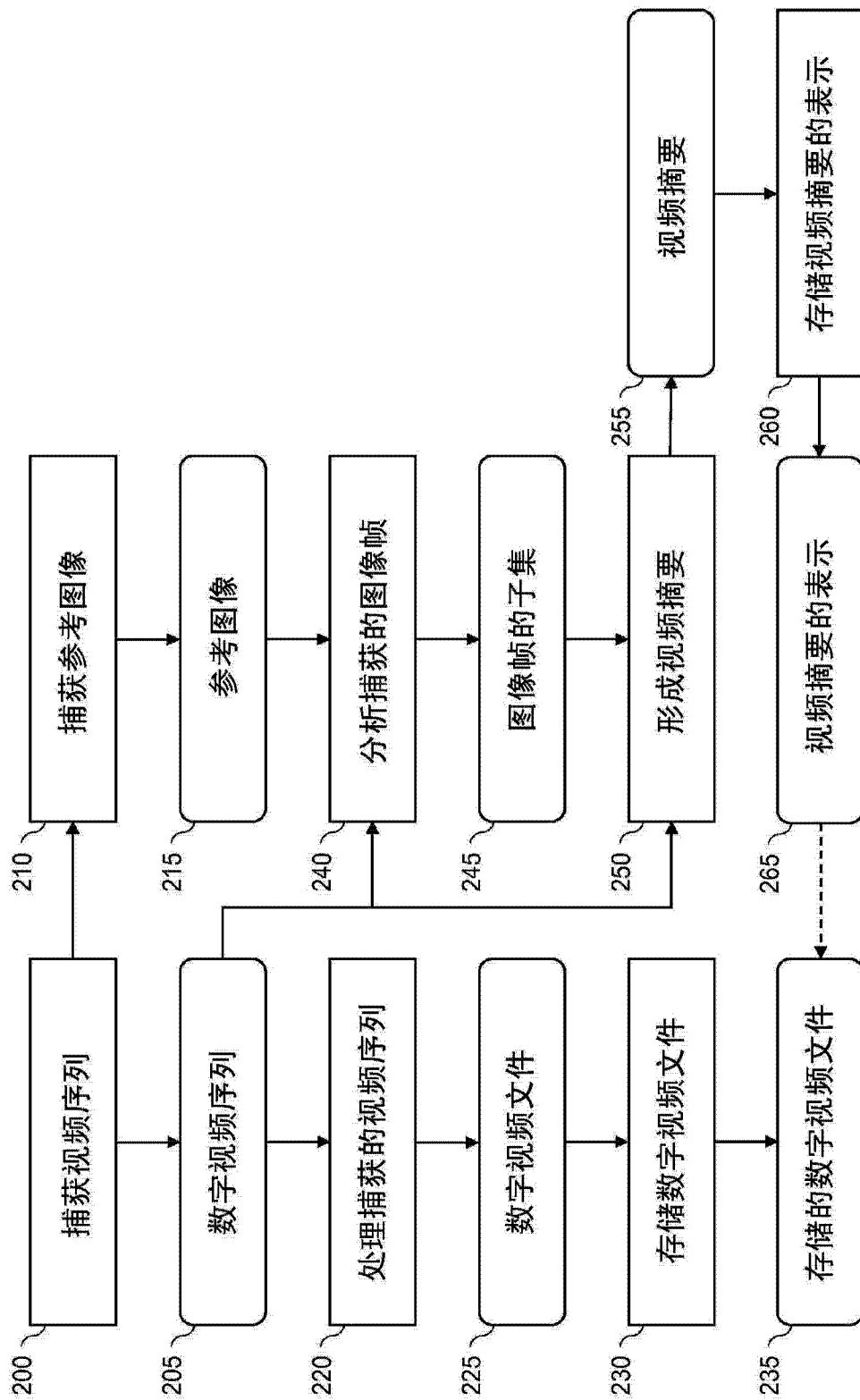


图 2

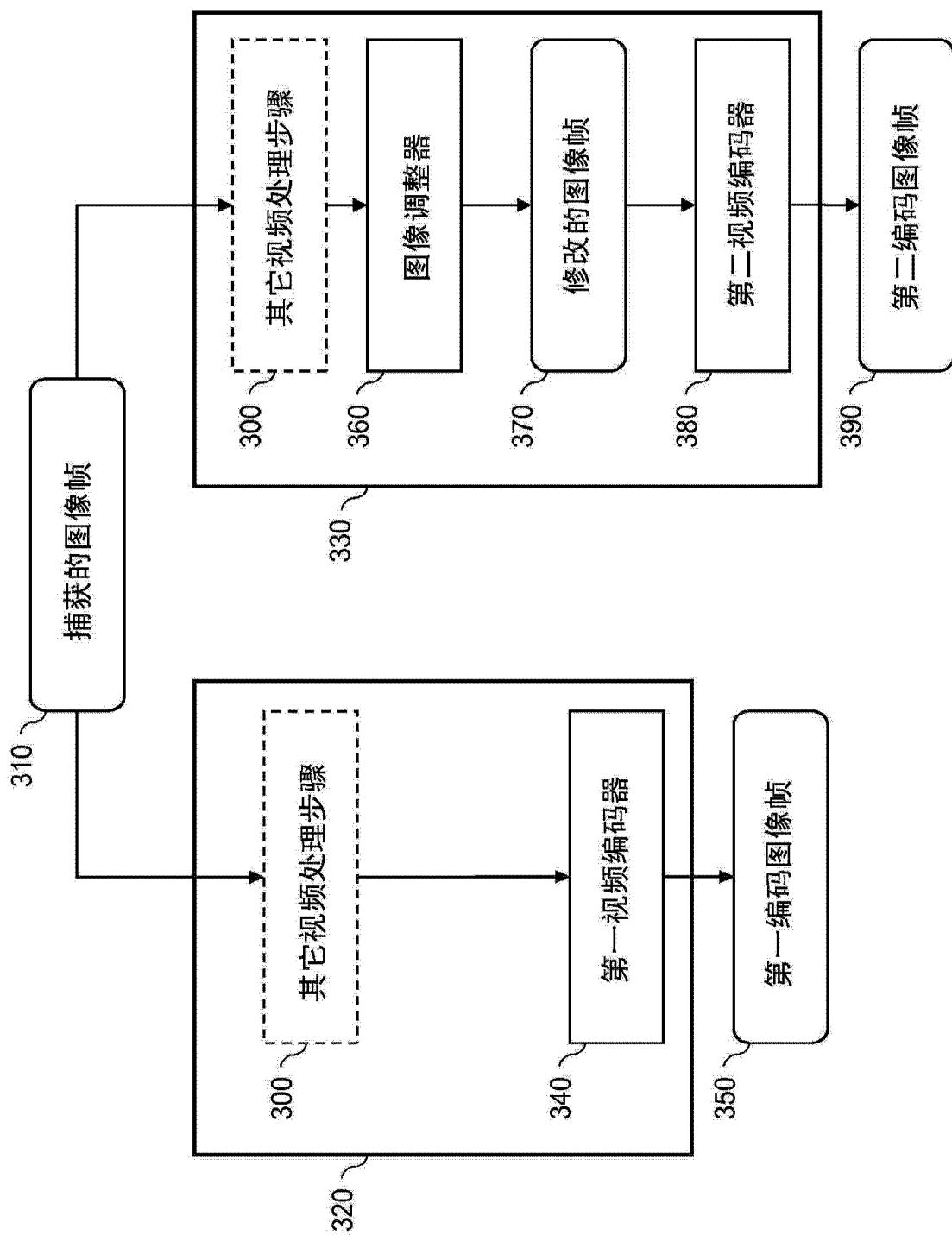


图 3

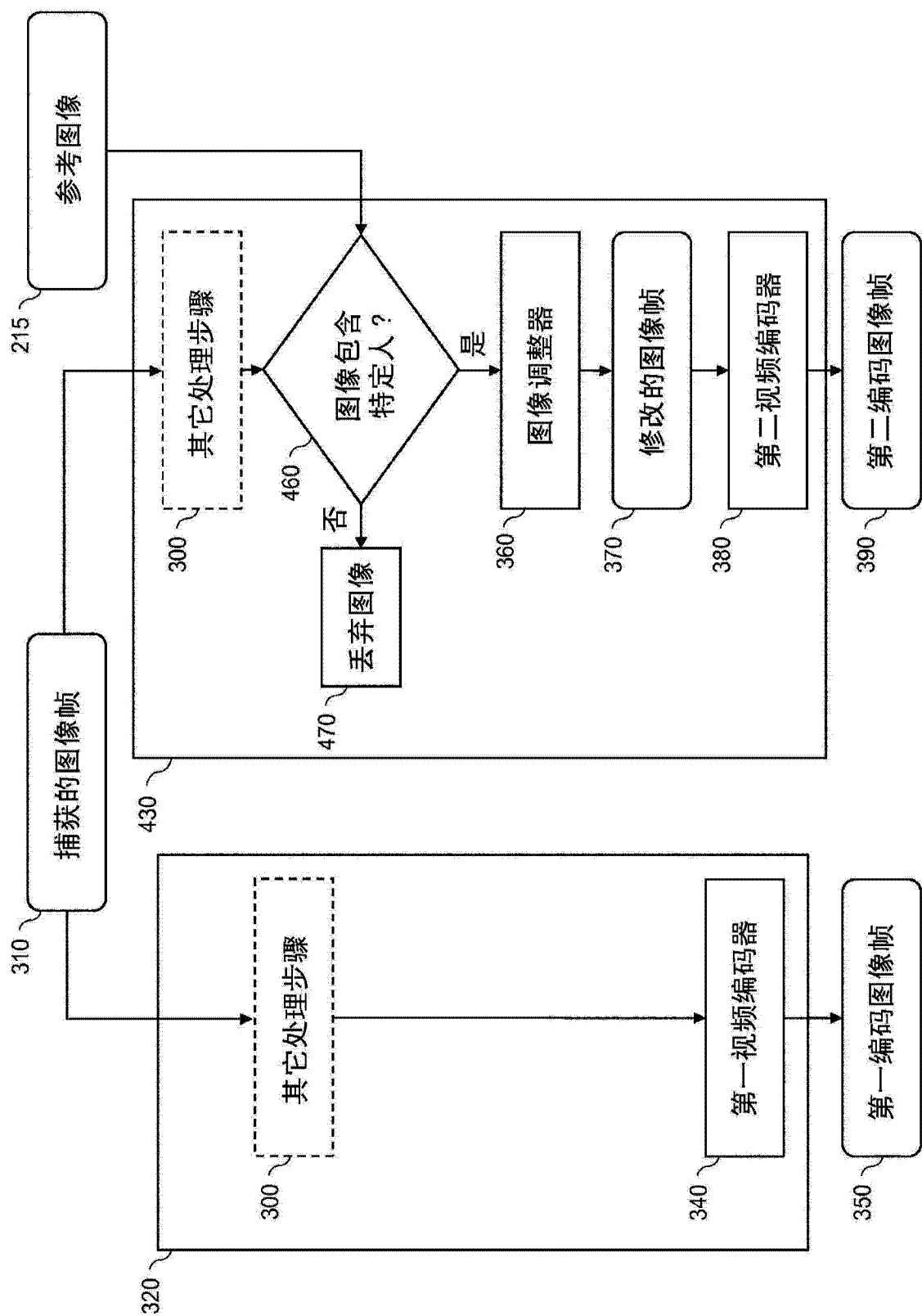


图 4

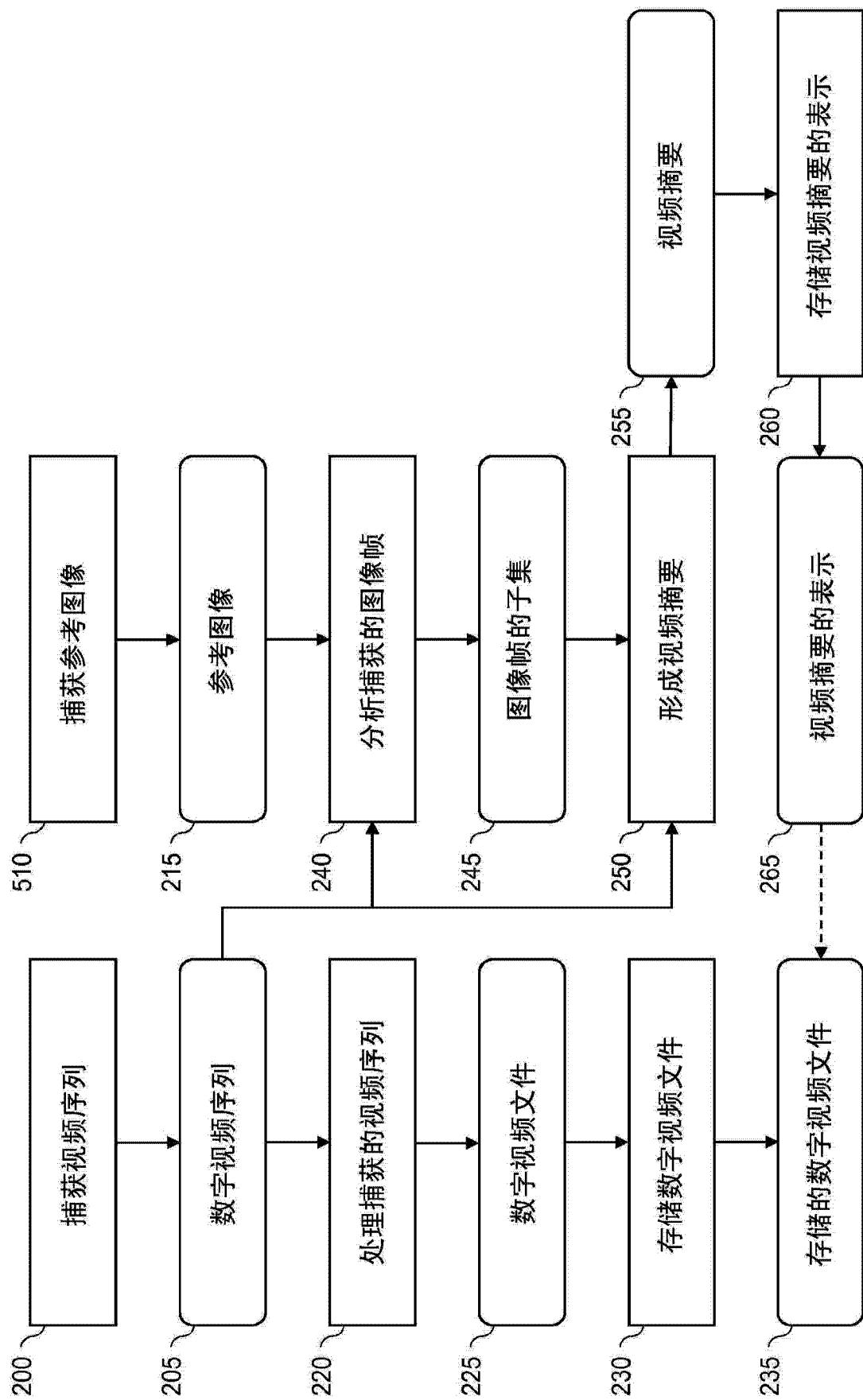


图 5

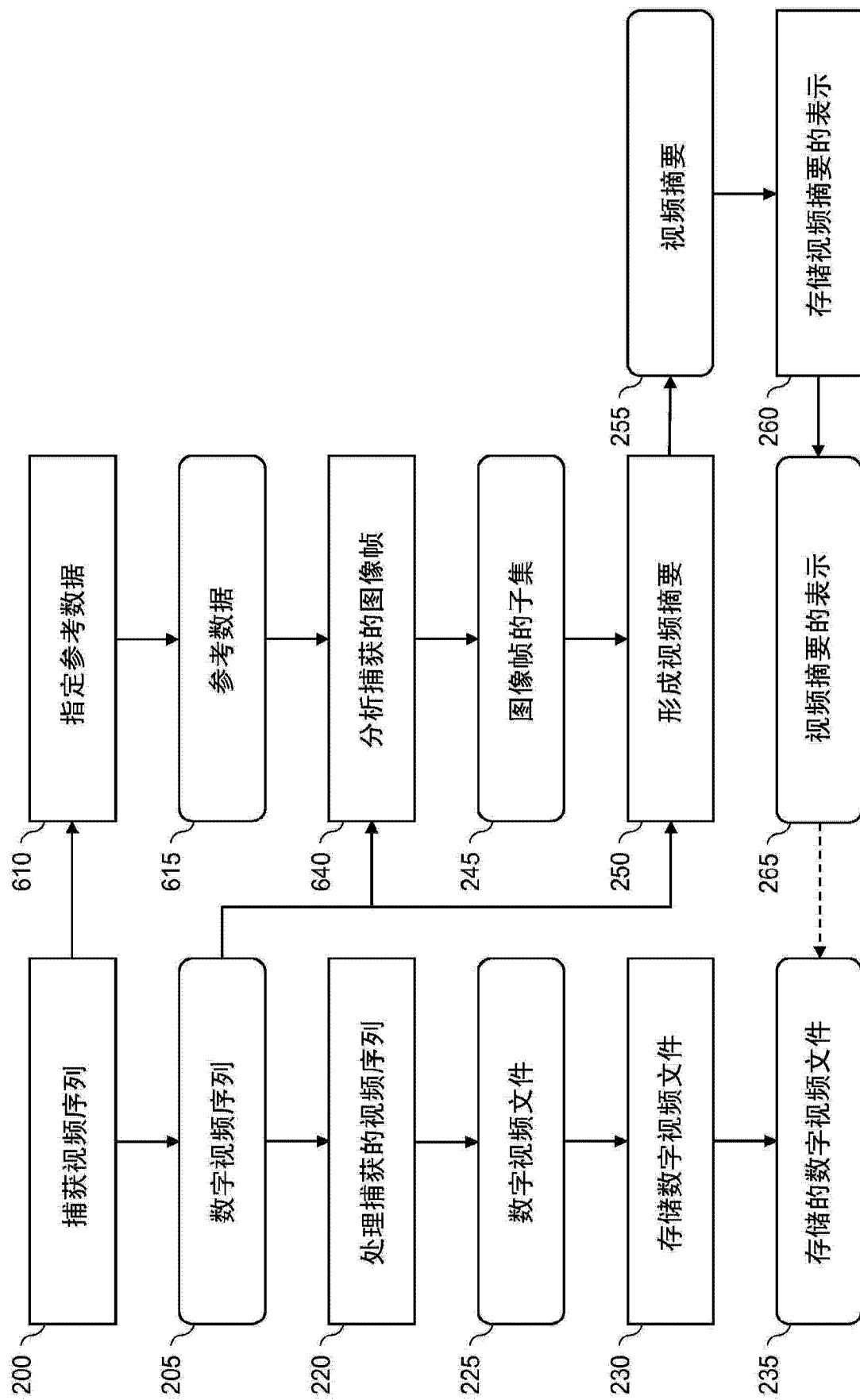


图 6