(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108140032 B
(45)授权公告日 2022.03.11

(21)申请号 201680056638.4
(22)申请日 2016.09.08
(65)同一申请的已公布的文献号
申请公布号 CN 108140032 A
(43)申请公布日 2018.06.08
(30)优先权数据
14/925,701 2015.10.28 US
(85)PCT国际申请进入国家阶段日
2018.03.28
(86)PCT国际申请的申请数据
PCT/US2016/050714 2016.09.08
(87)PCT国际申请的公布数据
WO2017/074562 EN 2017.05.04
(73)专利权人 英特尔公司
地址 美国加利福尼亚州
(72)发明人 R·Y·耶海兹凯罗厄卡 G·科伦
(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 高见 黄嵩泉
(51)Int.Cl.
G06F 16/738(2019.01)
G06V 20/40(2022.01)
G06V 10/764(2022.01)
G06V 10/82(2022.01)
(56)对比文件
CN 103440640 A,2013.12.11
CN 101299241 A,2008.11.05
CN 104520875 A,2015.04.15
CN 103533237 A,2014.01.22
CN 104182421 A,2014.12.03
US 2012106925 A1,2012.05.03
US 2013293776 A1,2013.11.07
US 2013293776 A1,2013.11.07
审查员 窦广健

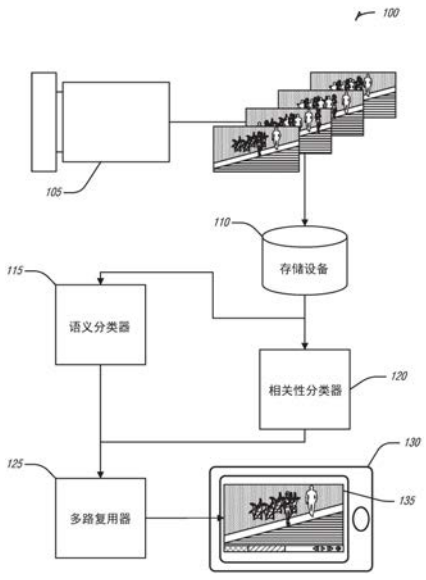
权利要求书2页 说明书27页 附图11页

(54)发明名称

用于自动视频概括的设备和方法

(57)摘要

本文描述了用于自动视频概括的系统和技术。可获得视频并且可从该视频的帧生成该视频的语义模型。可向帧指派相应的相关性评分。可使用该相应的相关性得分来初始化语义模型。该语义模型随后被迭代地处理以产生该视频的子场景,该子场景的集即为视频概括。



CN 108140032 B

1.一种用于自动视频概括的设备,其特征在于,所述设备包括:

存储设备,用于保存视频;

语义分类器,用于从所述视频的帧生成所述视频的语义模型;

相关性分类器,用于向所述帧指派相应的相关性得分;

多路复用器,用于:

使用所述相应的相关性得分初始化所述语义模型;以及

迭代地处理所述语义模型以产生子场景集合,每次迭代包括所述多路复用器,用于:

在初始化之后收敛所述语义模型;

在收敛之后选择具有最高相关性得分的帧序列;以及

通过固定所选帧序列的所述相关性得分来重新初始化所述语义模型。

2.如权利要求1所述的设备,其特征在于,生成所述语义模型包括提取所述帧的诸特征。

3.如权利要求2所述的设备,其特征在于,生成所述语义模型包括从所提取的诸特征生成伪语义域,其中所述伪语义域是从所述诸特征导出的n维空间。

4.如权利要求3所述的设备,其特征在于,生成所述语义模型包括通过为每个帧导出所述伪语义域中的坐标来将所述帧映射到所述伪语义域,所述坐标的每个元素与所述伪语义域的维度相对应并且是从所提取的诸特征中为所述帧所特有的特征的存在性导出的。

5.如权利要求4所述的设备,其特征在于,生成所述语义模型包括从映射到所述伪语义域的所述帧创建生成模型。

6.如权利要求5所述的设备,其特征在于,创建所述生成模型包括通过将具有最高得分的帧添加到关键帧集合来递归地标识所述帧中的所述关键帧集合,帧的所述得分是针对所述关键帧集合的所有成员就所述帧的坐标的平方范数乘以常数然后除以所述帧与所述关键帧集合中的帧之间的距离的范数的平方进行的求和的倒数。

7.如权利要求1所述的设备,其特征在于,向所述帧指派相应的相关性得分包括从所述视频提取拍摄所述视频的人的行为指示符。

8.如权利要求1所述的设备,其特征在于,使用所述相应的相关性得分来初始化所述语义模型包括构建一图形,在所述图形中,节点对应于所述帧,节点值对应于所述帧的相应的相关性得分,而边缘按所述语义模型中确定的帧之间的距离倒数来加权。

9.如权利要求8所述的设备,其特征在于,基于所述相应的相关性得分,关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。

10.如权利要求9所述的设备,其特征在于,与所述低相关性关键帧和所述高相关性关键帧相对应的节点被标记为固定值。

11.如权利要求10所述的设备,其特征在于,在初始化之后收敛所述语义模型包括为所述语义模型中不与所述低相关性关键帧或所述高相关性关键帧相对应的节点计算节点值。

12.如权利要求11所述的设备,其特征在于,节点之间沿边缘的消息传递被用于计算所述节点值。

13.一种用于自动视频概括的机器实现方法,其特征在于,所述方法包括:

获得视频;

从所述视频的帧生成所述视频的语义模型;

向所述帧指派相应的相关性得分；

使用所述相应的相关性得分初始化所述语义模型；以及

迭代地处理所述语义模型以产生子场景集合，每次迭代包括：

在初始化之后收敛所述语义模型；

在收敛之后选择具有最高相关性得分的帧序列；以及

通过固定所选帧序列的所述相关性得分来重新初始化所述语义模型。

14.如权利要求13所述的方法，其特征在于，生成所述语义模型包括提取所述帧的诸特征。

15.如权利要求14所述的方法，其特征在于，生成所述语义模型包括从所提取的诸特征生成伪语义域，其中所述伪语义域是从所述诸特征导出的n维空间。

16.如权利要求15所述的方法，其特征在于，生成所述语义模型包括通过为每个帧导出所述伪语义域中的坐标来将所述帧映射到所述伪语义域，所述坐标的每个元素与所述伪语义域的维度相对应并且是从所提取的诸特征中为所述帧所特有的特征的存在性导出的。

17.如权利要求16所述的方法，其特征在于，生成所述语义模型包括从映射到所述伪语义域的所述帧创建生成模型。

18.如权利要求17所述的方法，其特征在于，创建所述生成模型包括通过将具有最高得分的帧添加到关键帧集合来递归地标识所述帧中的所述关键帧集合，所述帧的得分是针对所述关键帧集合的所有成员就所述帧的坐标的平方范数乘以常数然后除以所述帧与所述关键帧集合中的帧之间的距离的范数的平方进行的求和的倒数。

19.如权利要求13所述的方法，其特征在于，向所述帧指派相应的相关性得分包括从所述视频提取拍摄所述视频的人的行为指示符。

20.如权利要求13所述的方法，其特征在于，使用所述相应的相关性得分来初始化所述语义模型包括构建一图形，在所述图形中，节点对应于所述帧，节点值对应于所述帧的相应的相关性得分，而边缘按所述语义模型中确定的帧之间的距离倒数来加权。

21.如权利要求20所述的方法，其特征在于，基于所述相应的相关性得分，关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。

22.如权利要求21所述的方法，其特征在于，与所述低相关性关键帧和所述高相关性关键帧相对应的节点被标记为固定值。

23.如权利要求22所述的方法，其特征在于，在初始化之后收敛所述语义模型包括为所述语义模型中不与所述低相关性关键帧或所述高相关性关键帧相对应的节点计算节点值。

24.如权利要求23所述的方法，其特征在于，节点之间沿边缘的消息传递被用于计算所述节点值。

25.一种机器可读介质，包括指令，其特征在于，所述指令在由机器执行时，使所述机器执行权利要求13-24的方法中的任一项。

用于自动视频概括的设备和方法

[0001] 优先权申请

[0002] 本申请要求2015年10月28日提交的美国申请S/N.14/925,701的优先权的权益,该申请通过引用整体结合于此。

技术领域

[0003] 本文描述的实施例一般涉及计算机视觉系统,且更具体地涉及自动视频概括。

背景技术

[0004] 计算机视觉系统一般涉及允许计算机处理图像数据以从该数据导出含义的进展。计算机视觉是人工智能的一方面,人工智能是涉及开发用于执行传统上需要活着的行动者(诸如人)的认知任务的人工系统的领域。视频一般由一系列静止图像组成。本文所使用的视频概括是指选择视频的子序列来创建该视频的子场景。这些子场景可被称作剪辑、集锦等。

附图说明

[0005] 在附图中(这些附图不一定是按比例绘制的),相同的数字可以描述在不同视图中的类似的组件。具有不同的字母后缀的相同的数字可以表示类似组件的不同实例。附图一般作为示例而非限制方式来示出在本文档中讨论的各种实施例。

[0006] 图1是根据一个实施例的环境的示例的框图,该环境包括用于自动视频概括的系统。

[0007] 图2是根据实施例的例示出用于自动视频概括的系统的通信流程的示例的框图。

[0008] 图3是根据实施例的例示出用于自动视频概括的系统的通信流程的示例的框图。

[0009] 图4例示出根据实施例的用于生成语义模型的系统通信流程的示例。

[0010] 图5例示出根据实施例的语义模型可视化的示例。

[0011] 图6例示出根据实施例的用于对帧指派相关度得分的系统通信流程的示例。

[0012] 图7例示出根据实施例的用于使用语义模型和所分配的相关性得分来产生子场景的系统通信流程的示例。

[0013] 图8例示出根据实施例的使用语义模型和所分配的相关性得分来产生子场景的可视化。

[0014] 图9例示出根据实施例的用于支持子场景选择的监督式学习的用户界面。

[0015] 图10例示出根据实施例的用于自动视频概括的方法的示例。

[0016] 图11是例示可在其上实现一个或多个实施例的机器的示例的框图。

具体实施方式

[0017] 概括视频是困难的计算机视觉问题,因为它通常不仅需要理解视频中显示的是什么,而且还要理解它是否重要。分类器是被设计成将语义上下文赋予视觉数据的系统(例

如,经训练的机器学习系统)。视频概括分类器通常在特定场景下经受训练,诸如穿行于城市环境中的行人手持相机。一般来说,这些分类器需要大量训练集(例如,在城市中行走的人的视频)并且常常要求显著的人类反馈(例如,选择相关帧、子场景等)是有用的。然而,这些分类器趋向受限于所使用的特定训练集和技术。例如,先前提到的城市行人分类器可将牲畜分类为相关(大多数城市环境中的不寻常事件),并由此将来自养牛场的视频镜头的牛分类为在很大程度上相关,尽管在该情境中,动物是司空见惯的东西。

[0018] 为了使上面讨论的视频概括问题复杂化,对于现存的计算机视觉系统来说,子场景边界也难以检测或以其他方式确定。也就是说,在没有全范围人类理解的情况下,当每个场景涉及相同主题、相同背景、相同面部等时,机器难以知晓持续三分钟发出歇斯底里的笑声的情节是一个子场景,而后续亚秒级喷嚏是分开的子场景。而且,对于子场景确定和一般的视频概括两者而言,如果视频是由人类拍摄(例如,控制、指导等)的,则底层概括是主观的。也就是说,两个不同的人会可能对相同的视频不同地做概括。在相机操作者是业余爱好者,或者给定原始镜头(例如,随意捕捉的连续镜头)而没有受益于可限定子场景边界的指导、编辑或其他技术时,这些问题可能恶化。

[0019] 为理解决以上注意到的问题,本文描述了一种自动视频概括系统。该系统从视频本身包含的信息生成视频的语义模型。以此方式创建语义模型允许子场景之间的固有差异来定义子场景边界,而不是依赖任意定时或经过特殊训练的分类器。该系统确实针对相关性问题使用分类器,但是语义模型允许使用更为不准确的相关性分类来产生有用的结果。因此,可使用在不同环境和场景下训练的分类器,因为结果不取决于人们认为相关的分类器的最终目标准确度,而是取决于视频中的比较相关性。最后,系统将生成的语义模型与不完善的相关性分类相组合,以从视频迭代地生成子场景,并由此自动地对视频做概括。

[0020] 图1是根据一实施例的环境的示例的框图,该环境包括用于自动视频概括的系统100。系统100可包括相机105(用于捕捉视频)、存储设备110(用于缓冲或储存视频)、语义分类器115、相关性分类器120以及多路复用器125。所有的这些组件以电磁硬件的形式实现,诸如电路(例如,下面所述的电路集)、处理器、内存、盘等。在示例中,这些组件中的一些或全部可共处于单个设备130中。

[0021] 存储设备110被安排成保持视频。在示例中,将该视频从相机105递送至存储设备110。在示例中,该视频由在某个时刻获得对视频的访问的另一个实体递送,该实体诸如移动电话、个人计算机等。存储设备110提供存储,系统100的其他组件可从该存储捡取和分析该视频的帧或其他数据。

[0022] 语义分类器115被安排成从视频的帧生成该视频的语义模型。如本文所使用的,语义模型是通过其表示帧之间的相似性的设备。在示例中,为了生成该模型,语义分类器115被安排成提取帧的特征。在示例中,该特征为低级特征。如本文所使用的,低级特征是帧的不需要任何语义理解的那些方面。例如,帧区域中为绿色的百分比不需要理解这样的特征意味着什么,而只需要简单的测量。然而,一般来说,低级特征是已被发现隐含语义信息的测量。与低级特征的测量相反,高级特征可包括来自底层测量的更好推断,诸如对面部的标识,其涉及按照各种当前形状和颜色对区域进行分类。

[0023] 在示例中,低级特征包括GIST描述符。可通过使用在不同尺度和方向上的多个Gabor滤波器对帧进行卷积来计算GIST描述符,以产生多个特征图。在示例中,存在32个

Gabor滤波器、4个尺度和8个方向，用于为GIST描述符产生三十二个特征图。然后可将这些特征图分成多个区域(例如，十六个区域或4乘4网格)，其中每个区域的平均特征值被计算。最后，平均值可被串接(例如，连接)以产生GIST描述符。可使用其他低级特征技术，诸如用于标识帧中的形状或线条的霍夫变换、基于颜色的度量等。在示例中，帧的元数据可被测量以进行特征提取，诸如帧捕获的地理位置。在示例中，可使用低级声音特征。在示例中，梅尔频率倒谱系数(MFCC)可被用作低级特征。一般而言，诸如存在喧闹噪声或不存在噪声之类的音频提示可能有助于标识视频的感兴趣(例如，相关)部分。

[0024] 在示例中，语义分类器115被安排成提取帧的高级特征。如上所述，高级特征涉及理解底层图像数据的某些内容，以例如确定背景(例如，室内、户外、家中、办公室中、剧院中、公园中、海上、海滩上、城市环境中、农村里；自然环境、山上等)、活动(例如，运动、游泳、诸如音乐会的事件、工作、聚会、烹饪、睡觉等)或物体(例如，地标、人、面部、动物、诸如汽车、卡车之类的装备、曲棍球棒或其他运动设备等)。在示例中，为了提取这些高级特征，语义分类器115可采用或实现分类器来标识场景(例如，多个帧)特征，在示例中，可使用针对场景分类训练的深度卷积网络来实现该分类器。

[0025] 一旦帧的特征被提取，该语义分类器115就基于所提取的特征组织数据结构中的帧。这种组织为模型提供了用于基于帧的相应特征的共性来表示帧的共性的有意义方式。在示例中，生成语义模型包括语义分类器115被安排为从所提取的帧特征生成伪语义域。这种伪语义领域是从所发现的特征导出的n维空间。例如，如果在三个特征上测量每个帧，则对每个特征的相应测量将是相应帧的三维空间中的坐标。其可视化在图5中给出并在下文中详细描述。一般来说，空间的维度等于或小于针对其作出所尝试提取的不同特征的数量。因此，给定一千个特征测试来从帧中提取特征，伪语义域的维度将是一千或更少。当例如特征提取揭示在视频中不存在该特征时，可以降低维度。在此示例中，对应于此特征的维度可被从n维空间中移除，从而使n维空间成为(n-1)维空间。例如，如果语义分类器115试图从帧中提取大象影像，并且不存在此类影像，则伪语义域的维度将被降低以消除大象影像。出于高效起见，也可使用其他降低机制来例如进行后续计算。在示例中，语义分类器115可采用谱嵌入(spectral embedding)来降低维度。

[0026] 在示例中，伪语义域可通过多个人工智能网络来处理和实现。例如，所提取的特征(例如，在帧中找到的那些特征)可被用于训练深度玻尔兹曼机(deep Boltzmann machine)——一种在没有监督的情况下被初始化和训练的神经网络。还可以使用各种其他无监督人工智能模型。然而，在示例中，仅从存在于视频帧中的特征而非从外部源创建伪语义域。如之后将讨论的，该特征衡量各帧之间的差异以允许跨比当前技术允许的更多种主题视频的子场景区分。可被使用的其他示例人工智能技术包括生成模型，诸如概率图模型或混合模型。

[0027] 在创建伪语义域之后，语义分类器115将个体帧映射到伪语义域如上所述，此类映射可包括使用个体特征提取值作为帧的坐标。这些值可被标准化，以便一起用作n维空间中的有效坐标。在示例中，不执行标准化，而使用原始值。在使用诸如深度玻尔兹曼网络之类的网络建立的伪语义域的例子中，映射个体帧可简单地涉及通过网络馈送每个帧以得出n维空间中为该帧所特有的结果坐标。

[0028] 当帧被置于n维度量空间中时，语义模型被生成，以使得空间中的帧之间的距离是

可计算的。作为简单示例，考虑二维空间(例如，由x和y标示的维度)中的欧几里德距离度量(Euclidean distance metric)，从一点(例如，帧)到另一点的距离遵循$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$=两点1和2之间的距离。在创建语义模型之后，任一帧与另一帧的相似性是n维空间中两帧之间的负平方距离的指数。也就是说，两帧越接近，它们越相似。

[0029] 在示例中，可标识来自所映射的帧的关键帧集合。所映射的帧或n维空间中的点表示n维空间中流形(manifold)的表面上的点。此流形最终是作为底层模型的流形，然而，该流形的确切定义并不是执行本文所描述的技术所必需的。事实上，可替代地使用帧的子集——关键帧。关键帧是代表语义概念的一组帧的单个帧。例如，n维空间中的帧群集表示相似的场景。来自该群集的帧可由此表示该群集，并且此帧是关键帧。可采用各种关键帧标识技术，诸如为群集寻找内核。在示例中，可通过对帧进行评分并相继取得最高得分帧来递归地标识关键帧，直到获取阈值数量个关键帧。在示例中，阈值由视频的长度决定。在示例中，阈值由n维空间中的多个经标识的群集来确定。在示例中，在帧的得分涉及关键帧之间的距离的情形中，阈值是帧之间的最小距离。也就是说，如果两帧之间的距离低于阈值，则递归搜索停止。

[0030] 在示例中，关键帧可按距离来被评分。此处，标识彼此远离的帧标识视频中显示不同事物的诸部分。为了对帧之间的距离进行评分，选择第一帧作为第一关键帧。在示例中，第一帧是基于离n维空间的原点最远来选择的。通过选择距离第一帧最远的帧，将第二帧选入关键帧集合中。所选的第三帧距离关键帧组中的第一帧和第二帧最远。如上所述，此过程可继续，直到第n帧之间的距离低于阈值。因此，该关键帧集合是通过将下一帧添加到在被映射的帧集合中具有最高得分的关键帧集合来对关键帧进行的递归标识，。在此示例中，帧的得分是针对关键帧组的所有成员就帧的坐标的平方范数乘以常数并除以该帧与关键帧集合中的另一帧之间的距离的范数的平方进行的求总的倒数。下式例示出该评分过程：

[0031] $$\tilde{X}_{R+1}=arg\max_{X_i}\left[\sum_{r=1}^{R}\frac{1+\gamma\|X_i\|^2}{\|X_i-\tilde{X}_r\|^2}\right]^{-1}$$

[0032] 其中$\tilde{X}$是键帧集合，X是所映射帧的集合，而γ是控制规则化水平的常数。

[0033] 如图1所示，相关性分类器120可独立于语义分类器115进行操作。此独立操作允许对视频的并行处理，因为语义分类器115的操作不干扰或以其他方式影响语义分类器115的操作。相关性分类器120被安排为向各帧分配相应的相关性得分。此类相关性分配可包括任何数量的分类器(例如，应用分类器库)，这些分类器对帧进行评分并且被组合来为每一帧提供得分或得分集合。示例分类器可涉及记录运动(例如，与静止相反的动作)、面部的存在等。分类器还可涉及关于视频的元数据，诸如视频被捕捉的时间或地点。在示例中，元数据与相机操作员的动作(例如，行为指示符)相对应。例如，相机保持稳定的时间或在一系列帧中使用变焦的时间。这些度量两者都暗示相机操作员的提高的兴趣，这可被用于增加相机保持稳定，或采用增加的变焦，或两者的帧的相关性。事实上，如下所述，单独使用这些相机操作员相关性分类器允许系统选择(与由其他人选择的其他子场景相比)与由操作员选择的那些更紧密对齐的子场景。在示例中，行为指示符和其他分类器的组合可被分别确定并

组合来为这些帧创建复合相关性得分。

[0034] 如上所述，系统100将语义模型与这些相关性得分组合以选择子场景，从而允许有效地使用较不准确的相关性得分。相应地，在示例中，相关性分类器可被过滤以将所应用的相关性分类器(例如，对相关性得分作出贡献的那些分类器)限制为具有大于百分之五十的几率(chance)指示相关性(例如，比随机结果更好)的那些分类器。在示例中，过滤受限于行为指示符。

[0035] 在示例中，基于用户反馈修改分类器库。例如，可将选定的子场景呈现给用户。用户可指示该子场景对于视频概括不是期望的。有助于子场景选择的相关性分类器可被从将来的分类中移除，或者可被修改(例如，加权)以对附加相关性得分具有较小的影响。

[0036] 在语义分类器115创建语义模型并且相关性分类器120将相关性得分指派给视频帧之后，多路复用器125被安排为获取这两个输入并产生对视频做概括的子场景集合。在示例中，多路复用器125被安排为利用各相关性得分来初始化语义模型并迭代地处理该模型以产生该子场景集合。在示例中，子场景选择的每次迭代包括在初始化之后收敛模型，在收敛之后选择具有最高相关性得分的帧序列，以及通过固定(例如，钳制、锚定等)此迭代中选择的帧序列的相关性得分来重新初始化模型。该过程继续进行，直到标识预定数量的子场景，或者直到达到质量阈值。在迭代处理结束时，可将子场景组合或以其他方式呈现(例如，经由文件格式、用户界面等)给用户作为经概括视频。

[0037] 在示例中，为了初始化模型，多路复用器125被安排为构建其中节点对应于视频的帧并且边缘被加权为语义模型中的帧之间的负平方距离的指数的图。因此，语义模型中帧越接近，连接帧的边缘的权重越大。节点的值是帧的对应的相关性得分。在示例中，当两个帧之间的距离超过阈值时，边缘被省略(例如，从未被置于图中)或被移除。也就是说，如果两帧的距离被移动得足够远，则在图中不会留下连接这些帧的相应节点的边缘。此类边缘减少可通过在每次迭代时减少计算次数来增加收敛模型的计算效率。在示例中，最小距离被确定成使得图被完全连接(例如，存在一系列边缘使得每个节点可到达每个其他节点)。

[0038] 在示例中，为了初始化模型，多路复用器125被安排成固定节点子集。在示例中，节点子集包括具有高(例如，大于中值、平均值或均值)相应相关性的那些节点和具有低(例如，小于中值、平均值或均值)相应相关性的那些节点。在示例中，节点的子集包含(例如，仅具有)关键帧。在示例中，所有的关键帧被固定并构成节点子集的整体。在示例中，关键帧的子集包括具有最高相关性得分的单个关键帧和具有最低相关性得分的单个关键帧。

[0039] 在示例中，为了收敛模型，多路复用器125被安排成计算不在该节点子集中的节点的节点值。也就是说，未被固定的每个节点在收敛之前具有最终未知的值。因此，每个节点在每次收敛迭代时被更新。该计算基于相邻(例如，由单个边缘连接的)节点的值。在示例中，该计算经由消息传递算法沿节点边缘在节点之间进行。在示例中，在每次收敛迭代时，节点的值被设置为相邻节点的加权平均值。因此，相邻节点的值被边缘权重修改(例如，乘以边缘权重)并与其他相邻节点的加权值相加并最终除以相邻节点的数量。收敛迭代继续直至达到收敛阈值。

[0040] 在示例中，收敛阈值定义节点值的变化率，低于该变化率时，收敛被认为达成。因此，如果初始迭代导致对节点值的大变化，则成比例地小得多的连续变化指示收敛完成。这种类型的缩放测量可帮助动态地处理(例如，在具有很多变化的视频与在整个视频中其内

容的整体变化很小的那些视频之间的)不同的语义模型。在示例中,收敛阈值是绝对值。也就是说,只要在迭代中节点中的最大变化落在阈值之下,收敛就完成。

[0041] 当模型对于节点具有高固定值和低固定值时,收敛过程导致接近高固定值的非固定节点朝向那些值移动,而接近低固定值的其他非固定节点朝向那些节点移动。因此,如语义模型中所表示的,框架的语义接近度固有地选择子场景边界。

[0042] 一般来说,非固定节点一般将倾向于双峰分布,如图8所示以及以下所描述的那样。因此,在给定复用器125迭代期间选择子场景涉及选择足够相关并包括高相关性关键帧的连续帧序列。在此上下文中,足够相关是基于结果节点值的分布来确定的阈值问题。如在图8的元素815中所例示的,简单地确定与低得分节点相对应的值并取得与具有比所确定的值更大的值的节点相对应的帧。因此,在示例中,在收敛之后选择具有最高相关性得分的帧序列包括多路复用器125被安排成选择带有具有大于选择阈值的值的对应节点的帧。在示例中,选择阈值是从节点值整体来确定。在示例中,选择阈值是节点值的平均值。

[0043] 在示例中,在每次迭代时重新初始化模型可包括复用器125固定为子场景选择的节点的节点值(例如,相关性得分)。因此,在之后的迭代中,同样的子场景将不被选择。在示例中,这些节点被固定为低相关性。在示例中,该低相关性是相关性下限(例如,最低相关性)。通过这样做,在后续迭代中将选择下一最相关子场景。而且,后续子场景将与所选择的子场景尽可能在语义上不同。因此,该系列子场景不仅体现视频的最相关部分,而且还将避免冗余子场景。在示例中,先前选择的节点将被固定为高相关性。在示例中,该高相关性是相关性上限(例如,最高相关性)。诸如经由以下所描述的用户界面135的这样的设置允许用户在子场景选择中请求更多相同的设置。然而,应该理解,在后续子场景选择迭代中将不选择先前选择的相同节点。在示例中,节点可被固定在低相关性与高相关性之间的中性相关性。在示例中,中性相关性级别是节点的相关性值的均值或中值中的至少一个。此设置使先前选择的子场景的效果最小化,使得它既不鼓励类似的后续子场景也不鼓励不同的后续子场景。

[0044] 可用多种方法来设置所产生的子场景的总数。在示例中,子场景的总数基于视频的长度。在示例中,子场景的总数基于语义模型中群集的数目。在示例中,子场景的总数基于用户设定值。在示例中,如果子场景被用户取消选择(例如,移除、删除等),则可产生附加的子场景来填补其位置。

[0045] 在示例中,随着每一子场景被选择,子场景的剪辑被创建。创建剪辑可涉及简单地标识作为剪辑的部分的帧。在示例中,创建剪辑包括复制帧序列以创建剪辑。在示例中,系统100包括子场景(例如,剪辑)被呈现于其中的用户界面135。在示例中,子场景以它们被产生的次序来呈现。在此示例中,产生的顺序也是概括视频的子场景的重要性的次序。在示例中,用户界面135可被安排成从用户接收关于子场景的丢弃选择。所丢弃的子场景随后被从最终视频概括中移除。如上所述,系统可利用此类用户参与到对系统100的输出的校正中来调整相关性得分机制。

[0046] 如以上所描述的系统100的操作将视频自动地概括为多个子场景剪辑。通过使用所描述的机制,系统100不需要如一些现存系统所需要的高准确或上下文相关性分类器。相反,系统100使用固有语义信息以及帧之间的相对相关性值来在没有人类参与的情况下有效地对视频做概括。

[0047] 而且，当在相关性得分过程中使用相机操作员的行为指示符时，在与由相机操作员所期望的结果比较时，由系统100产生的结果优于陌生人的那些结果。这是使用五个不同相机操作员在十个视频剪辑上被实验验证。在实验中，系统100的简单版本被使用并与作为参考的、未经训练的人类表现进行比较。仅使用伪语义域中的变化来确定相关性得分。比较精确度/召回(recall)，并对照并非相机操作员的人类比较系统100概括的F1得分(信息检索)。当与参考人员比较时，系统100对于视频中的七个视频里的相机操作员的子场景选择具有较好的一致性，在一个视频中在统计上相等，而在两个视频中更差。还示出系统100的性能优于对子场景的随机选择(超过100000次实验)以及其他固定选择方法(例如，暂时选择中心/最先/最后子场景)。因此，系统100在概括相机操作员的视频以使该操作员满意上出乎意料地比随机人类执行得更好。

[0048] 图2是根据实施例的例示出用于自动视频概括的系统200的通信流程的示例的框图。图2中所例示的通信流程是用于实现自动视频概括的各个组件的概览。系统200的组件由机器实现。

[0049] 在组件205处，视频数据(例如，帧)和可选传感器数据(例如，用于确定相机是否在移动的加速度计数据等)被获得。组件205被安排成提取帧的特征，例如，如上文所描述的那样。所提取(发现)的特征被置于所提取的特征的池210中。

[0050] 在组件225处，系统200学习固有语义域。因此，组件225操作成使得视频本身(被做概括的视频)为完整域。也就是说，组件225操作成如同视频包含全部所需要的语义信息(非固有)那样，并创建(例如，学习、训练等)用于视频内容的生成模型(例如，实现机器学习)。这将导致子场景解析，其中系统200仅从视频学习什么是不同子场景。组件225随后产生伪语义域230作为输出。

[0051] 在示例中，伪语义域230可在其他情境中被重新使用。例如，诸如位置、时间(例如，当日时间、假日等)以及环境条件(例如，室内/室外标识)之类的情境信息可被合并到伪语义域230以重新使用(例如，精炼化)先前从在相同(或相似)情境下可能由不同用户(例如，相机操作员)拍摄的视频生成的伪语义域230。此类伪语义域改进一般不会干扰系统200根据特定用户的偏好来对视频做概括的能力，因为这是在此阶段之后完成的。

[0052] 在组件215处，系统200在视频帧中寻找一般相关性线索。如上所述，这可包括对帧评分。在示例中，组件215标识相机操作员自己的行为并采用其作用子场景相关性(例如，兴趣性、重要性)的线索。预定义摄影师行为列表比预定义视频内容描述列表简单得多，从而允许系统200以比其他系统更少的初步设置来操作。该行为相关性可作为诸如物体、面部、人类、动作或场景标识算法之类的其他相关性标识符的附加来使用。组件215的输出是相关性标签域220(例如，应用于帧的相关性得分)。

[0053] 相关性推断引擎235将伪语义域230和相关性标签域220进行多路复用来作为输入并产生经概括视频作为输出。因此，相关性推断引擎235融合了语义信息与一般相关性线索。相关性推断引擎235在从固有语义信息产生的语义模型上应用直推式学习并用一般相关性线索来初始化该语义模型。经初始化的模型随后被迭代地收敛。在收敛之后，该模型为视频的各个部分提供固有相关性得分。隐式地获得时间平滑度，以及如图8的元素815中所例示的相关性感知的子场景边界检测。

[0054] 相关性推断引擎235重复模型初始化和模型收敛以迭代地提取具有减小的相关性

的子场景。这可经由组件240来完成。在每次模型收敛时，组件240“学习”视频段的相对相关性(例如，固有相关性)。组件240随后以减小的相关性次序从视频提取子场景。例如，一旦最重要的子场景被标识，它被反馈到推断引擎作为具有低相关性(或替代地移除/屏蔽它)。当模型被再次收敛时，下一最相关子场景被揭露。在示例中，该迭代过程允许用户或自动程序在任何时间停止，从而控制概括压缩率(例如，多少剪辑将被用于概括)。如所注意到的，最相关子场景首先被标识。后续标识的子场景比先前所标识的子场景具有更低相关性/兴趣性/重要性。经概括视频是系统200的输出。

[0055] 在示例中，系统200采用用户界面来呈现所标识的子场景并接受用户输入。该用户界面可向系统200提供附加特征，这些附加特征允许用户从相关性方面的有序表(与搜索引擎结果列表或决策支持系统相似)手动选择相关子场景。由于组件240的操作是迭代的，因此该列表可在运行中实时地增长。而且，倘若视频的全自动概括与半自动概括(例如，包括人类输入)不同，则系统200可更新其语义模型(线上或活跃学习方案)以通过调整相关性线索(例如，相关性标签域220)或固有模型(例如，伪语义域230)来结合用户反馈。

[0056] 图3是根据实施例的例示出用于自动视频概括的系统300的通信流程的示例的框图。图3-4和6-7提供系统300的细节。图3提供各组件的概览而剩余附图描述这些组件的子组件。所讨论的所有组件和子组件均由遍及本文档而描述的机器来实现。

[0057] 组件305使用无监督学习机制来学习用于视频的固有伪语义域。组件305还在此伪语义域中学习用于剪辑的生成模型(例如，通过针对全部子场景标识关键帧)。

[0058] 组件310关于一般相关性(即，个体帧的重要性或兴趣性)来对视频帧评分。此得分充当相关性的初始证据并使用硬编码(专家知识)或先前(例如，通过向系统馈送视频及其手动概括)学习的规则。这些规则无需是十分精确的，然而，由于之后对伪语义域和一般相关性建模的组合，因此应当比随机猜测要好(例如，准确度>0.5)。组件310的输出使用不同视频部分的相对的、非绝对的得分。也就是说，一般相关性得分被衡量以使得在帧群体内存在最大和最小相关性得分。

[0059] 给定来自组件310的初始相关性得分和来自组件305的语义域，组件315生成用于推断子场景的相关性的模型。该推断可以是具有软标签(例如，来自组件310或先前迭代的证据)的半监督学习算法。

[0060] 图4例示出根据实施例的用于生成语义模型的系统通信流程400的示例。系统400是上文所讨论的组件305的示例实现。

[0061] 要被做概括的视频被置于存储设备405中。场景特征由组件410来提取。这些特征是用于对场景类型(例如，室内/室外、海滩/日落/聚会等)进行分类的共同特征。示例特征可包括GIST描述符或用于场景分类的深度卷积网络的第一层的输出。所提取的特征可被置于存储设备415中以供其他组件使用。

[0062] 组件420学习伪语义域。这使用无监督学习算法来完成，诸如训练深度玻尔兹曼机、谱嵌入、自动编码器、稀疏滤波等。“语义”自然源于所使用的特征类型。无监督学习算法在维持语义解释的同时减少噪声。因此，相对于其他视频帧(例如，直推式推断)具有类似“语义”信息的两帧相对于其他视频帧的映射被映射到其间具有小距离的点。可选输入2允许通过使用来自在相同或类似环境下(例如，相机操作员的偏好、时间、地点、事件等)捕捉的视频的语义域来进行众包。

[0063] 存储设备425包含将被用于把视频映射到域的伪语义模型(例如,所学习的域)。组件430将视频帧映射到所学习的域。如果深度波尔兹曼机被用于产生所学习的域,则映射涉及通过该波尔兹曼机前馈这些帧。在其中使用谱嵌入的情形中,样本外扩展技术可被用于执行映射。

[0064] 组件435学习用于映射视频的生成模型(例如,概率图模型、混合模型等)。也就是说,组件435学习被假定来"生成"视频帧的模型。在系统400中,组件435标识关键帧——例如,散布在由映射视频创建的流形上的帧集合。在示例中,组件435使用其中具有最高得分的帧被添加到关键帧集合的递归过程:

[0065] $$\tilde{X}_{R+1} = arg\max_{X_i}\left[\sum_{r=1}^{R}\frac{1+\gamma\|X_i\|^2}{\|X_i-\tilde{X}_r\|^2}\right]^{-1}$$

[0066] 其中$\tilde{X}$为关键帧集合,X是映射帧结合,而γ是控制规则化的常数。到组件435的输入3是其中来自具有相同或类似环境的其他视频的数据可被用于例如标识关键帧的可选输入。组件435的输出是生成模型,该生成模型在被查询时,提供关于视频中的帧的语义相似性度量(例如,对两帧在语义上有多相似的测量)。

[0067] 图5例示出根据实施例的语义模型可视化500的示例。如上文所多方面描述的,语义模型或伪语义域可以是其中维度与所提取的帧特征相对应的n维空间。为简单起见,可视化500例示出三维空间。小黑点标示映射到该域的视频的帧。大白点表示所选择的关键帧。注意,可视化500左侧的帧彼此的邻近度大于右侧的帧彼此的邻近度。如所例示的,例如,左侧场景中的女子正在跳入水中,并且因此与右侧场景中沿池奔跑的人相比,被群集化。

[0068] 如上所述,点之间的距离衡量与这些点相对应的帧的相似性。在选择关键帧上,如上文参考图4所描述的,初始帧被选择,该初始帧很能是离空间中的原点最远的帧。所选择的下一帧是距离第一场景最远的那个帧。后续帧是距离全部先前所选择的关键帧最远的那些帧。因此,映射帧提供了对建模帧之间的语义关系的空间中的流形的直推式推断。关键帧是流形的模型,并且之后被用于生成用于新输入的流形。

[0069] 图6例示出根据实施例的用于对帧支配相关性得分的系统通信流程600的示例。系统600是上文所讨论的组件310的示例实现。系统300学习/应用一般相关性规则(例如,具有分类准确度>0.5)。这些规则可从先前(手动或半自动)经概括的视频学习或由人类专家(例如,程序员)直接添加。系统600使用全局规则来标识视频中具有相对高相关性得分的帧。

[0070] 存储设备605和625存储要做概括的视频。存储设备625还可包括先前经概括的子场景和伴随的相关性信息,例如,经由与输入1的简单视频相对照的输入3。

[0071] 组件610和635以相同方式操作。组件615和640也以相同方式操作。两个所例示流程之间的主要差异是组件630的帧标示和组件645的一般相关性规则学习,该一般相关性规则学习可包括经由存储设备625的附加子场景信息而来自用户的反馈信息。

[0072] 在组件610处,使用诸如相机运动的标准偏差之类的特征或通过检测变焦(例如,推近或拉远)来提取相机操作员的行为。这提供了正从相机操作员的视点捕捉的场景的相关性上的线索。例如,如果相机抖动较低,或者存在推近动作,则可推断出(例如,具有大于百分之五十的准确度)相机操作员认为在场景中发生了一些感兴趣的事情(例如,相关性分类器620应该将这些帧分类为相关)。在示例中,(例如,如组件615所提取的)内容特征,诸如

面部识别，也可充当用于相关性分类器620的相关性的线索。

[0073] 一般相关性规则650可外部地(例如，经由输入4)提供，或者从先前手动/半自动做概括的视频(例如，经由输入3经由用户界面)学习。在后一种情形中，提取相机操作员的行为(例如，组件635)和内容特征(例如，组件640)的相同特征。通过标记先前概括中包括哪些帧以及不包括哪些帧——例如，经由组件630进行帧标示，诸如将完整视频中的每个帧标示为不相关(如果其不在概括中)或者相关(否则的话)——监督可被用于通过相关性分类器620利用诸如支持向量机、深度度神经网络之类的公共监督学习算法来学习或补充一般分类规则650。输出2是视频帧的相关性分类(例如，得分)。

[0074] 图7例示出根据实施例的用于使用语义模型和所指派的相关性得分来产生子场景的系统通信流程700的示例。系统700是上文所讨论的组件315的示例实现。

[0075] 组件705使用来自系统400(例如，经由输入1)的映射视频剪辑并生成用于半监督学习的模型710。在示例中，模型是其节点与帧相对应且其边缘与伪语义域中的相似性相对应(例如，使用n维空间中的点之间的距离倒数来对边缘加权)的图形。

[0076] 使用经由输入2(例如，来自系统600或来自先前迭代的)可用相关性证据来初始化组件715。在示例中，组件715选择最确信的证据(例如，具有最高和最低相关性得分的关键帧)并固定(例如，钳制)图形中相应节点(帧)的状态。剩余节点不被固定并被视为其状态将在之后被确定(例如，推断)的隐藏变量。

[0077] 例如通过组件720推断隐藏变量的状态可使用诸如迭代标签传播之类的消息传递机制来完成。在每次迭代中，迭代标签传播涉及图形中每个节点的状态被设置为其相邻节点的加权平均值。一旦被收敛，具有最高和最低相关性得分的帧就被视为证据并被用于重新初始化该图形(例如，从组件720被传达回组件715)。

[0078] 图8例示出根据实施例的使用语义模型和所指派的相关性得分来产生子场景的可视化800。可视化800例示出两个图表：初始化数据805，其中相机运动变化是y轴而帧编号(例如，每个帧的有序编号)是x轴；以及收敛相关性图表815，其中x轴依旧是帧编号而y轴是所推断的兴趣性。图表815是模型810收敛之后。如上所述，关键帧(例如，图表815中的较大圆)已将剩余节点拉到高状态(例如，关键帧位于其顶点的峰值)和低状态(例如，组成平稳段的剩余帧)。而且，如由图表815所证实的，子场景选择是直接的，那就是用于到达子场景间隔820的对相关性得分的取阈值。也就是说，通过确定平稳段的最大值，并取超过此最大值的任何帧作为子场景，子场景边界可被确定。

[0079] 再次地，图表805是每一帧中(x轴)相机操作员的运动变化(y轴)；此处，低运动意味着较高相关性。图表中无阴影和有阴影的圆是分别用低和高相关性得分固定的所选择的关键帧。迭代消息传递被应用于图形810(暗阴影与亮阴影节点是所固定的关键帧)，其中消息在连接节点的边缘上被传递。在收敛之后，在图表815上，节点(x轴)的相关性得分(y轴)被示出。

[0080] 在示例中，一旦被收敛，最高得分视频段(例如，子场景)被添加到经概括的剪辑缓冲器并且还被设置为用于重新初始化图形810的低相关性证据。因此，后续视频段可获得最高相关性得分并且可被检测到而不被已发现的最相关视频段屏蔽。例如，为了在初始图形810收敛之后迭代地标识相关子场景，段被添加到经概括的视频。接着，此段作为低相关性证据被添加以重新初始化该图形来供下一次迭代。图形810再次被收敛并且第二最相关段

被标识并被添加到视频概括。

[0081] 图9例示出根据实施例的用于支持子场景选择的监督式学习的用户界面900。以上讨论的系统的视频概括的独特能力允许直观用户界面最大程度地审阅或编辑概括中的子场景包含物，以及提供一些反馈以助益对相关性线索或语义建模的监督式学习。例如，当请求用户干预时(例如，决策支持系统操作模式)，可采用用户界面900。如果用户对全自动概括结果不满意，则可移除特定的子场景或将其重新排序成在概要中并不显著。用户界面900例示出类搜索引擎界面，其中所呈现的子场景以(如自动检测的)递减相关性来排序。用户可用其手指在屏幕上上下滑动来浏览子场景。如果例如用户浏览超过已生成的子场景的末尾，则系统可产生附加子场景来填充菜单(例如，通过激活从组件720至组件715的反馈)。用户还可通过将其手指例如分别滑动到左侧和右侧来选择在概要中要包括或不包括哪个呈现的子场景。

[0082] 图10例示出根据实施例的用于自动视频概括的方法1000的示例。方法1000的操作由诸如上文或下文所描述的机器(例如，计算机硬件)来执行(例如，由电路集实现)。

[0083] 在操作1005处，可获得(例如，接收或检取)视频。该视频包括组成视频的帧。

[0084] 在操作1010处，可从视频的帧生成视频的语义模型。在示例中，生成语义模型包括提取帧的特征。在示例中，提取特征包括寻找低级特征。在示例中，低级特征包括GIST描述符。在示例中，低级特征包括位置。

[0085] 在示例中，提取特征包括寻找高级特征。在示例中，寻找高级特征包括将分类器应用于帧以标识场景特性。在示例中，应用分类器包括使用为场景分类训练的深度卷积网络。在示例中，场景特性包括一个或多个背景、活动或物体。在示例中，背景是室内、室外、海滩上、森林中或城市环境中的一个。在示例中，活动是聚会、运动、事件或工作中的一个。在示例中，物体是面部、动物、地标、或运动设备中的至少一个。

[0086] 在示例中，生成语义模型包括从所提取的特征生成伪语义域，其中伪语义域是从特征导出的n维空间。在示例中，从所提取的特征生成伪语义域包括使用所提取的特征训练深度波尔兹曼机。在示例中，从所提取的特征生成伪语义域包括将谱嵌入应用于所提取的特征以减少维度。

[0087] 在示例中，生成语义模型包括通过为每个帧导出伪语义域中的坐标来将帧映射到伪语义域，该坐标的每个元素对应于伪语义域的维度并且是从特定于该帧的所提取特征的特征的存在导出的。在示例中，将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。在示例中，将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0088] 在示例中，生成语义模型包括从映射到伪语义域的帧创建生成模型。在示例中，创建生成模型包括创建概率图模型。在示例中，创建生成模型包括创建混合模型。在示例中，创建生成模型包括通过将具有最高得分的帧添加到帧集合中来递归地标识帧中的关键帧集合，帧的得分是帧的坐标的平方范数乘以常数并且除以该帧与(对于关键帧组的所有成员的)关键帧集合中另一帧之间的距离的范数的平方的总和的倒数。

[0089] 在操作1015处，可向帧指派相应的相关性得分。在示例中，向帧指派相应的相关性得分包括从视频中提取拍摄该视频的人的行为指示符。在示例中，行为指示符包括用于捕捉该视频的相机运动的缺少。在示例中，行为指示符包括用于捕捉该视频的相机的增加的

变焦。在示例中，从视频中提取拍摄该视频的人的行为指示符包括限制具有百分之五十或更大几率指示相关性的行为指示符。

[0090] 在示例中，向帧指派相应的相关性得分包括将相关性分类器库应用于帧。在示例中，通过用户反馈修改相关性分类器库。在示例中，应用相关性分类器库的输出得分与来自从视频中提取的拍摄该视频的人的行为指示符的输出组合以创建复合相关性得分。在示例中，相关性分类器库被过滤以排除具有少于百分之五十指示相关性的的几率的分类器。

[0091] 在操作1020处，可使用该相应的相关性得分来初始化语义模型。在示例中，使用相应的相关性得分来初始化语义模型包括构建一图形，在该图形中，节点对应于帧，节点值对应于帧的相应相关性得分，而边缘通过语义模型中确定的帧之间的距离倒数来加权。在示例中，当两帧之间的距离低于阈值时，边缘被省略或移除。

[0092] 在示例中，基于相应的相关性得分，关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。在示例中，与低相关性关键帧和高相关性关键帧相对应的节点被标记为固定值。

[0093] 在操作1025处，可迭代地处理模型以产生子场景集合。该组子场景是视频的概括。操作1030－1040描述作为子场景产生的每次迭代的部分的操作。在示例中，迭代地处理语义模型以产生子场景集合继续进行，直到标识预定数目的子场景。在示例中，预定数目是用户输入。在示例中，预定数目基于视频的长度。在示例中，预定数目基于从语义模型确定的帧的群集的数目。

[0094] 在操作1030处，可在初始化后收敛语义模型。在示例中，在初始化后收敛语义模型包括为语义模型中不与低相关性关键帧或高相关性关键帧相对应的的节点计算节点值。在示例中，节点之间沿边缘的消息传递被用于计算节点值。在示例中，用于计算节点值的节点之间沿边缘的消息传递包括将每个节点值设置为相邻节点的加权平均值，相邻节点通过边缘连接到该节点，连接该节点与相邻节点的边缘的权重在对相邻节点的值求平均以创建加权平均值之前修改该相邻节点的值。在示例中，将每个节点值设置为相邻节点的加权平均值继续进行直到达到收敛阈值。在示例中，收敛阈值定义节点值的变化率，低于该变化率时，收敛被认为完成。

[0095] 在操作1035处，可选择收敛之后具有最高相关性得分的帧序列。在示例中，在收敛之后选择具有最高相关性得分的帧序列包括选择具有带有大于选择阈值的值的对应节点的帧。在示例中，选择阈值是从节点值的整体性来确定的。在示例中，选择阈值是节点值的平均值。在示例中，选择帧序列包括创建具有该帧序列的视频剪辑。

[0096] 在操作1040处，可通过固定所选帧序列的相关性得分来重新初始化语义模型。在示例中，通过固定所选帧序列的相关性得分来重新初始化语义模型包括在语义模型中将所选帧序列标记为固定值。在示例中，固定值被设置为最低可能相关性得分。在示例中，高相关性关键帧的新子集被选择并被设置为固定值。

[0097] 方法1000还可包括以下任选操作：以产生子场景的顺序呈现包括这些子场景的用户界面；从用户接收对子场景的丢弃选择；以及从视频概括中丢弃子场景。在示例中，指派相应相关性得分的相关性标准可被修改以降低在所丢弃子场景中发现的相关性标准的重要性。

[0098] 图11例示出本文中所讨论的技术(例如，方法)中的任意一个或多个可在其上执行

的示例机器1100的框图。在替代实施例中，机器1100可以作为独立设备来操作或者可以连接(例如，联网)到其他机器。在联网部署中，机器1100可在服务器-客户端网络环境中的服务器、客户端或其两者的能力内操作。在示例中，机器1100可以充当对等(P2P)(或其他分布式)网络环境中的对等机器。机器1100可以是个人计算机(PC)、平板PC、机顶盒(STB)、个人数字助理(PDA)、移动电话、web应用、网络路由器、交换机或桥接器、或者能够执行指定该机器要采取的行动的指令(顺序的或者以其他方式)的任何机器。此外，虽然只例示出单个机器，但是术语“机器”也应当包括单独或联合地执行一组(或多组)指令以执行本文所讨论的方法的任何一种或多种的机器的任意集合，所述方法诸如云计算、软件即服务(SaaS)、其他计算机集群配置。

[0099] 如本文中所述示例可包括逻辑或多个组件、模块或机制，或可通过逻辑或多个组件、模块或机制来操作。电路集是实现于包括硬件(例如，简单电路、门、逻辑等)的有形实体中的一批电路。电路集成员可能随着时间的推移以及底层硬件变化而变化。电路集包括在操作时可单独或组合地执行指定操作的成员。在示例中，电路集的硬件可被不可变地设计为执行特定操作(例如，硬连线)。在示例中，电路集的硬件可包括可变地连接的物理组件(例如，执行单元、晶体管、简单电路等)，包括被物理地(例如，对不变聚集粒子的磁性地、电气地、可移动地布置等)修改的计算机可读介质，用于编码特定操作的指令。在连接物理组件时，硬件组成成分的底层电特性例如从绝缘体改变为导体或反之亦然。这些指令使嵌入式硬件(例如，执行单元或加载机构)能够经由可变连接在硬件中创建电路集的成员以在操作时执行特定操作的部分。相应地，当设备运行时，计算机可读介质通信地耦合到电路集成员的其他组件。在示例中，物理组件中的任一个可在多于一个电路集的多于一个成员中使用。例如，在操作下，执行单元可在时间上的某一点在第一电路集的第一电路中使用而在不同的时间被第一电路集中的第二电路或被第二电路集中的第三电路使用。

[0100] 机器(例如，计算机系统)1100可包括硬件处理器1102(例如，中央处理单元(CPU)、图形处理单元(GPU)、硬件处理器核或者其任何组合)、主存储器1104以及静态存储器1106，这些部件中的一些或全部经由互连链路(例如，总线)1108彼此进行通信。机器1100还可包括显示设备1110、输入设备1112(例如，键盘)以及用户界面(UI)导航设备1114(例如，鼠标)。在示例中，显示设备1110、输入设备1112以及UI导航设备1114可以是触摸屏显示器。机器1100可以附加地包括大容量存储(例如，驱动单元)1116、信号生成设备1118(例如，扬声器)，网络接口设备1120以及一个或多个传感器1121(诸如，全球定位系统(GPS)传感器、罗盘、加速度计或其他传感器)。机器1100可包括与一个或多个外围设备(例如，打印机、读卡器等)连通或者控制这些外围设备的输出控制器1128，诸如串行(例如，通用串行总线(USB))、并行、或者其他有线或无线(例如，红外线(IR)、近场通信(NFC)等)连接。

[0101] 大容量存储1116可包括在其上存储由本文中所描述的技术或功能中的任意一种或多种体现或利用的一组或者多组数据结构或指令1124(例如，软件)的机器可读介质1122。指令1124还可在机器1100执行它的期间完全地或至少部分地驻留在主存储器1104内、驻留在静态存储器1106内、或者驻留在硬件处理器1102内。在示例中，硬件处理器1102、主存储器1104、静态存储器1106或大容量存储1116中的一个或任何组合都可以构成机器可读介质。

[0102] 尽管机器可读介质1122被示为单一介质，但是术语“机器可读介质”可包括被配置

成存储一个或多个指令1124的单一介质或多个介质(例如,集中式或分布式数据库、和/或相关联的高速缓存或服务器)。

[0103] 术语“机器可读介质”可包括能够存储、编码或携带供机器1100执行并且使机器1100执行本公开的任何一项或多项技术的指令,或者能够存储、编码或携带此类指令所使用的或与此类指令相关联的数据结构的任何介质。非限制性机器可读介质示例可以包括固态存储器以及光和磁介质。在示例中,大容量机器可读介质包括具有多个粒子的机器可读介质,这些粒子具有不变(例如,静止)质量。相应地,大容量机器可读介质是非瞬态传播信号。大容量机器可读介质的具体事例可包括:非易失性存储器,诸如,半导体存储器设备(例如,电可编程只读存储器(EPROM)、电可擦除可编程只读存储器(EEPROM))以及闪存设备;磁盘,诸如内部硬盘和可移动盘;磁光盘;以及CD-ROM和DVD-ROM盘。

[0104] 还可以经由利用许多传输协议(例如,帧中继、网际协议(IP)、传输控制协议(TCP)、用户数据报协议(UDP)、超文本传输协议(HTTP)等等)中的任何一种协议的网络接口设备1120,通过使用传输介质的通信网络1126来进一步发送或接收指令1124。示例通信网络可包括局域网(LAN)、广域网(WAN)、分组数据网络(例如,互联网)、移动电话网络(例如,蜂窝网络)、普通老式电话(POTS)网络、以及无线数据网络(例如,称为Wi-Fi®的电气与电子工程师协会(IEEE)802.11系列标准、称为WiMax®的IEEE 802.16系列标准、IEEE 802.15.4系列标准、点对点(P2P)网络等)。在示例中,网络接口设备1120可包括用于连接到通信网络1126的一个或多个物理插口(jack)(例如,以太网、共轴、或电话插口)或者一个或多个天线。在示例中,网络接口设备1120可包括使用单输入多输出(SIMO)、多输入多输出(MIMO),或多输入单输出(MISO)技术中的至少一种来进行无线通信的多个天线。术语“传输介质”应当包括任何无形的介质,所述任何无形的介质能够存储、编码或携带由计算机1100执行的指令,并且“传输介质”包括数字或模拟通信信号或者用于促进此类软件的通信的其他无形的介质。

[0105] 附加注释和示例

[0106] 示例1是一种用于自动视频概括的设备,方法包括:存储设备,用于保存视频;语义分类器,用于从视频的帧生成视频的语义模型;相关性分类器,用于向帧指派相应的相关性得分;多路复用器,用于:使用相应的相关性得分初始化语义模型;以及迭代地处理语义模型以产生子场景集合,每次迭代包括多路复用器,用于:在初始化之后收敛语义模型;在收敛之后选择具有最高相关性得分的帧序列;以及通过固定所选帧序列的相关性得分来重新初始化语义模型。

[0107] 在示例2中,示例1的主题任选地包括,其中生成语义模型包括提取帧的诸特征。

[0108] 在示例3中,示例2的主题任选地包括,其中提取特征包括寻找低级特征。

[0109] 在示例4中,示例3的主题任选地包括,其中低级特征包括GIST描述符。

[0110] 在示例5中,示例3-4中的任何一个或多个的主题任选地包括,其中低级特征包括位置。

[0111] 在示例6中,示例2-5中的任何一个或多个的主题任选地包括,其中提取特征包括寻找高级特征。

[0112] 在示例7中,示例6的主题任选地包括,其中寻找高级特征包括将分类器应用于帧以标识场景特性。

[0113] 在示例8中，示例7的主题任选地包括，应用分类器包括使用为场景分类训练的深度卷积网络。

[0114] 在示例9中，示例7-8中的任何一个或多个的主题任选地包括，其中场景特性包括背景、活动或物体中的一个或多个。

[0115] 在示例10中，示例9的主题任选地包括，背景是室内、室外、海滩上、森林中或城市环境中的一个。

[0116] 在示例11中，示例9-10中的任何一个或多个的主题任选地包括，其中活动是聚会、运动、事件或工作中的一个。

[0117] 在示例12中，示例9-11中的任何一个或多个的主题任选地包括，其中物体是面部、动物、地标或运动设备中的至少一个。

[0118] 在示例13中，示例2-12中的任何一个或多个的主题任选地包括，其中生成语义模型包括从所提取的诸特征生成伪语义域，其中伪语义域是从特征导出的的n维空间。

[0119] 在示例14中，示例13的主题任选地包括，其中从所提取的诸特征生成伪语义域包括使用所提取的特征训练深度波尔兹曼机。

[0120] 在示例15中，示例13-14中的任何一个或多个的主题任选地包括，从所提取的诸特征生成伪语义域包括将谱聚类应用于所提取的诸特征以减少维度。

[0121] 在示例16中，示例13-15中的任何一个或多个的主题任选地包括，其中生成语义模型包括通过为每个帧导出伪语义域中的坐标来将帧映射到伪语义域，坐标的每个元素与伪语义域的维度相对应并且是从所提取的诸特征中为帧所特有的特征的存在性导出的。

[0122] 在示例17中，示例16的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0123] 在示例18中，示例16-17中的任何一个或多个的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0124] 在示例19中，示例16-18中的任何一个或多个的主题任选地包括，其中生成语义模型包括从映射到伪语义域的帧创建生成模型。

[0125] 在示例20中，示例19的主题任选地包括，其中创建生成模型包括创建概率图模型。

[0126] 在示例21中，示例19-20中的任何一个或多个的主题任选地包括，其中创建生成模型包括创建混合模型。

[0127] 在示例22中，示例19-21中的任何一个或多个的主题任选地包括，其中创建生成模型包括通过将具有最高得分的帧添加到帧集合来递归地标识帧中的关键帧集合，帧的得分是针对关键帧集合的所有成员就帧的坐标的平方范数乘以常数然后除以帧与关键帧集合中的帧之间的距离的范数的平方进行的求和的倒数。

[0128] 在示例23中，示例1-22中的任何一个或多个的主题任选地包括，其中向帧分配相应的相关性得分包括从视频提取拍摄该视频的人的行为指示符。

[0129] 在示例24中，示例23的主题任选地包括，其中行为指示符包括缺少用于捕捉该视频的相机的运动。

[0130] 在示例25中，示例23-24中的任何一个或多个的主题任选地包括，其中行为指示符包括用于捕捉该视频的增加的变焦。

[0131] 在示例26中，示例23-25中的任何一个或多个的主题任选地包括，从视频中提取拍摄该视频的人的行为指示符包括限制具有百分之五十或更大几率指示相关性的行为指示符。

[0132] 在示例27中，示例1-26中的任何一个或多个的主题任选地包括，其中向帧指派相应的相关性得分包括将相关性分类器的库应用于帧。

[0133] 在示例28中，示例27的主题任选地包括，其中通过用户反馈修改相关性分类器库。

[0134] 在示例29中，示例27-28中的任何一个或多个的主题任选地包括，其中应用相关性分类器库的输出得分与来自从视频中提取的拍摄该视频的人的行为指示符的输出组合以创建复合相关性得分。

[0135] 在示例30中，示例27-29中的任何一个或多个的主题任选地包括，其中相关性分类器库被过滤以排除具有少于百分之五十指示相关性的的几率的分类器。

[0136] 在示例31中，示例1-30中的任何一个或多个的主题任选地包括，其中使用相应的相关性得分来初始化语义模型包括构建一图形，在该图形中，节点对应于帧，节点值对应于帧的相应的相关性得分，而边缘按语义模型中确定的帧之间的距离倒数来加权。

[0137] 在示例32中，示例31的主题任选地包括，其中边缘在两帧之间的距离超过阈值时被省略或移除。

[0138] 在示例33中，示例31-32中的任何一个或多个的主题任选地包括，其中基于相应的相关性得分，关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。

[0139] 在示例34中，示例33的主题任选地包括，其中与低相关性关键帧和高相关性关键帧相对应的节点被标记为固定值。

[0140] 在示例35中，示例34的主题任选地包括，其中在初始化后收敛语义模型包括为语义模型中不与低相关性关键帧或高相关性关键帧相对应的的节点计算节点值。

[0141] 在示例36中，示例35的主题任选地包括，其中节点之间沿边缘的消息传递被用于计算节点值。

[0142] 在示例37中，示例36的主题任选地包括，其中用于计算节点值的节点之间沿边缘的消息传递包括将每个节点值设置为相邻节点的加权平均值，相邻节点通过边缘连接到该节点，连接该节点与相邻节点的边缘的权重在对相邻节点的值求平均以创建加权平均值之前修改相邻节点的值。

[0143] 在示例38中，示例37的主题任选地包括，其中将每个节点值设置为相邻节点的加权平均值继续进行直到达到收敛阈值。

[0144] 在示例39中，示例38的主题任选地包括，其中收敛阈值定义节点值的变化率，低于该变化率时，收敛被认为完成。

[0145] 在示例40中，示例38-39中的任何一个或多个的主题任选地包括，其中在收敛之后选择具有最高相关性得分的帧序列包括选择具有带有大于选择阈值的值的对应节点的帧。

[0146] 在示例41中，示例40的主题任选地包括，其中选择阈值从节点值的整体来确定的。

[0147] 在示例42中，示例41的主题任选地包括，其中选择阈值是节点值的平均值。

[0148] 在示例43中，示例40-42中的任何一个或多个的主题任选地包括，其中选择帧序列包括创建具有该帧序列的视频剪辑。

[0149] 在示例44中，示例43的主题任选地包括，其中通过固定所选帧序列的相关性得分来重新初始化语义模型包括在语义模型中将所选帧序列标记为固定值。

[0150] 在示例45中，示例44的主题任选地包括，其中固定值被设置为最低可能相关性得分。

[0151] 在示例46中，示例44-45中的任何一个或多个的主题任选地包括，其中高相关性关键帧的新子集被选择并被设置为固定值。

[0152] 在示例47中，示例1-46中的任何一个或多个的主题任选地包括，其中迭代地处理语义模型以产生子场景集合继续进行，直到标识预定数目的子场景。

[0153] 在示例48中，示例47的主题任选地包括，其中预定数据是用户输入。

[0154] 在示例49中，示例47-48中的任何一个或多个的主题任选地包括，其中预定数目基于视频的长度。

[0155] 在示例50中，示例47-49中的任何一个或多个的主题任选地包括，其中预定数目基于从语义模型确定的帧的群集的数目。

[0156] 在示例51中，示例1-50中的任何一个或多个的主题任选地包括控制器，用于：以产生子场景的次序呈现包括这些子场景的用户界面；从用户接收对子场景的丢弃选择；以及从视频概括中丢弃子场景。

[0157] 在示例52中，示例1-51中的任何一个或多个的主题任选地包括，其中相关性分类器用于修改分配相应相关性得分的相关性标准以降低在所丢弃的子场景中发现的相关性标准的重要性。

[0158] 示例53是一种用于自动视频概括的机器实现方法，所述方法包括：获取视频；从该视频的帧生成该视频的语义模型；向帧指派相应的相关性得分；使用相应的相关性得分初始化语义模型；迭代地处理语义模型以产生子场景集合，每次迭代包括：在初始化之后收敛语义模型；在收敛之后选择具有最高相关性得分的帧序列；以及通过固定所选择的帧序列的相关性得分来重新初始化语义模型。

[0159] 在示例54中，示例53的主题任选地包括，其中生成语义模型包括提取帧的诸特征。

[0160] 在示例54中，示例54的主题任选地包括，其中提取诸特征包括寻找低级特征。

[0161] 在示例56中，示例55的主题任选地包括，其中低级特征包括GIST描述符。

[0162] 在示例57中，示例55-56中的任何一个或多个的主题任选地包括，其中低级特征包括位置。

[0163] 在示例58中，示例54-57中的任何一个或多个的主题任选地包括，其中提取诸特征包括寻找高级特征。

[0164] 在示例59中，示例58的主题任选地包括，其中寻找高级特征包括将分类器应用于帧以标识场景特性。

[0165] 在示例60中，示例59的主题任选地包括，其中应用分类器包括使用为场景分类训练的深度卷积网络。

[0166] 在示例61中，示例59-60中的任何一个或多个的主题任选地包括，其中场景特性包括背景、活动或物体中的一个或多个。

[0167] 在示例62中，示例61的主题任选地包括，其中背景是室内、室外、海滩上、森林中或城市环境中的一个。

[0168] 在示例63中，示例61-62中的任何一个或多个的主题任选地包括，其中活动是聚会、运动、事件或工作中的一个。

[0169] 在示例64中，示例61-63中的任何一个或多个的主题任选地包括，其中物体是面部、动物、地标或运动设备中的至少一个。

[0170] 在示例65中，示例54-64中的任何一个或多个的主题任选地包括，其中生成语义模型包括从所提取的诸特征生成伪语义域，其中伪语义域是从特征导出的n维空间。

[0171] 在示例66中，示例65的主题任选地包括，其中从所提取的特征生成伪语义域包括使用所提取的特征训练深度波尔兹曼机。

[0172] 在示例67中，示例65-66中的任何一个或多个的主题任选地包括，其中从所提取的特征生成伪语义域包括将谱聚类应用于所提取的特征以减少维度。

[0173] 在示例68中，示例65-67中的任何一个或多个的主题任选地包括，其中生成语义模型包括通过为每个帧导出伪语义域中的坐标来将帧映射到伪语义域，坐标的每个元素与伪语义域的维度相对应并且是从所提取的诸特征中为帧所特有的特征的存在性导出的。

[0174] 在示例69中，示例68的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0175] 在示例70中，示例68-69中的任何一个或多个的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0176] 在示例71中，示例68-70中的任何一个或多个的主题任选地包括，其中生成语义模型包括从映射到伪语义域的帧创建生成模型。

[0177] 在示例72中，示例71的主题任选地包括，其中创建生成模型包括创建概率图模型。

[0178] 在示例73中，示例71-72中的任何一个或多个的主题任选地包括，其中创建生成模型包括创建混合模型。

[0179] 在示例74中，示例71-73中的任何一个或多个的主题任选地包括，其中创建生成模型包括通过将具有最高得分的帧添加到帧集合来递归地标识帧中的关键帧集合，帧的得分是针对对于关键帧集合的所有成员就帧的坐标的平方范数乘以常数然后除以帧与关键帧集合中的帧之间的距离的范数的平方进行的求和的倒数。

[0180] 在示例75中，示例53-74中的任何一个或多个的主题任选地包括，其中向帧指派相应的相关性得分包括从视频提取拍摄该视频的人的行为指示符。

[0181] 在示例76中，示例75的主题任选地包括，其中行为指示符包括缺少用于捕捉该视频的相机的运动。

[0182] 在示例77中，示例75-76中的任何一个或多个的主题任选地包括，其中行为指示符包括用于捕捉该视频的增加的变焦。

[0183] 在示例78中，示例75-77中的任何一个或多个的主题任选地包括，其中从视频中提取拍摄该视频的人的行为指示符包括限制具有百分之五十或更大几率指示相关性的行为指示符。

[0184] 在示例79中，示例53-78中的任何一个或多个的主题任选地包括，其中向帧分配相应的相关性得分包括将相关性分类器的库应用于帧。

[0185] 在示例80中，示例79的主题任选地包括，其中通过用户反馈修改相关性分类器库。

[0186] 在示例81中，示例79-80中的任何一个或多个的主题任选地包括，其中应用相关性分类器库的输出得分与来自从视频中提取的拍摄该视频的人的行为指示符的输出组合以创建复合相关性得分。

[0187] 在示例82中，示例79-81中的任何一个或多个的主题任选地包括，其中相关性分类器库被过滤以排除具有少于百分之五十指示相关性的的几率的分类器。

[0188] 在示例83中，示例53-82中的任何一个或多个的主题任选地包括，其中使用相应的相关性得分来初始化语义模型包括构建一图形，在该图形中，节点对应于帧，节点值对应于帧的相应的相关性得分，而边缘按语义模型中确定的帧之间的距离倒数来加权。

[0189] 在示例84中，示例83的主题任选地包括，其中边缘在两帧之间的距离超过阈值时被省略或移除。

[0190] 在示例85中，示例83-84中的任何一个或多个的主题任选地包括，其中基于相应的相关性得分，关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。

[0191] 在示例86中，示例85的主题任选地包括，其中与低相关性关键帧和高相关性关键帧对应的节点被标记为固定值。

[0192] 在示例87中，示例86的主题任选地包括，其中在初始化后收敛语义模型包括为语义模型中不与低相关性关键帧或高相关性关键帧相对应的的节点计算节点值。

[0193] 在示例88中，示例87的主题任选地包括，其中节点之间沿边缘的消息传递被用于计算节点值。

[0194] 在示例89中，示例88的主题任选地包括，其中用于计算节点值的节点之间沿边缘的消息传递包括将每个节点值设置为相邻节点的加权平均值，相邻节点通过边缘连接到该节点，连接该节点与相邻节点的边缘的权重在对相邻节点的值求平均以创建加权平均值之前修改相邻节点的值。

[0195] 在示例90中，示例89的主题任选地包括，其中将每个节点值设置为相邻节点的加权平均值继续进行直到达到收敛阈值。

[0196] 在示例91中，示例90的主题任选地包括，其中收敛阈值定义节点值的变化率，低于该变化率时，收敛被认为完成。

[0197] 在示例92中，示例90-91中的任何一个或多个的主题任选地包括，其中在收敛之后选择具有最高相关性得分的帧序列包括选择具有带有大于选择阈值的值的对应节点的帧。

[0198] 在示例93中，示例92的主题任选地包括，其中选择阈值从全部节点值确定。

[0199] 在示例94中，示例93的主题任选地包括，其中选择阈值是节点值的平均值。

[0200] 在示例95中，示例92-94中的任何一个或多个的主题任选地包括，其中选择帧序列包括创建具有该帧序列的视频剪辑。

[0201] 在示例96中，示例95的主题任选地包括，其中通过固定所选帧序列的相关性得分来重新初始化语义模型包括在语义模型中将所选帧序列标记为固定值。

[0202] 在示例97中，示例96的主题任选地包括，其中固定值被设置为最低可能相关性得分。

[0203] 在示例98中，示例96-97中的任何一个或多个的主题任选地包括，其中高相关性关键帧的新子集被选择并被设置为固定值。

[0204] 在示例99中，示例53-98中的任何一个或多个的主题任选地包括，其中迭代地处理语义模型以产生子场景集合继续进行，直到标识预定数目的子场景。

[0205] 在示例100中，示例99的主题任选地包括，其中预定数据是用户输入。

[0206] 在示例101中，示例99-100中的任何一个或多个的主题任选地包括，其中预定数目基于视频的长度。

[0207] 在示例102中，示例99-101中的任何一个或多个的主题任选地包括，其中预定数目基于从语义模型确定的帧的群集的数目。

[0208] 在示例103中，示例53-102中的任何一个或多个的主题任选地包括：以子场景产生的顺序呈现包括这些子场景的用户界面；从用户接收对子场景的丢弃选择；以及从视频概括中丢弃子场景。

[0209] 在示例104中，示例53-103中的任何一个或多个的主题任选地包括，修改分配相应相关性得分的相关性标准以降低在所丢弃的子场景中发现的相关性标准的重要性。

[0210] 示例105是一种系统，包括用于执行示例53-104的方法中的任一方法的装置。

[0211] 示例106是包括指令的至少一种机器可读介质，该指令在被机器执行时，使机器执行示例53-104的方法中的任一方法。

[0212] 示例107是至少一种机器可读介质，包括指令，该指令在被机器执行时，使机器执行用于自动视频概括的操作，该操作包括：获得视频；从该视频的帧生成该视频的语义模型；向帧分配相应的相关性得分；使用相应的相关性得分初始化语义模型；迭代地处理语义模型以产生子场景集合，每次迭代包括：在初始化之后收敛语义模型；在收敛后选择具有最高相关性得分的帧序列；以及通过固定所选择的帧序列的相关性得分来重新初始化语义模型。

[0213] 在示例108中，示例107的主题任选地包括，其中生成语义模型包括提取帧的诸特征。

[0214] 在示例109中，示例108的主题任选地包括，其中提取诸特征包括寻找低级特征。

[0215] 在示例110中，示例109的主题任选地包括，其中低级特征包括GIST描述符。

[0216] 在示例111中，示例109-110中的任何一个或多个的主题任选地包括，其中低级特征包括位置。

[0217] 在示例112中，示例108-111中的任何一个或多个的主题任选地包括，其中提取诸特征包括寻找高级特征。

[0218] 在示例113中，示例112的主题任选地包括，其中寻找高级特征包括将分类器应用于帧以标识场景特性。

[0219] 在示例114中，示例113的主题任选地包括，应用分类器包括使用为场景分类训练的深度卷积网络。

[0220] 在示例115中，示例113-114中的任何一个或多个的主题任选地包括，其中场景特性包括背景、活动或物体中的一个或多个。

[0221] 在示例116中，示例115的主题任选地包括，其中背景是室内、室外、海滩上、森林中或城市环境中的一个。

[0222] 在示例117中，示例115-116中的任何一个或多个的主题任选地包括，其中活动是聚会、运动、事件或工作中的一个。

[0223] 在示例118中，示例115-117中的任何一个或多个的主题任选地包括，其中物体是面部、动物、地标或运动设备中的至少一个。

[0224] 在示例119中，示例108-118中的任何一个或多个的主题任选地包括，其中生成语义模型包括从所提取的特征生成伪语义域，其中伪语义域是从特征导出的n维空间。

[0225] 在示例120中，示例119的主题任选地包括，其中从所提取的特征生成伪语义域包括使用所提取的特征训练深度波尔兹曼机。

[0226] 在示例121中，示例119-120中的任何一个或多个的主题任选地包括，其中从所提取的特征生成伪语义域包括将谱聚类应用于所提取的特征以减少维度。

[0227] 在示例122中，示例119-121中的任何一个或多个的主题任选地包括，其中生成语义模型包括通过为每个帧导出伪语义域中的坐标来将帧映射到伪语义域，坐标的每个元素与伪语义域的维度相对应并且是从所提取的诸特征中为帧所特有的特征的存在性导出的。

[0228] 在示例123中，示例122的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0229] 在示例124中，示例122-123中的任何一个或多个的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0230] 在示例125中，示例122-124中的任何一个或多个的主题任选地包括，其中生成语义模型包括从映射到伪语义域的帧创建生成模型。

[0231] 在示例126中，示例125的主题任选地包括，其中创建生成模型包括创建概率图模型。

[0232] 在示例127中，示例125-126中的任何一个或多个的主题任选地包括，其中创建生成模型包括创建混合模型。

[0233] 在示例128中，示例125-127中的任何一个或多个的主题任选地包括，其中创建生成模型包括通过将具有最高得分的帧添加到帧集合来递归地标识帧中的关键帧集合，帧的得分是针对关键帧集合的所有成员就帧的坐标的平方范数乘以常数然后除以帧与关键帧集合中的帧之间的距离的范数的平方进行的求和的倒数。

[0234] 在示例129中，示例107-128中的任何一个或多个的主题任选地包括，其中向帧分配相应的相关性得分包括从视频提取拍摄该视频的人的行为指示符。

[0235] 在示例130中，示例129的主题任选地包括，其中行为指示符包括缺少用于捕捉该视频的相机的运动。

[0236] 在示例131中，示例129-130中的任何一个或多个的主题任选地包括，其中行为指示符包括用于捕捉该视频的增加的变焦。

[0237] 在示例132中，示例129-131中的任何一个或多个的主题任选地包括，其中从视频中提取拍摄该视频的人的行为指示符包括限制具有百分之五十或更大几率指示相关性的行为指示符。

[0238] 在示例133中，示例107-132中的任何一个或多个的主题任选地包括，其中向帧指派相应的相关性得分包括将相关性分类器的库应用于帧。

[0239] 在示例134中，示例133的主题任选地包括，其中通过用户反馈修改相关性分类器库。

[0240] 在示例135中，示例133-134中的任何一个或多个的主题任选地包括，其中应用相关性分类器库的输出得分与来自从视频中提取的拍摄该视频的人的行为指示符的输出组合以创建复合相关性得分。

[0241] 在示例136中，示例133-135中的任何一个或多个的主题任选地包括，其中相关性分类器库被过滤以排除具有少于百分之五十指示相关性的的几率的分类器。

[0242] 在示例137中，示例107-136中的任何一个或多个的主题任选地包括，其中使用相应的相关性得分来初始化语义模型包括构建以图形，在该图形中，节点对应于帧，节点值对应于帧的相应的相关性得分，而边缘按语义模型中确定的帧之间的距离倒数来加权。

[0243] 在示例138中，示例137的主题任选地包括，其中边缘在两帧之间的距离超过阈值时被省略或移除。

[0244] 在示例139中，示例137-138中的任何一个或多个的主题任选地包括，其中基于相应的相关性得分，关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。

[0245] 在示例140中，示例139的主题任选地包括，其中与低相关性关键帧和高相关性关键帧对应的节点被标记为固定值。

[0246] 在示例141中，示例140的主题任选地包括，其中在初始化后收敛语义模型包括为语义模型中不与低相关性关键帧或高相关性关键帧相对应的的节点计算节点值。

[0247] 在示例142中，示例141的主题任选地包括，其中节点之间沿边缘的消息传递被用于计算节点值。

[0248] 在示例143中，示例142的主题任选地包括，其中用于计算节点值的节点之间沿边缘的消息传递包括将每个节点值设置为相邻节点的加权平均值，相邻节点通过边缘连接到该节点，连接该节点与相邻节点的边缘的权重在对相邻节点的值求平均以创建加权平均值之前修改相邻节点的值。

[0249] 在示例144中，示例143的主题任选地包括，其中将每个节点值设置为相邻节点的加权平均值继续进行直到达到收敛阈值。

[0250] 在示例145中，示例144的主题任选地包括，其中收敛阈值定义节点值的变化率，低于该变化率时，收敛被认为完成。

[0251] 在示例146中，示例144-145中的任何一个或多个的主题任选地包括，其中在收敛之后选择具有最高相关性得分的帧序列包括选择具有带有大于选择阈值的值的对应节点的帧。

[0252] 在示例147中，示例146的主题任选地包括，其中选择阈值从全部节点值确定。

[0253] 在示例148中，示例147的主题任选地包括，其中选择阈值是节点值的平均值。

[0254] 在示例149中，示例146-148中的任何一个或多个的主题任选地包括，其中选择帧序列包括创建具有该帧序列的视频剪辑。

[0255] 在示例150中，示例149的主题任选地包括，其中通过固定所选帧序列的相关性得分来重新初始化语义模型包括在语义模型中将所选帧序列标记为固定值。

[0256] 在示例151中，示例150的主题任选地包括，其中固定值被设置为最低可能相关性得分。

[0257] 在示例152中，示例150-151中的任何一个或多个的主题任选地包括，其中高相关

性关键帧的新子集被选择并被设置为固定值。

[0258] 在示例153中，示例107-152中的任何一个或多个的主题任选地包括，其中迭代地处理语义模型以产生子场景集合继续进行，直到标识预定数目的子场景。

[0259] 在示例154中，示例153的主题任选地包括，其中预定数据是用户输入。

[0260] 在示例155中，示例153-154中的任何一个或多个的主题任选地包括，其中预定数目基于视频的长度。

[0261] 在示例156中，示例153-155中的任何一个或多个的主题任选地包括，其中预定数目基于从语义模型确定的帧的群集的数目。

[0262] 在示例157中，示例107-156中的任何一个或多个的主题任选地包括，其中指令包括：以子场景产生的顺序呈现包括这些子场景的用户界面；从用户接收对子场景的丢弃选择；以及从视频概括中丢弃子场景。

[0263] 在示例158中，示例107-157中的任何一个或多个的主题任选地包括，其中指令包括修改分配相应相关性得分的相关性标准以降低在所丢弃的子场景中发现的相关性标准的重要性。

[0264] 示例159是一种用于自动视频概括的系统，该系统包括：用于获取视频的装置；用于从该视频的帧生成该视频的语义模型的装置；用于向帧指派相应的相关性得分的装置；用于使用相应的相关性得分初始化语义模型的装置；以及用于迭代地处理语义模型以产生子场景集合的装置，每次迭代包括：用于在初始化之后收敛语义模型的装置；用于在收敛后选择具有最高相关性得分的帧序列的装置；以及用于通过固定所选择的帧序列的相关性得分来重新初始化语义模型的装置。

[0265] 在示例160中，示例159的主题任选地包括，其中生成语义模型包括提取帧的诸特征。

[0266] 在示例161中，示例160的主题任选地包括，其中提取诸特征包括寻找低级特征。

[0267] 在示例162中，示例161的主题任选地包括，其中低级特征包括GIST描述符。

[0268] 在示例163中，示例161-162中的任何一个或多个的主题任选地包括，其中低级特征包括位置。

[0269] 在示例164中，示例160-163中的任何一个或多个的主题任选地包括，其中提取诸特征包括寻找高级特征。

[0270] 在示例165中，示例164的主题任选地包括，其中寻找高级特征包括将分类器应用于帧以标识场景特性。

[0271] 在示例166中，示例165的主题任选地包括，其中应用分类器包括使用为场景分类训练的深度卷积网络。

[0272] 在示例167中，示例165-166中的任何一个或多个的主题任选地包括，其中场景特性包括背景、活动或物体中的一个或多个。

[0273] 在示例168中，示例167的主题任选地包括，其中背景是室内、室外、海滩上、森林中或城市环境中的一个。

[0274] 在示例169中，示例167-168中的任何一个或多个的主题任选地包括，其中活动是聚会、运动、事件或工作中的一个。

[0275] 在示例170中，示例167-169中的任何一个或多个的主题任选地包括，其中物体是

面部、动物、地标或运动设备中的至少一个。

[0276] 在示例171中，示例160-170中的任何一个或多个的主题任选地包括，其中生成语义模型包括从所提取的特征生成伪语义域，其中伪语义域是从特征导出的n维空间。

[0277] 在示例172中，示例171的主题任选地包括，其中从所提取的诸特征生成伪语义域包括使用所提取的诸特征训练深度波尔兹曼机。

[0278] 在示例173中，示例171-172中的任何一个或多个的主题任选地包括，其中从所提取的诸特征生成伪语义域包括将谱聚类应用于所提取的特征以减少维度。

[0279] 在示例174中，示例171-173中的任何一个或多个的主题任选地包括，其中生成语义模型包括通过为每个帧导出伪语义域中的坐标来将帧映射到伪语义域，坐标的每个元素与伪语义域的维度相对应并且是从所提取的诸特征中为帧所特有的特征的存在性导出的。

[0280] 在示例175中，示例174的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0281] 在示例176中，示例174-175中的任何一个或多个的主题任选地包括，其中将帧映射到伪语义域(其中深度波尔兹曼机被训练以创建伪语义域)包括通过深度玻尔兹曼机向前馈送帧。

[0282] 在示例177中，示例174-176中的任何一个或多个的主题任选地包括，其中生成语义模型包括从映射到伪语义域的帧创建生成模型。

[0283] 在示例178中，示例177的主题任选地包括，其中创建生成模型包括创建概率图模型。

[0284] 在示例179中，示例177-178中的任何一个或多个的主题任选地包括，其中创建生成模型包括创建混合模型。

[0285] 在示例180中，示例177-179中的任何一个或多个的主题任选地包括，其中创建生成模型包括通过将具有最高得分的帧添加到帧集合来递归地标识帧中的关键帧集合，帧的得分是针对关键帧集合的所有成员就帧的坐标的平方范数乘以常数然后除以帧与关键帧集合中的帧之间的距离的范数的平方进行的求和的倒数。

[0286] 在示例181中，示例159-180中的任何一个或多个的主题任选地包括，其中向帧分配相应的相关性得分包括从视频提取拍摄该视频的人的行为指示符。

[0287] 在示例182中，示例181的主题任选地包括，其中行为指示符包括缺少用于捕捉该视频的相机的运动。

[0288] 在示例183中，示例181-182中的任何一个或多个的主题任选地包括，其中行为指示符包括用于捕捉该视频的增加的变焦。

[0289] 在示例184中，示例181-183中的任何一个或多个的主题任选地包括，其中从视频中提取拍摄该视频的人的行为指示符包括限制具有百分之五十或更大几率指示相关性的行为指示符。

[0290] 在示例185中，示例159-184中的任何一个或多个的主题任选地包括，其中向帧分配相应的相关性得分包括将相关性分类器的库应用于帧。

[0291] 在示例186中，示例185的主题任选地包括，其中通过用户反馈修改相关性分类器库。

[0292] 在示例187中，示例185-186中的任何一个或多个的主题任选地包括，其中应用相

关性分类器库的输出得分与来自从视频中提取的拍摄该视频的人的行为指示符的输出组合以创建复合相关性得分。

[0293] 在示例188中，示例185-187中的任何一个或多个的主题任选地包括，其中相关性分类器库被过滤以排除具有少于百分之五十指示相关性的的几率的分类器。

[0294] 在示例189中，示例159-188中的任何一个或多个的主题任选地包括，其中使用相应的相关性得分来初始化语义模型包括构建一图形，在该图形中，节点对应于帧，节点值对应于帧的相应的相关性得分，而边缘按语义模型中确定的帧之间的距离倒数来加权。

[0295] 在示例190中，示例189的主题任选地包括，其中边缘在两帧之间的距离超过阈值时被省略或移除。

[0296] 在示例191中，示例189-190中的任何一个或多个的主题任选地包括，其中基于相应的相关性得分，关键帧的第一子集被标识为低相关性帧而关键帧的第二子集被标识为高相关性帧。

[0297] 在示例192中，示例191的主题任选地包括，其中与低相关性关键帧和高相关性关键帧对应的节点被标记为固定值。

[0298] 在示例193中，示例192的主题任选地包括，其中在初始化后收敛语义模型包括为语义模型中不与低相关性关键帧或高相关性关键帧相对应的的节点计算节点值。

[0299] 在示例194中，示例193的主题任选地包括，其中节点之间沿边缘的消息传递被用于计算节点值。

[0300] 在示例195中，示例194的主题任选地包括，其中用于计算节点值的节点之间沿边缘的消息传递包括将每个节点值设置为相邻节点的加权平均值，相邻节点通过边缘连接到该节点，连接该节点与相邻节点的边缘的权重在对相邻节点的值求平均以创建加权平均值之前修改相邻节点的值。

[0301] 在示例196中，示例195的主题任选地包括，其中将每个节点值设置为相邻节点的加权平均值继续进行直到达到收敛阈值。

[0302] 在示例197中，示例196的主题任选地包括，其中收敛阈值定义节点值的变化率，低于该变化率时，收敛被认为完成。

[0303] 在示例198中，示例196-197中的任何一个或多个的主题任选地包括，其中在收敛之后选择具有最高相关性得分的帧序列包括选择具有带有大于选择阈值的值的对应节点的帧。

[0304] 在示例199中，示例198的主题任选地包括，其中选择阈值从节点值的整体来确定。

[0305] 在示例200中，示例199的主题任选地包括，其中选择阈值是节点值的平均值。

[0306] 在示例201中，示例198-200中的任何一个或多个的主题任选地包括，其中选择帧序列包括创建具有该帧序列的视频片段。

[0307] 在示例202中，示例201的主题任选地包括，其中通过固定所选帧序列的相关性得分来重新初始化语义模型包括在语义模型中将所选帧序列标记为固定值。

[0308] 在示例203中，示例202的主题任选地包括，其中固定值被设置为最低可能相关性得分。

[0309] 在示例204中，示例202-203中的任何一个或多个的主题任选地包括，其中高相关性关键帧的新子集被选择并被设置为固定值。

[0310] 在示例205中，示例159-204中的任何一个或多个的主题任选地包括，其中迭代地处理语义模型以产生子场景集合继续进行，直到标识预定数目的子场景。

[0311] 在示例206中，示例205的主题任选地包括，其中预定数据是用户输入。

[0312] 在示例207中，示例205-206中的任何一个或多个的主题任选地包括，其中预定数目基于视频的长度。

[0313] 在示例208中，示例205-207中的任何一个或多个的主题任选地包括，其中预定数目基于从语义模型确定的帧的群集的数目。

[0314] 在示例209中，示例159-208中的任何一个或多个的主题任选地包括：用于以产生子场景的次序呈现包括这些子场景的用户界面的装置；用于从用户接收对子场景的丢弃选择的装置；以及用于从视频概括中丢弃子场景的装置。

[0315] 在示例210中，示例159-209中的任何一个或多个的主题任选地包括用于修改分配相应相关性得分的相关性标准以降低在所丢弃的子场景中发现的相关性标准的重要性的装置。

[0316] 以上具体实施方式包括对附图的引用，附图形成具体实施方式的部分。附图通过说明来示出可实践的特定实施例。这些实施例在本文中也称为“示例”。此类示例可包括除所示或所述的那些元件以外的元件。然而，本申请发明人还构想其中只提供所示或所描述的那些元素的示例。而且，本申请发明人还构想相对于特定示例(或者其一个或多个方面)或者相对于本文中所示或所描述的其他示例(或者其一个或多个方面)使用所示或所描述的那些元素(或者其一个或多个方面)的组合或置换的示例。

[0317] 本文献中所涉及的所有公开、专利、和专利文献通过引用整体结合于此，好像通过引用单独地结合。本文献和通过引用所结合的那些文献之间的不一致的用法的情况，在结合的引用中的用法应当被认为是对本文献的用法的补充；对于不可调和的不一致，以本文献中的用法为准。

[0318] 在此文档中，如在专利文档中常见的那样，使用术语“一”(“a”或“an”)以包括一个或多于一个，这独立于“至少一个”或“一个或多个”的任何其他实例或用法。在此文档中，使用术语“或”来指非排他性的“或”，使得“A或B”包括“A但非B”、“B但非A”以及“A和B”，除非另外指示。在所附权利要求书中，术语“包括(including)”和“其中(in which)”被用作相应的术语“包括(comprising)”和“其中(wherein)”的普通英语等价词。此外，在所附权利要求书中，术语“包括”和“包含”是开放式的，也就是说，在权利要求中除此类术语之后列举的那些元件之外的元件的系统、设备、制品或过程仍被视为落在那项权利要求的范围内。此外，在所附权利要求书中，术语“第一”、“第二”、“第三”等仅用作标记，而不旨在对他们的对象施加数值要求。

[0319] 以上描述旨在是说明性的，而非限制性的。例如，上述示例(或者其一个或多个方面)可相互组合使用。诸如，本领域普通技术人员中的一个可通过回顾以上描述来使用其他实施例。摘要用于允许读者快速地确认本技术公开的性质，并且提交此摘要需理解：它不用于解释或限制权利要求书的范围或含义。此外，在以上具体实施方式中，各种特征可以共同成组以使本公开流畅。但这不应被解释为意指未要求保护的所公开特征对任何权利要求而言是必要的。相反，发明性主题可在于少于特定的所公开实施例的所有特征。因此，所附权利要求在此被结合到具体描述中，其中每个权利要求独立成为单独实施例。各实施例的范

围应当参考所附权利要求连同这些权利要求赋予的等同物的全部范围而确定。

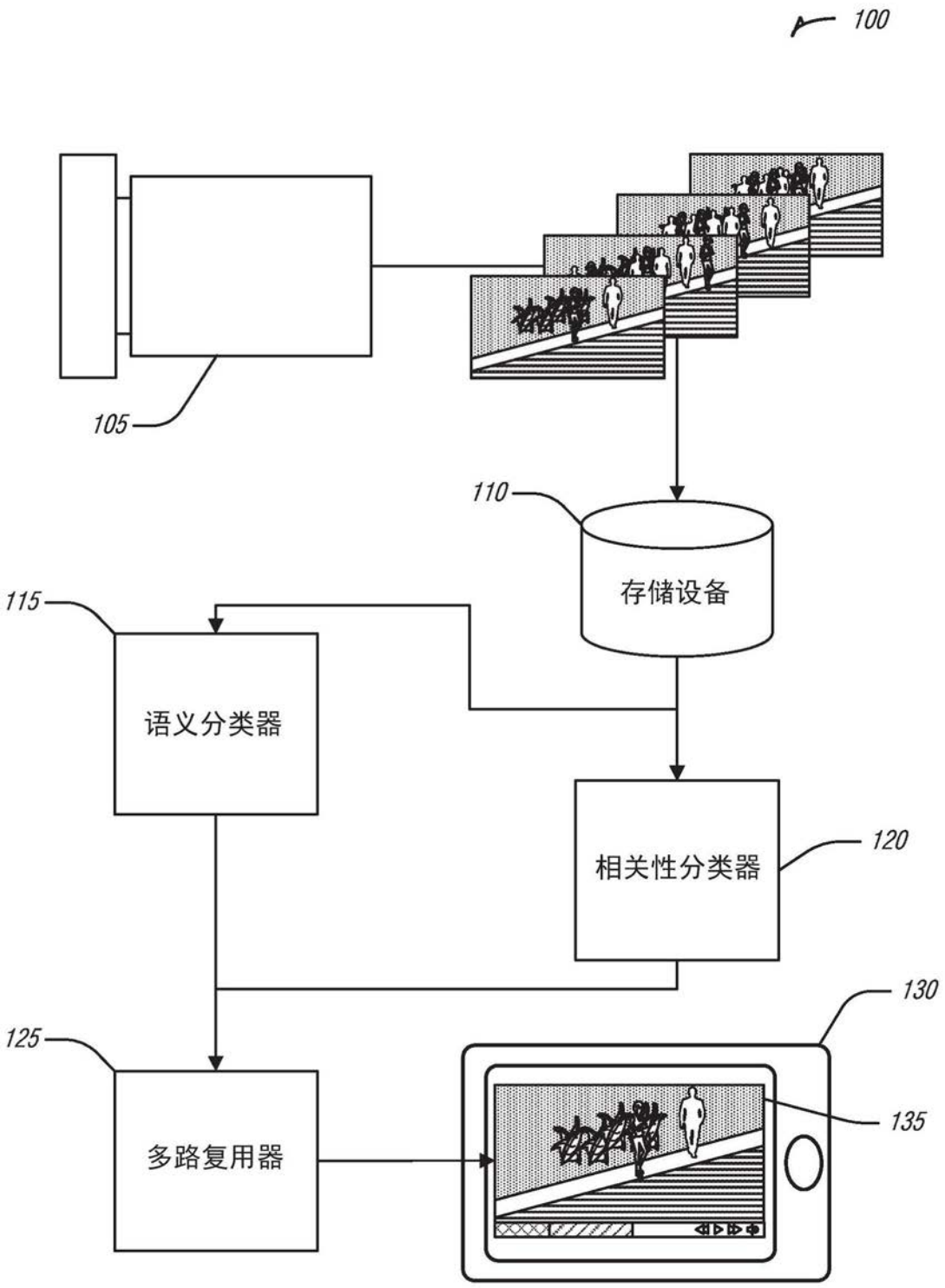


图1

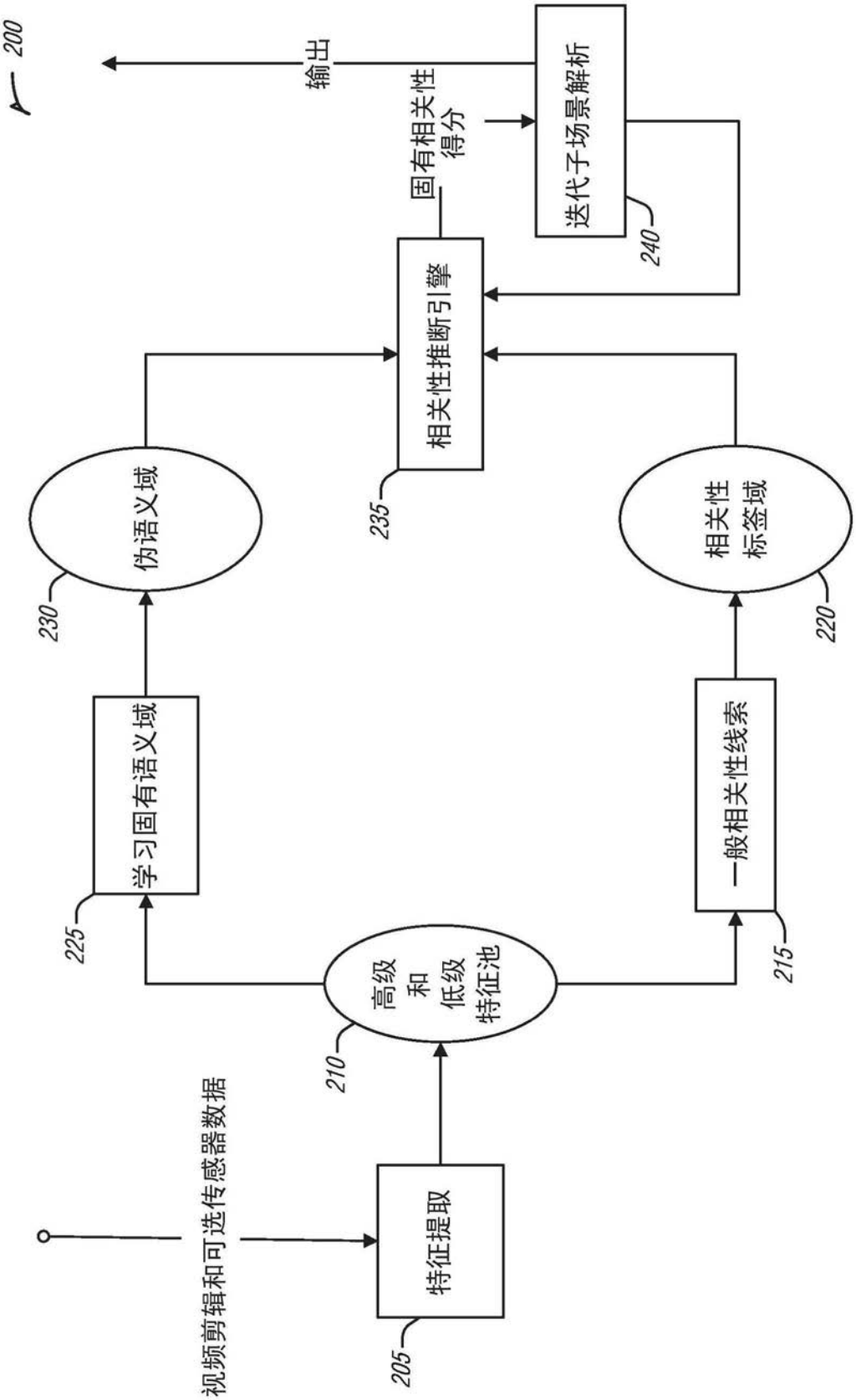

200
输出
固有相关性得分
迭代子场景解析
240
相关性推断引擎
235
伪语义域
230
相关性标签域
220
学习固有语义域
225
一般相关性线索
215
高级和低级特征池
210
特征提取
205
视频剪辑和可选传感器数据

图2

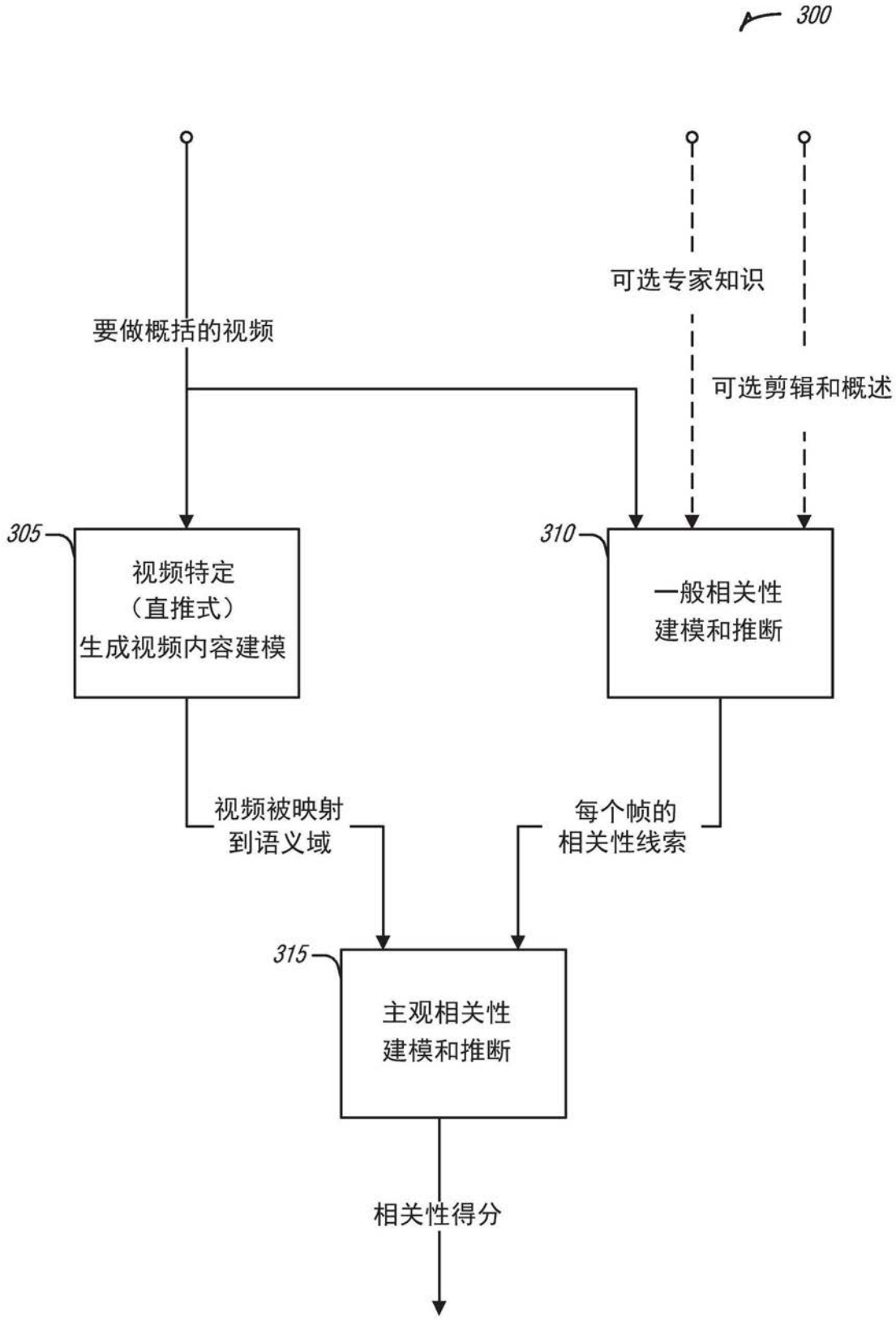

300
可选专家知识
要做概括的视频
可选剪辑和概述
305
视频特定
（直推式）
生成视频内容建模
310
一般相关性
建模和推断
视频被映射
到语义域
每个帧的
相关性线索
315
主观相关性
建模和推断
相关性得分

图3

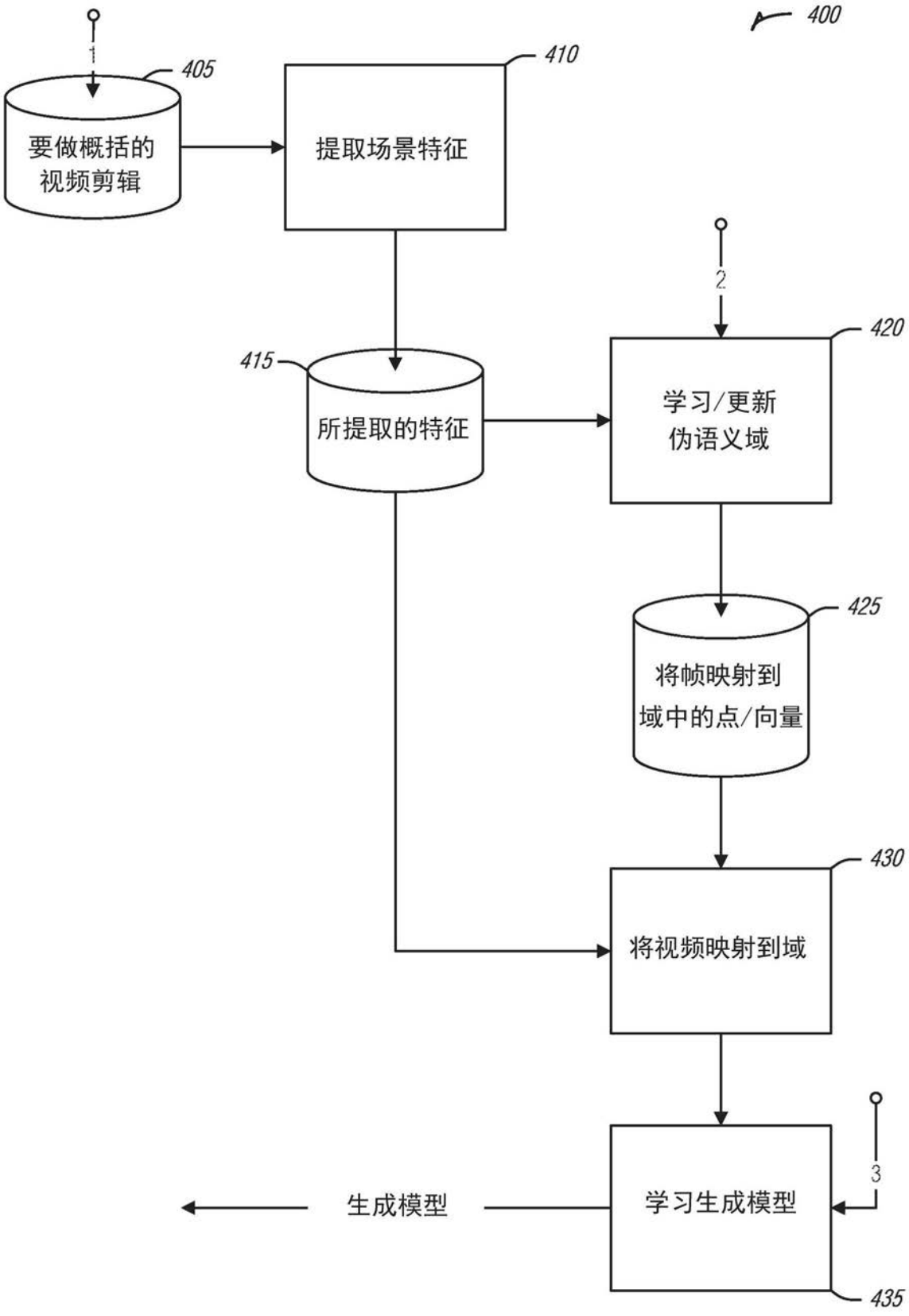

400
1
405
要做概括的
视频剪辑
410
提取场景特征
2
420
415
所提取的特征
学习/更新
伪语义域
425
将帧映射到
域中的点/向量
430
将视频映射到域
3
学习生成模型
435
生成模型

图4

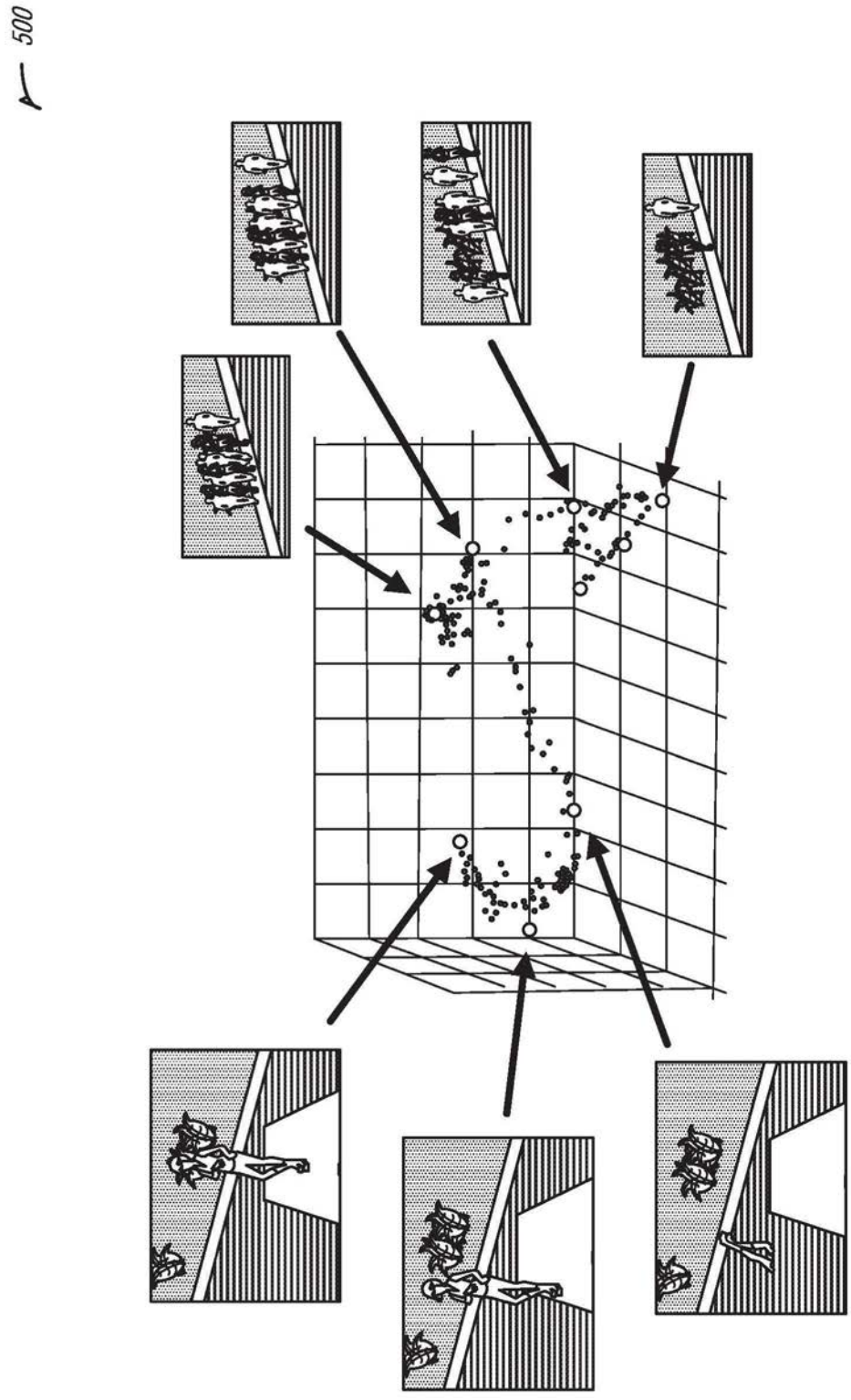


图5

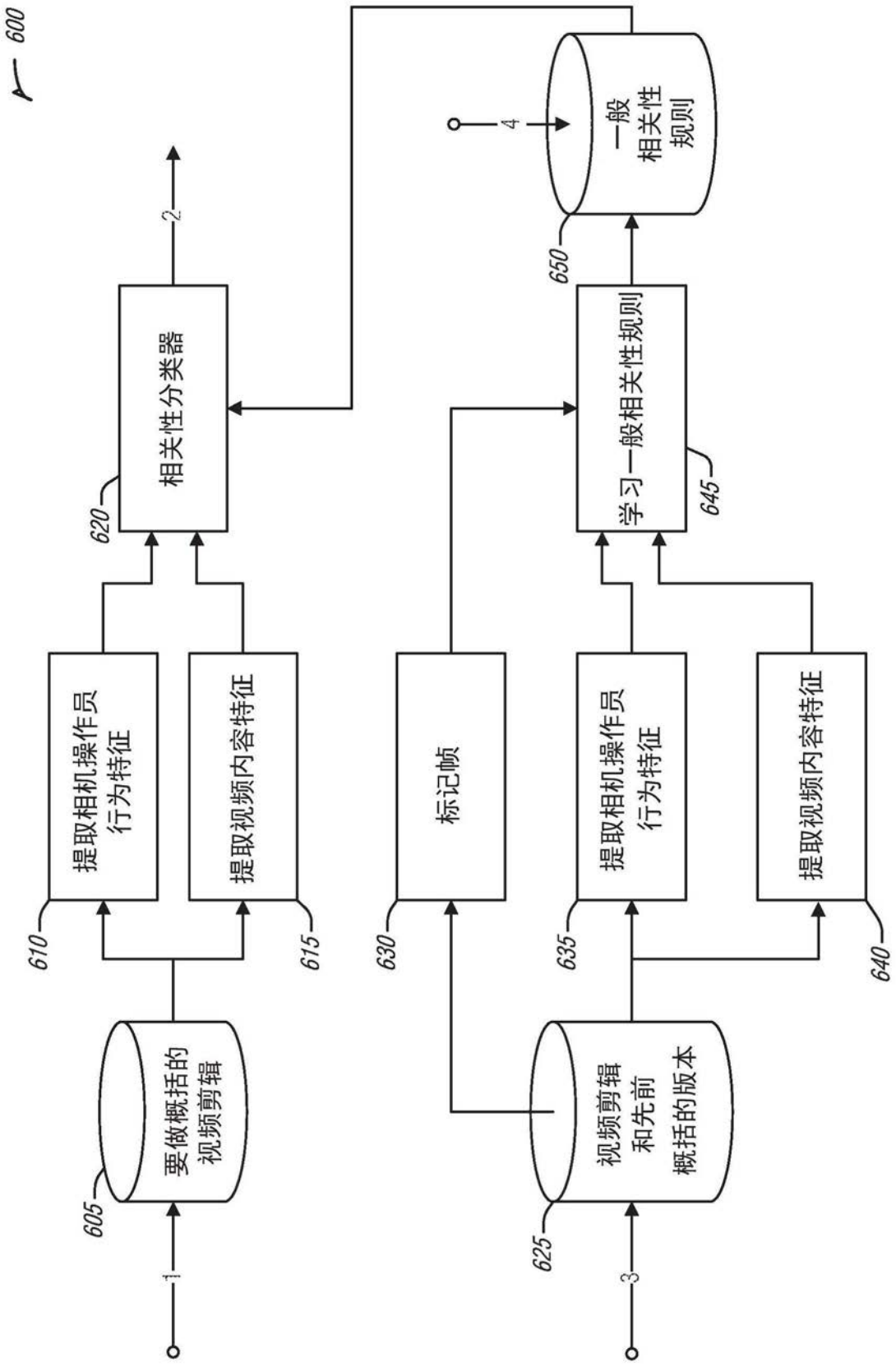

600
要做概括的视频剪辑
605
提取相机操作员行为特征
610
提取视频内容特征
615
相关性分类器
620
1
2
视频剪辑和先前概括的版本
625
标记帧
630
提取相机操作员行为特征
635
提取视频内容特征
640
学习一般相关性规则
645
一般相关性规则
650
3
4

图6

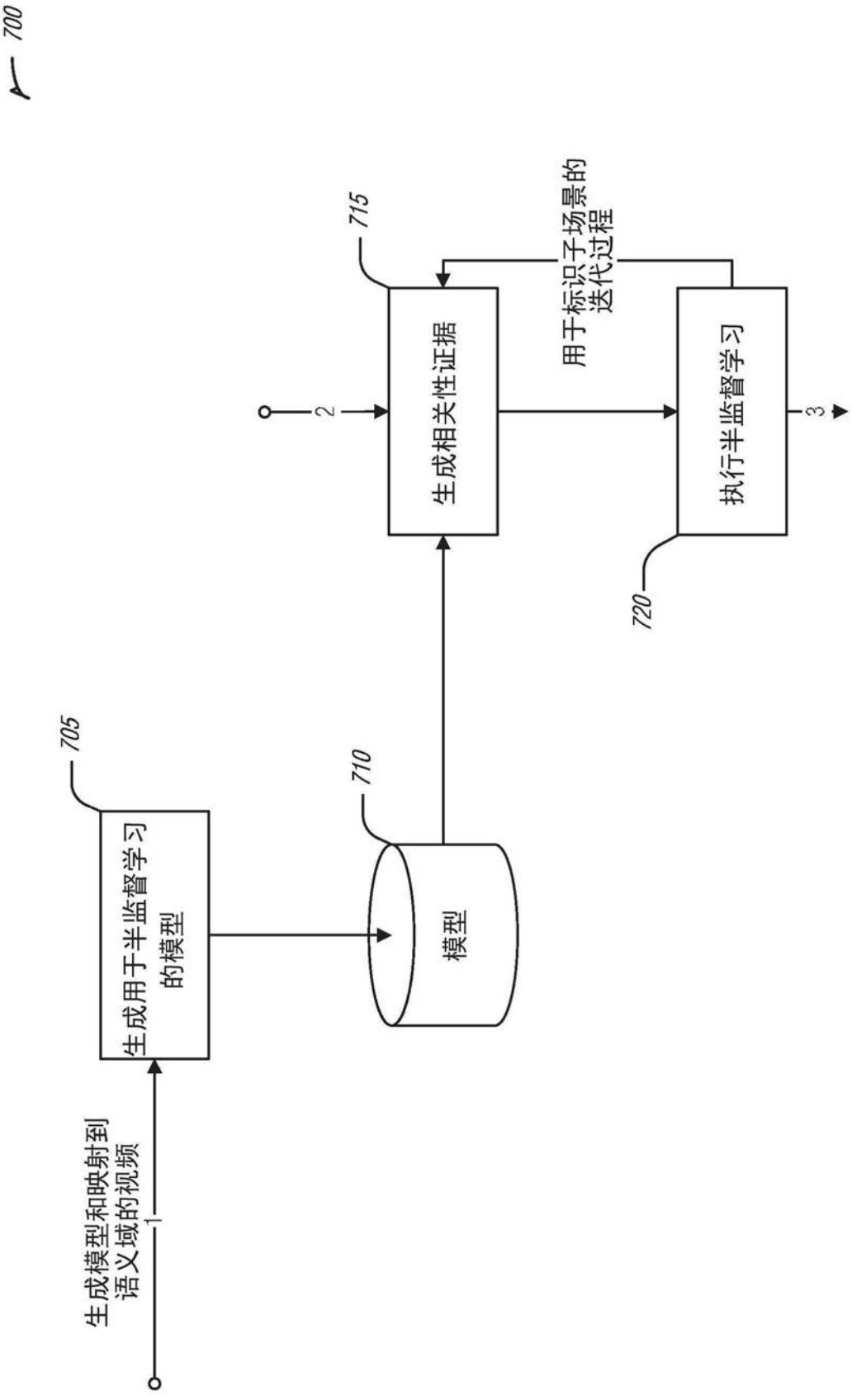

700
生成模型和映射到语义域的视频
1
生成用于半监督学习的模型
705
模型
710
2
生成相关性证据
715
用于标识子场景的迭代过程
执行半监督学习
720
3

图7

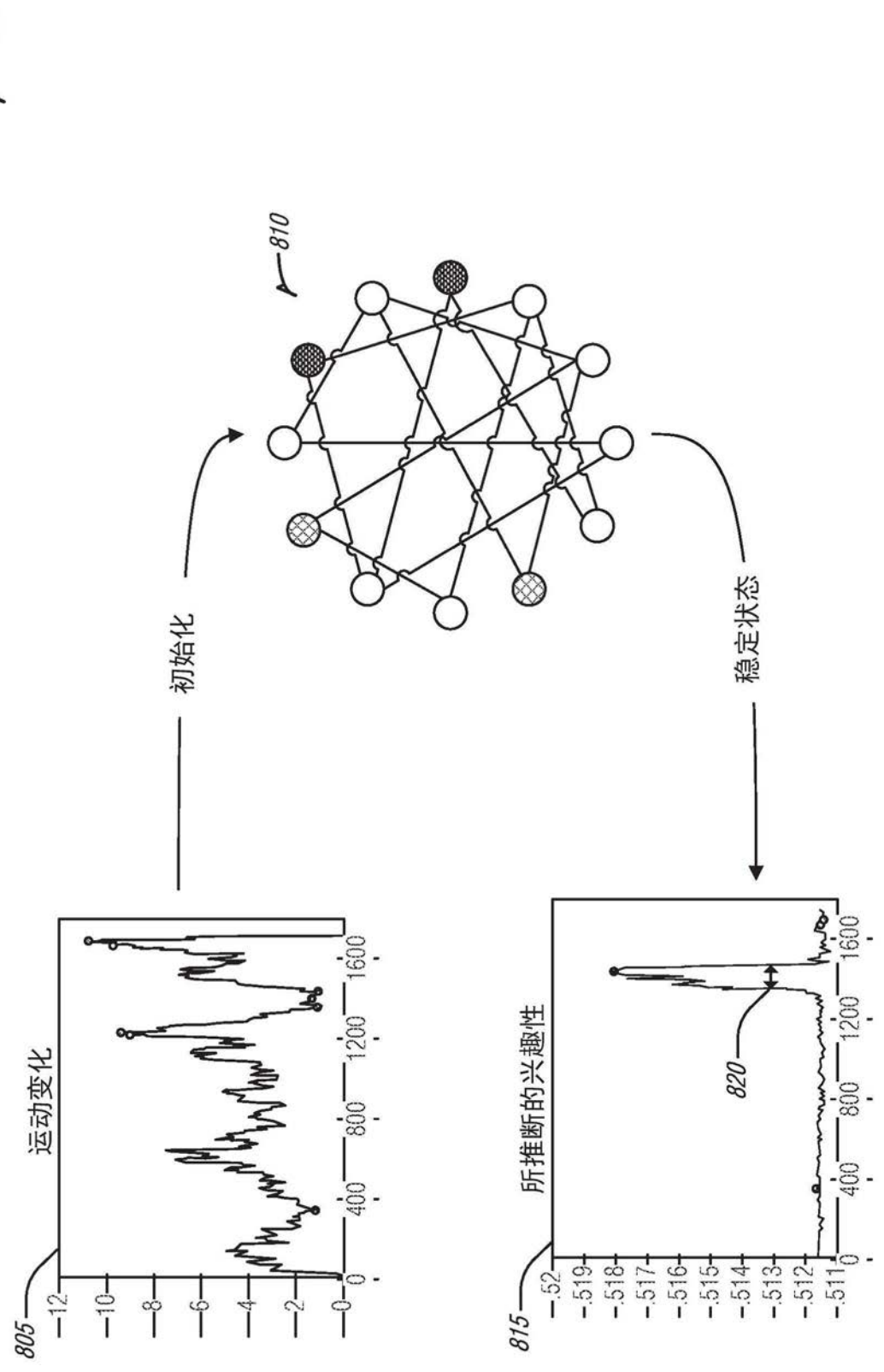

800
810
初始化
稳定状态
运动变化
805
12
10
8
6
4
2
0
0
400
800
1200
1600
所推断的兴趣性
815
820
.52
.519
.518
.517
.516
.515
.514
.513
.512
.511
0
400
800
1200
1600

图8

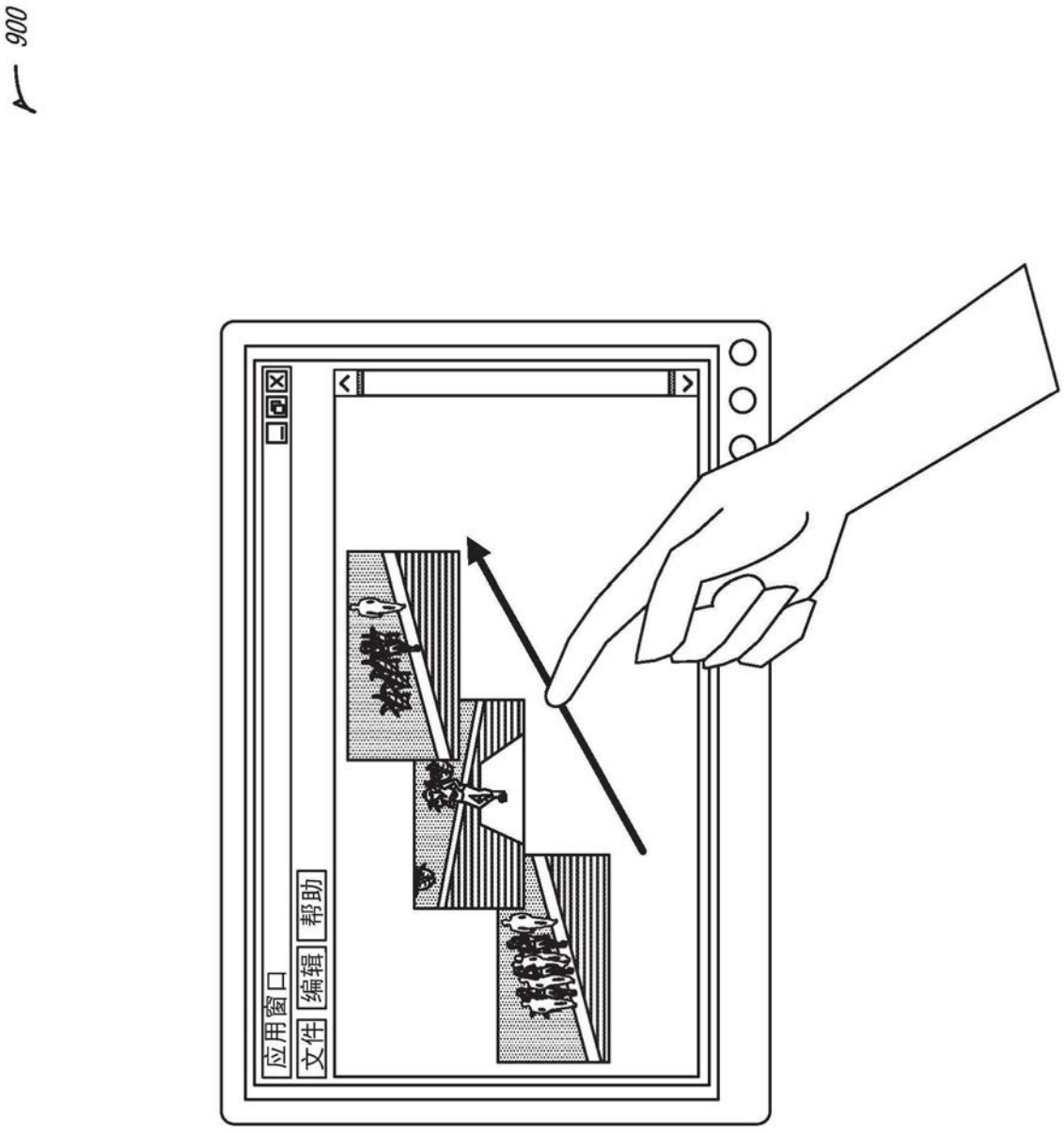


图9

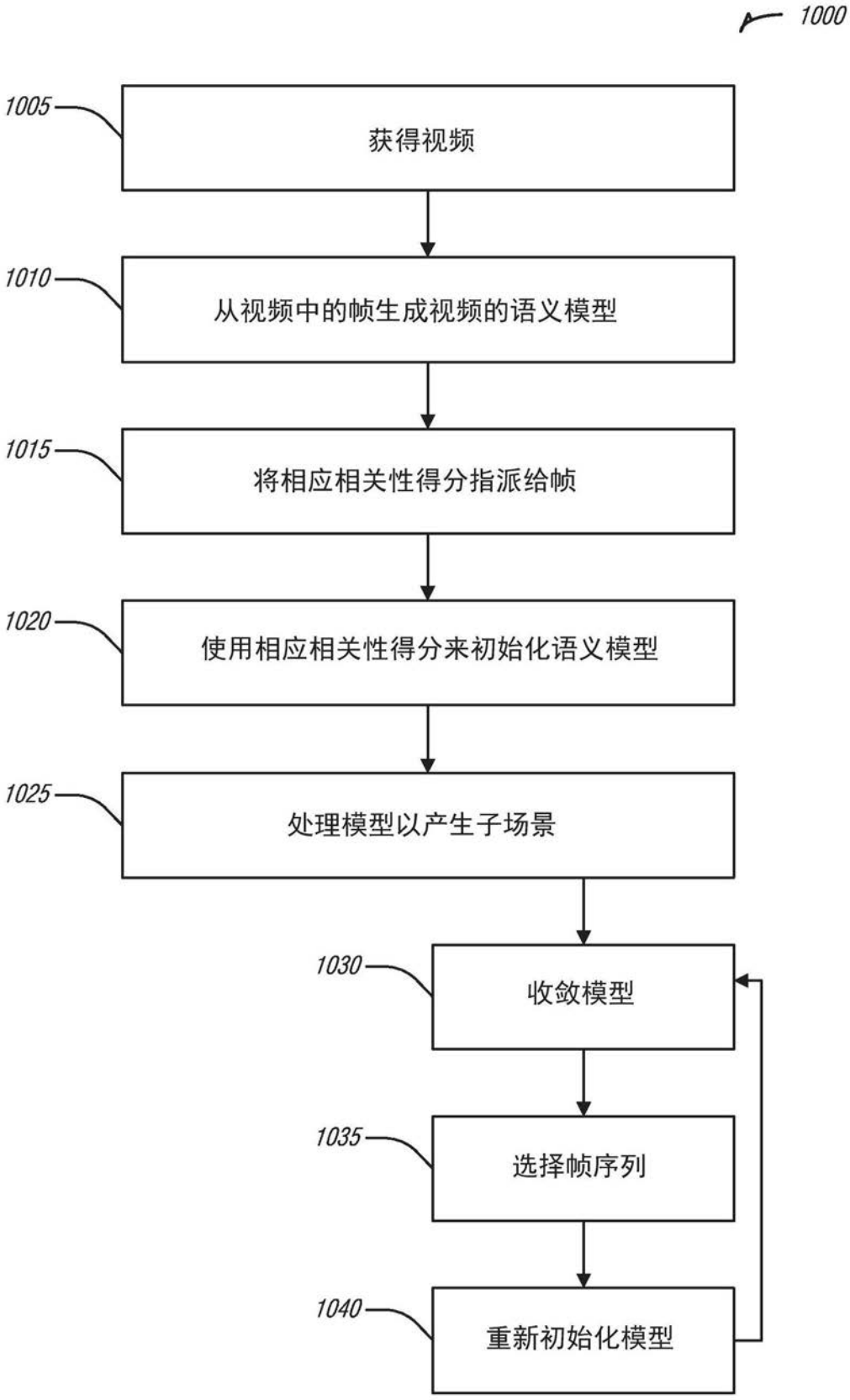

1000
1005
获得视频
1010
从视频中的帧生成视频的语义模型
1015
将相应相关性得分指派给帧
1020
使用相应相关性得分来初始化语义模型
1025
处理模型以产生子场景
1030
收敛模型
1035
选择帧序列
1040
重新初始化模型

图10

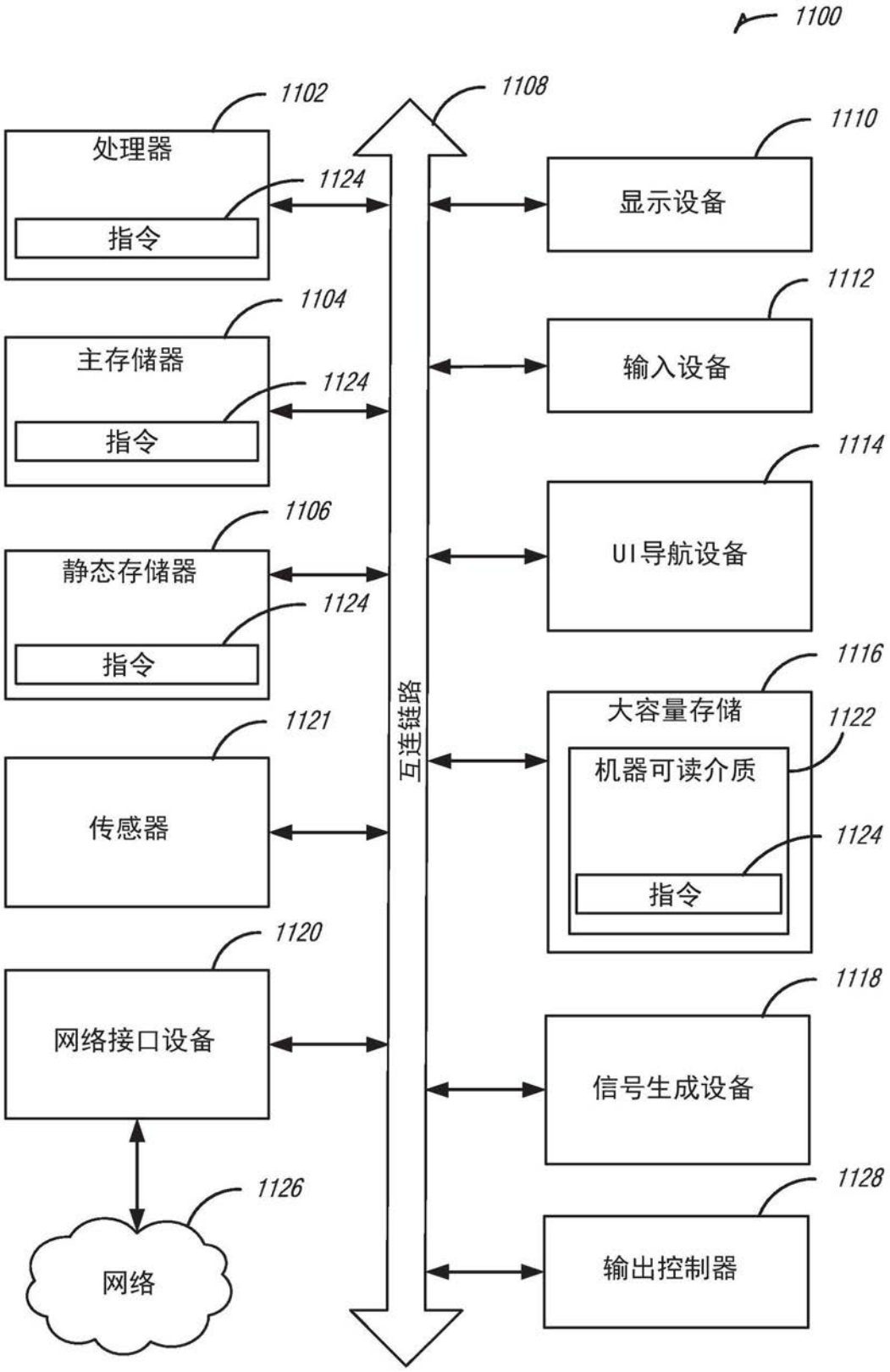

1100
1102
处理器
1124
指令
1104
主存储器
1124
指令
1106
静态存储器
1124
指令
1121
传感器
1120
网络接口设备
1126
网络
1108
互连链路
1110
显示设备
1112
输入设备
1114
UI导航设备
1116
大容量存储
1122
机器可读介质
1124
指令
1118
信号生成设备
1128
输出控制器

图11