



(12) 发明专利

(10) 授权公告号 CN 108376242 B

(45) 授权公告日 2023. 11. 14

(21) 申请号 201810089779.7

HO4N 21/845 (2011.01)

(22) 申请日 2018.01.30

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 108376242 A

CN 101398934 A, 2009.04.01

CN 103559504 A, 2014.02.05

CN 103210651 A, 2013.07.17

(43) 申请公布日 2018.08.07

US 2013216135 A1, 2013.08.22

(30) 优先权数据

CN 104487915 A, 2015.04.01

15/419,281 2017.01.30 US

CN 104756155 A, 2015.07.01

(73) 专利权人 黑莓有限公司

CN 104981680 A, 2015.10.14

地址 加拿大安大略省沃特卢市

CN 102959951 A, 2013.03.06

(72) 发明人 何大可 陈军

穆哈马德·拉比亚·M·阿拉贝亚

尹婷 乔盈婵 王益中 李良燕

P. H. S. TORR ET AL. Feature Based Methods for Structure and Motion Estimation.《Vision Algorithms: Theory and Practice》.SPRINGER BERLIN HEIDELBERG, 2000,第1883卷281.

(74) 专利代理机构 北京市金杜律师事务所

11256

专利代理师 王茂华

审查员 邢雲峰

(51) Int. Cl.

G06V 20/40 (2022.01)

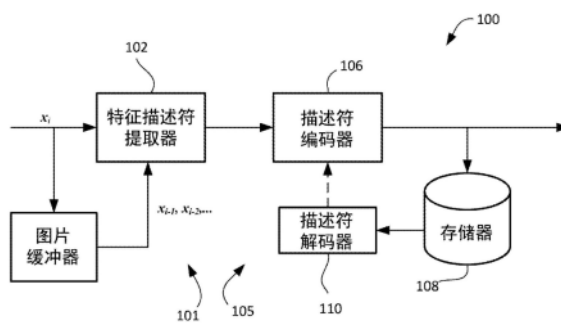
权利要求书3页 说明书10页 附图8页

(54) 发明名称

针对视频的持续性特征描述符

(57) 摘要

用于提取视频的特征描述符的方法和设备, 该视频具有画面序列。该方法包括识别第一关键画面和与序列中迟于第一关键画面的第二关键画面; 从第一关键画面中提取特征描述符的第一集合, 并从第二关键画面中提取特征描述符的第二集合; 识别特征描述符对的集合, 其中每对包括来自第一集合的一个描述符和来自第二集合的一个描述符; 产生描述第一关键画面与第二关键画面之间的运动场的运动信息; 以及基于与运动信息的相关性来过滤特征描述符对的集合以产生和输出持续性描述符的子集。



1. 一种在视频特征描述符提取器中提取视频的特征描述符的方法,所述视频包括画面序列,所述方法包括:

识别第一关键画面和所述序列中迟于所述第一关键画面的第二关键画面;

从所述第一关键画面中提取特征描述符的第一集合,并从所述第二关键画面中提取特征描述符的第二集合;

识别特征描述符对的集合,其中,每对包括来自所述第一集合的一个特征描述符和来自所述第二集合的一个特征描述符;

产生描述所述第一关键画面与所述第二关键画面之间的运动场的运动信息;以及

基于与所述运动信息的相关性过滤所述特征描述符对的集合以产生并输出作为持续性特征描述符的特征描述符对的子集,其中过滤所述特征描述符对的集合包括,针对每一对:

基于所述运动信息和来自所述第一关键画面的对的特征描述符的位置来确定所述第二关键画面中的估计位置;

确定来自所述第二关键画面的所述对的特征描述符是否位于以所述估计位置为中心的搜索窗内;以及

如果是,则将所述对保留在所述持续性特征描述符的所述子集中,以及

如果不是,则将所述对从所述持续性特征描述符的所述子集中排除。

2. 根据权利要求1所述的方法,其中,识别特征描述符对的集合包括:针对所述第一集合中的每个特征描述符,基于将来自所述第一集合的该特征描述符与来自所述第二集合的特征描述符之间的欧氏距离最小化来识别所述第二集合中的特征描述符,并将它们指定为所述特征描述符对的集合中的一对。

3. 根据权利要求1所述的方法,其中,识别特征描述符对的集合包括:针对所述第一集合中的每个特征描述符,确定所述第二集合中是否存在匹配特征描述符,并且如果是,则将它们指定为所述特征描述符对中的一对。

4. 根据权利要求3所述的方法,其中,确定是否存在匹配特征描述符包括:针对所述第一集合中的特征描述符,基于距所述第一集合中的该特征描述符的第一欧氏距离来识别所述第二集合中的最接近特征描述符;基于距所述第一集合中的该特征描述符的第二欧氏距离来识别所述第二集合中的第二最接近特征描述符;以及如果所述第一欧氏距离与所述第二欧氏距离的比值小于预设最大值,则将所述最接近特征描述符指定为所述匹配特征描述符。

5. 根据权利要求4所述的方法,其中,所述预设最大值为0.8。

6. 根据权利要求4所述的方法,其中,所述预设最大值小于0.8。

7. 根据权利要求1所述的方法,其中,识别特征描述符对的集合包括:确定两个或更多个特征描述符对包括所述第二关键画面中的相同特征描述符,并且基于所述确定,对所述两个或更多个特征描述符对中的每一个特征描述符对的质量进行评分,保留最高质量对,并丢弃所述两个或更多个特征描述符对中的其余对。

8. 根据权利要求1所述的方法,其中,产生运动场信息包括:使用光流算法来确定所述第一关键画面中的区域与所述第二关键画面中的区域之间的相对移动。

9. 根据权利要求1所述的方法,其中,提取包括:对所述第一关键画面和所述第二关键

画面应用尺度不变特征变换“SIFT”算法。

10. 根据权利要求1所述的方法,其中,识别包括:将所述画面序列成分段,每个分段具有相应的第一关键画面和相应的第二关键画面。

11. 根据权利要求10所述的方法,其中,每个分段包含相应的一系列画面,每个分段的所述相应的第一关键画面是其系列中的第一画面,并且每个分段的所述相应的第二关键画面是所述序列中的随后分段中的第一画面。

12. 一种用于提取视频的特征描述符的视频特征描述符提取器,所述视频包括画面序列,所述视频特征描述符提取器包括:

处理器;

存储器;以及

编码组件,包含可由所述处理器执行的指令,所述指令当被执行时使得所述处理器:

识别第一关键画面和所述序列中迟于所述第一关键画面的第二关键画面;

从所述第一关键画面中提取特征描述符的第一集合,并从所述第二关键画面中提取特征描述符的第二集合;

识别特征描述符对的集合,其中,每对包括来自所述第一集合的一个特征描述符和来自所述第二集合的一个特征描述符;

产生描述所述第一关键画面与所述第二关键画面之间的运动场的运动信息;以及

基于与所述运动信息的相关性过滤所述特征描述符对的集合以产生并输出作为持续性特征描述符的特征描述符对的子集,

其中所述处理器通过以下操作来过滤所述特征描述符对的集合:针对每一对

基于所述运动信息和来自所述第一关键画面的对的特征描述符的位置来确定所述第二关键画面中的估计位置;

确定来自所述第二关键画面的所述对的特征描述符是否位于以所述估计位置为中心的搜索窗内;以及

如果是,则将所述对保留在所述持续性特征描述符的所述子集中,以及

如果不是,则将所述对从所述持续性描述符的所述子集中排除。

13. 根据权利要求12所述的视频特征描述符提取器,其中,所述处理器通过以下操作来识别所述特征描述符对的集合:针对所述第一集合中的每个特征描述符,基于将来自所述第一集合的该特征描述符与来自所述第二集合的特征描述符之间的欧氏距离最小化来识别所述第二集合中的特征描述符,并将它们指定为所述特征描述符对中的一对。

14. 根据权利要求12所述的视频特征描述符提取器,其中,所述处理器通过以下操作来识别所述特征描述符对的集合:针对所述第一集合中的每个特征描述符,确定所述第二集合中是否存在匹配特征描述符,并且如果是,则将它们指定为所述特征描述符对中的一对。

15. 根据权利要求14所述的视频特征描述符提取器,其中,所述处理器通过以下操作来确定是否存在匹配特征描述符:针对所述第一集合中的特征描述符,基于距所述第一集合中的该特征描述符的第一欧氏距离来识别所述第二集合中的最接近特征描述符;基于距所述第一集合中的该特征描述符的第二欧氏距离来识别所述第二集合中的第二最接近特征描述符;以及如果所述第一欧氏距离与所述第二欧氏距离的比值小于预设最大值,则将所述最接近特征描述符指定为所述匹配特征描述符。

16. 根据权利要求15所述的视频特征描述符提取器,其中,所述预设最大值为0.8。

17. 根据权利要求15所述的视频特征描述符提取器,其中,所述预设最大值小于0.8。

18. 根据权利要求12所述的视频特征描述符提取器,其中,所述处理器通过以下操作来识别所述特征描述符对的集合:确定两个或更多个特征描述符对包括所述第二关键画面中的相同特征描述符,并且基于所述确定,对所述两个或更多个特征描述符对中的每一个的质量进行评分,保留最高质量对,并丢弃所述两个或更多个特征描述符对中的其余对。

19. 根据权利要求12所述的视频特征描述符提取器,其中,所述处理器通过以下操作来产生运动场信息:使用光流算法来确定所述第一关键画面中的区域与所述第二关键画面中的区域之间的相对移动。

20. 根据权利要求12所述的视频特征描述符提取器,其中,所述处理器通过以下操作来提取特征描述符:对所述第一关键画面和所述第二关键画面应用尺度不变特征变换“SIFT”算法。

21. 根据权利要求12所述的视频特征描述符提取器,其中,所述处理器通过以下操作来识别第一关键画面和第二关键画面:将所述画面序列成分段,每个分段具有相应的第一关键画面和相应的第二关键画面。

22. 根据权利要求21所述的视频特征描述符提取器,其中,每个分段包含相应的一系列画面,每个分段的相应的第一关键画面是其系列中的第一画面,并且每个分段的相应的第二关键画面是所述序列中的随后分段中的第一画面。

23. 一种非瞬时性处理器可读介质,其存储用于提取视频的特征描述符的处理器可执行指令,所述视频包括画面序列,其中,所述处理器可执行指令当由视频特征描述符提取器中的处理器执行时使得所述处理器:

识别第一关键画面和所述序列中迟于所述第一关键画面的第二关键画面;

从所述第一关键画面中提取特征描述符的第一集合,并从所述第二关键画面中提取特征描述符的第二集合;

识别特征描述符对的集合,其中,每对包括来自所述第一集合的一个特征描述符和来自所述第二集合的一个特征描述符;

产生描述所述第一关键画面与所述第二关键画面之间的运动场的运动信息;以及

基于与所述运动信息的相关性过滤所述特征描述符对的集合以产生并输出作为持续性特征描述符的特征描述符对的子集,其中过滤所述特征描述符对的集合包括,针对每一对:

基于所述运动信息和来自所述第一关键画面的对的特征描述符的位置来确定所述第二关键画面中的估计位置;

确定来自所述第二关键画面的所述对的特征描述符是否位于以所述估计位置为中心的搜索窗内;以及

如果是,则将所述对保留在所述持续性特征描述符的所述子集中,以及

如果否,则将所述对从所述持续性特征描述符的所述子集中排除。

针对视频的持续性特征描述符

技术领域

[0001] 本申请一般地涉及针对视频的特征描述符,诸如可以用于对象识别或分析,并且特别地涉及用于提取和识别持续性特征描述符的方法和系统。

背景技术

[0002] 特征检测和提取被开发用于图像处理以识别图像内的对象(特征)。这作为计算机视觉、图像搜索、对象识别和其它这类图像分析系统的一部分可能是有用的。特征检测的动作涉及通常使用梯度分析来识别图像中潜在受关注的点或区域(例如边缘、角落等)。特征提取的动作涉及产生包含受关注特征的图像中的一小片的多维向量表示。特征提取的所得输出是特征描述符。

[0003] 然后,在例如对象识别或图像匹配过程中使用特征描述符。例如,来自图像的特征描述符的集合可被存储在包含用于大量图像的特征描述符的集合的查询数据库中。例如,这使得能够实现如面部识别之类的功能。

[0004] 特征检测和提取的概念还可以扩展到视频,其中,视频可被认为是一系列画面。运动图像专家组(MPEG)旨在将用于视频的描述符标准化为用于视频分析的紧凑描述符(CDVA)。MPEG正在探索用于视频处理和分析的工具,目的是减少“所得视频描述符传输或存储的数据量”,并实现“实施方式、应用和数据库的互操作性”。注意,在MPEG CDVA中,针对视频序列的描述符是以逐个图像(或逐帧)方式提取的,然后被压缩。

[0005] 可以从视频的每个画面提取的大量特征描述符导致识别和存储以及传输特征描述符的过程中数据负担较大,即使在经过压缩后仍是如此。提供以下特征描述符提取的系统和方法将是有利的,其保持或改善特征描述符的区分能力(即,准确性),同时减少或保持描述符的数量和复杂度。

附图说明

[0006] 作为示例,现在将参考示出本申请的示例实施例的附图,其中:

[0007] 图1示出了用于视频分析系统的简化紧凑描述符;

[0008] 图2以框图形式示出了视频特征描述符提取器系统的示例;

[0009] 图3以框图形式示出了用于视频中的对象识别的查询系统的示例;

[0010] 图4以框图形式示出了视频特征描述符提取器的示例实施例;

[0011] 图5示出了图示用于视频特征描述符提取的示例过程的流程图;

[0012] 图6示出了图示用于视频特征描述符对的运动一致性评估的示例过程的流程图;

[0013] 图7示出了视频特征描述符提取器的示例简化框图;

[0014] 图8示出了从画面对SIFT描述符的示例提取;以及

[0015] 图9示出了持续性描述符的示例。

[0016] 在不同的附图中可能使用类似的附图标记来表示相似的组件。

具体实施方式

[0017] 本申请描述了用于从视频中提取特征描述符的方法和设备。这些方法和设备依靠于时间冗余来识别“持续性描述符”，其中，持续性描述符是出现在多个连续帧中的特征的特征描述符。这一特征可以通过确定从多个连续帧提取的特征描述符彼此足够匹配以致它们可能涉及相同特征来识别。运动一致性检查可用来确保持续性描述符与视频中对象的运动一致。在一些情况下，持续性描述符可以通过匹配与视频分段(segment)有关的两个关键帧的所提取的描述符来识别，其中关键帧是位于分段的任一端的画面。

[0018] 在第一方面，本申请描述了一种在视频特征描述符提取器中提取针对视频的特征描述符的方法，所述视频包括画面序列。该方法包括识别第一关键画面和序列中迟于第一关键画面的第二关键画面；从第一关键画面中提取特征描述符的第一集合，并从第二关键画面中提取特征描述符的第二集合；识别特征描述符对的集合，其中每对包括来自第一集合的一个描述符和来自第二集合的一个描述符；产生描述第一关键画面与第二关键画面之间的运动场的运动信息；以及基于与运动信息的相关性来过滤特征描述符对的集合以产生和输出持续性描述符的子集。

[0019] 在另一方面，本申请描述了一种被配置成实施这类视频特征描述符提取方法的视频特征提取器。

[0020] 在又一方面，本申请描述了存储计算机可执行程序指令的非瞬时性计算机可读介质，所述计算机可执行程序指令当被执行时使一个或多个处理器执行所描述的视频特征描述符提取方法。

[0021] 本领域普通技术人员通过结合附图回顾以下示例的描述，将会理解本申请的其它方面和特征。

[0022] 在下面的描述中，一些示例实施例可以参考各种特征提取过程，即用于检测和/或表示图像中的局部化特征的算法。这类过程的常见示例包括尺度不变特征变换(SIFT)。其它示例包括加速鲁棒特征(SURF)、梯度位置和方向直方图(GLOH)、局部基于能量的形状直方图(LESH)、定向梯度直方图(HOG)。通常，特征提取器或特征提取过程产生特征描述符，该特征描述符是围绕特定局部特征的图像数据的多维表示。确定要通过特征描述符提取或描述哪些特征可以依赖于许多特征检测过程或算法中的任何一个或多个。示例特征检测过程包括边缘检测、角落检测(有时被称为“受关注”检测)、斑点检测、脊线检测或被设计用于识别图像中潜在受关注的区域的其它算法，所述图像中潜在受关注的区域在识别该图像中的特定对象中具重要性。许多这类算法可以使用梯度分析来识别潜在受关注或具重要性的点或区域。当本申请提及“特征提取”(即分析图像或画面以提取一个或多个特征描述符)时，其并不旨在仅涉及提取(确定描述符)，还可能将特征检测包括为提取过程的一部分。

[0023] 例如，SIFT描述符是检测到的潜在受关注点周围的区域。该区域被划分成十六个矩形，根据这十六个矩形，构造了关于每个矩形内的梯度方向和幅值的16个直方图。从而，描述符形成为这十六个直方图按照矩形的规定扫描顺序的级联。

[0024] 图8示出了画面800以及从该画面800提取的示例SIFT描述符的一部分的示例。特征被识别并且其周围的区域被划分成十六个矩形b1、b2、...、b16。示出了级联直方图的一部分(具体地，是b1和b2的示例直方图)。

[0025] 在下面的描述中，当涉及视频时，术语帧、画面、切片、贴片、量化组和矩形切片组

可以在某种程度上互换使用。本领域技术人员将认识到,画面或帧可以包含一个或多个切片或分段。在一些情况下,连续的一系列帧/画面可被称为“序列”。其它术语可以用在其它视频实施方式中。

[0026] 在本申请中,术语“和/或”旨在涵盖列出的元件的全部可能的组合和子组合,包括单独列出的元件中的任何一个、任何子组合或全部元件,而并不排除附加元件。

[0027] 在本申请中,短语“…或…中的至少一个”旨在涵盖列出的元件中的任何一个或多个,包括单独列出的元件中的任何一个、任何子组合或全部元件,而不一定排除任何附加元件,并且不一定需要全部元件。

[0028] 特征检测和提取被开发用于图像处理以识别图像内的对象(特征)。这可能作为计算机视觉、图像搜索、对象识别和其它这类图像分析系统的一部分是有用的。特征检测的动作涉及通常使用梯度分析来识别图像中潜在受关注的点或区域(例如边缘、角落等)。特征提取的动作涉及产生包含受关注特征的图像中的小片的多维向量表示。特征提取的所得输出是特征描述符。如上所述,SIFT是特征提取器的一个示例。

[0029] 为了从画面中定位和提取描述符,SIFT使用以下步骤:

[0030] 1. 根据局部变化在区域和尺度,识别和定位关键点。

[0031] 2. 在每个关键点处,通过使用表示多个定向中的每一个的多个图像来表示局部图像区域。这类表示对于局部几何失真是鲁棒的。应注意,可能会执行采样以适应关键点的尺度级别。

[0032] 3. 在每个关键点处,从表示关键点尺度上的局部图像区域的多个图像的样本中提取描述符。

[0033] 给定图像作为输入,SIFT可以产生大量关键描述符。例如,可以从尺寸为 384×512 (196,608个像素)的图像中提取数百个关键描述符。

[0034] 那些特征描述符然后被用于例如对象识别或图像匹配过程。例如,来自图像的特征描述符的集合可被存储在包含大量图像的特征描述符的集合的查询数据库中。新图像可以经历特征提取,并且查询系统可以尝试通过比较来自新图像的新提取的特征描述符与数据库中存储的特征描述符的集合来找到查询数据库中的那些特征的匹配。例如,这使得能够实现如面部识别之类的功能。

[0035] 特征检测和提取的概念还可以扩展到视频,这里,视频通常是一系列画面。然而,如将理解的,视频包括大量的画面,导致大量的特征描述符。MPEG旨在将用于视频的描述符标准化为用于视频分析的紧凑描述符(CDVA)。MPEG正在探索用于视频处理和工具,目的是减少“所得视频描述符传输或存储的数据量”,并实现“实施方式、应用和数据库的互操作性”。注意,在MPEG CDVA中,视频序列的描述符是以逐个图像(或逐帧)方式提取的,然后被压缩。也就是说,视频序列中的每个画面被视为单个图像,通过使用类似于SIFT的方法从该图像定位和提取描述符。在从单独画面中提取描述符之后,在描述符的压缩中使用画面间的相关性。

[0036] 图1以框图形式示出了CDVA系统10的示例。CDVA系统10使用类似SIFT的特征描述符提取器12并将视频视为一系列视频帧或画面。来自特征描述符提取器12的输出是每一帧/画面的描述符14的集合。然后,在该实例中,由描述符编码器16对每个描述符集合进行编码,以产生紧凑描述符18的比特流,然后可将其存储在数据库或其它存储器20中。

[0037] 系统10可以在编码过程中包括反馈回路,以通过预测编码来改善描述符的压缩。在这类示例实施方式中,系统10包括描述符解码器22,其解码序列中先前画面的紧凑描述符。在图1中,虚线表示在从当前画面 x_i 提取的描述符的编码中使用来自先前画面 x_{i-1} 的描述符是可选的。进一步注意,在图1中,描述符的编码在一些示例中可能是有损的并且涉及量化。

[0038] 改善特征描述符的准确性(区别能力)和特征描述符的压缩之间的平衡可能是有利的。特征描述符的准确性可以通过评估框架来衡量。例如,MPEG在2015年发布了视频描述符的统一评估框架。在MPEG的示例框架中,可以通过比较两个比值来评估区别准确度(即,性能)。一个比值评估真正比伪负(true positive to false negative)。另一个比值评估伪正比真负(false positive to true negative)。改善的准确性可以改善查询响应准确性。另外,作为示例,就可以保持或改善准确性同时减少每个图像或每个视频的特征描述符的数量而言,导致更高的性能并节省带宽(用于传输特征描述符)、计算资源(用于处理查询)以及存储要求(对于查询数据库或本地图像表示)。

[0039] 本申请提出了旨在改善视频的特征描述符的区别准确度并潜在地减少表征视频序列的特征描述符的数量的方法和系统。为此,本申请提出在特征提取过程中利用视频的时间性质,以便识别和提取更鲁棒和有区别的特征描述符。

[0040] 本申请引入了“持续性”特征描述符的概念。持续性特征描述符(在本文中也称为“持续性描述符”)是在视频序列中的两个或更多个连续画面中持续存在的特征描述符。也就是说,如果序列中的描述符彼此匹配(即,它们在彼此的规定距离内),则来自连续视频帧的描述符序列被统称为这些视频帧的持续性描述符。在一些实施例中,持续性特征描述符可以进一步针对图像的运动场进行测试,并且偏离运动场的那些特征描述符可从持续性描述符排除。典型CDVA测试视频序列的实验结果示出从视频序列提取的大部分描述符是持续性的。

[0041] 图2以框图形式示出了用于产生表征视频序列的持续性描述符的集合的视频特征提取器100的一个示例。视频特征提取器100包括提取组件101和编码组件105。在一些实施例中,提取器100可以排除编码组件105,并且可以简单地输出未编码的描述符集合以供本地或远程存储,或者用于产生一个或多个查询。

[0042] 编码组件105包括描述符编码器106、本地存储或存储器108以及描述符解码器110。包含描述符解码器110的反馈回路使得编码过程能够预测地对描述符进行编码。也就是说,描述符编码器106可以基于来自序列中较早图像的解码描述符来对来自当前图像的当前描述符进行编码。应注意,这种预测编码操作涉及利用视频图像的时间性质,以便在识别和提取描述符之后改善描述符本身的编码和压缩。

[0043] 提取组件101包括描述符提取器102和画面缓冲器104。画面缓冲器104使多个画面对描述符提取器102可用,从而使得描述符提取器102能够在多个连续画面上执行描述符匹配,以识别持续性描述符。

[0044] 在一个实施例中,通过描述符提取器102来实施在视频序列中找到持续性描述符的直接方法:

[0045] 1. 从序列的每一帧中提取全部特征描述符。

[0046] 2. 匹配特征描述符,以识别覆盖序列的持续性特征(如果有的话)。

[0047] 这一两步方法代表了用来获得给定的视频序列的持续性描述符的集合的一种简单和准确的方法。它基于时空特征和与内容的相关性来定位和识别描述符；然而，这在计算上是昂贵的。复杂性的起因是需要独立处理每一帧来提取全部特征描述符并在提取的描述符中针对匹配进行搜索。

[0048] 在另一实施例中，首先分析序列以识别关键帧，而不是搜索视频序列的每一帧或画面。通过分析帧相关性，可以将该序列划分为多个分段，在所述分段中的帧之间呈现出高的帧相关性。关键帧可以是分段中的第一个和最后一个画面。在另一实施例中，关键帧是分段中的第一个画面，使得对于任何一个分段，该分段的关键帧是其第一个画面和来自序列中的下一个分段的第一个画面。高的帧相关性表明两个帧可以产生包含大多数持续性描述符的集合。这已经在实验上证明是正确的。因此，该解决方案从分段中的两个帧（被称为关键帧）而不是全部帧中提取描述符。然后可以在这两个关键帧的描述符之间搜索持续性描述符。在一些实施例中，搜索可以利用几何信息和运动信息两者。术语“关键帧”、“关键帧”和“关键画面”在本文中可互换使用。

[0049] 通常在场景中的不同对象上检测到持续性描述符，其中一些正在移动而另一些则是静止的。描述符跟踪实验已经表明，在这两种情况下，持续性描述符呈现出与其对象相同的运动场，例如，如果在出现在少数连续帧中的静止对象上检测到描述符，则其在这些帧中将不会呈现任何位置变化。这种运动顺从性 (compliance) 有助于检测和保存与显著对象有关的描述符—显著对象是存在于多个帧中并传达场景主要信息的那些对象。

[0050] 现在简要参考图9，图9示出了持续性描述符的一个示例。图9示出了两个画面：第一画面902和第二画面904。在一些示例中，第一画面902和第二画面904是视频中的连续画面。在一些示例中，第一画面902可以是第一关键帧，并且第二画面904是针对视频的给定分段的第二关键帧904。在画面下面示出了两个相应画面902、904中的所识别的特征及其区域。从相应的画面902、904中提取的特征描述符（例如，SIFT）的开始部分被示出为描述符1和描述符2。描述符2基于其与描述符1的相似性（不等同，但位于特定的欧几里德距离内）而与描述符1匹配（配对）。可以针对与运动场的一致性来测试与相同特征有关的两个描述符的配对，以对配对进行验证。

[0051] 在该示例实施例中，由特征提取器102实施的特征提取过程可以由以下给出：

[0052] S1. 从视频序列中选择一对帧 x_{k_1} 和 x_{k_2} 作为关键帧，其中 $k_2 > k_1$ 是两个视频帧的索引。

[0053] S2. 从 x_{k_1} 和 x_{k_2} 中提取特征描述符（例如，通过使用SIFT）。

[0054] S3. 产生从 x_{k_1} 到 x_{k_2} 的运动信息（例如，通过使用运动估计）。

[0055] S4. 根据上述步骤S3中产生的运动信息匹配上述步骤S2中提取的描述符。匹配的描述符被标记为视频帧 $x_{k_1}x_{k_1+1}\cdots x_{k_2}$ 的持续性描述符。

[0056] 应注意，在上述步骤S1中，对 x_{k_1} 和 x_{k_2} 的选择可能取决于运动信息，以确保运动场在选定的一对关键帧之间不中断。换句话说，S1可能包括场景变化检测，以选择关键帧。

[0057] 在上述S3中，可以通过使用在视频编码中通常使用的传统的基于块的运动估计方法来产生从 x_{k_1} 到 x_{k_2} 的运动信息，其中画面被分成块并且针对每一块识别运动向量。在一些实施例中，可以使用光流算法来确定要在S4中使用的运动信息。

[0058] 上述解决方案分析运动场并且通过追随运动场来相对于空间和时间特征定义描

述符。因此,该解决方案能够提取捕获时间结构的一般化描述符。通过使用一般化描述符,该解决方案允许定位和识别关键区域,例如具有时间波动的平坦区域,所述关键区域在其他情况中可能不能通过使用基于图像的描述符提取方案来定位。运动场是用于描述视频序列或分段中的对象的运动的2D向量场。在一些情况下,可以使用光流算法来确定或估计运动场。

[0059] 如上所述,由视频特征提取器100输出的持续性描述符的集合可以用于查询系统中。现在参考图3,图3以框图形式示出了用于视频中的识别的查询系统200的一个示例实施例。在该示例中,系统200包括视频数据库202,该视频数据库202包含针对多个视频分段的持续性特征描述符集合。系统200可以用于确定视频分段在数据库202中是否有匹配;也就是说,是否存在展示出相同功能的视频分段。在一些实施方式中,这可以用于对象识别。

[0060] 系统200包括接收查询数据的查询处理器204,在这种情况下,该查询处理器204包括作为查询的主题的视频分段的持续性描述符。如果持续性描述符被编码,则系统200可以包括解码器206。查询处理器204将形成查询的所接收的(和解码的,如果需要的话)持续性描述符与数据库中的持续性描述符进行比较,以确定是否存在用来声明两个视频段之间的匹配的足够接近的匹配。比较的过程可以包括识别查询分段与数据库分段之间的持续性描述符对,确定配对描述符之间的欧氏距离,以及评估持续性描述符之间的差异是否低于阈值水平。该阈值水平可以基于特定应用所要求的一致性程度来设定。

[0061] 将会理解,从画面中提取的局部描述符可以被概括性地表示为称为画面的全局描述符的单个向量。例如,画面的全局描述符中的每个元素可以指示来自画面的局部描述符的数量,所述局部描述符被量化为已知描述符集合(即,词汇表)中的成员。在这种情况下,全局描述符被构造为给定词汇表的局部描述符的直方图。例如,在一些实施方式中,使用从每个局部描述符获得的差异向量及其在词汇表集合中的匹配来形成全局描述符。全局描述符是重塑为长向量的最终矩阵。

[0062] 具有针对画面的全局描述符的好处是允许在大数据库中针对匹配进行搜索时采用两阶段过程,其中,在第一阶段专门使用较简单、较短的全局描述符,并且只有在第一阶段中找到匹配的情况下才使用局部描述符。

[0063] 针对画面的全局描述符的概念可以被扩展为视频分段的全局描述符。例如,在给定的针对视频分段的词汇表的情况下,可以构建分段中的持续性描述符的直方图并将其用作视频分段的全局描述符。相应地,可以使用以下过程来针对匹配进行搜索:

[0064] 1. 在第一阶段中,将来自查询的全局描述符与数据库中存储的全局描述符进行比较,得到视频序列的集合作为与查询相匹配的候选。

[0065] 2. 在第二阶段中,持续性描述符然后被用于与来自数据库的参考视频的解码描述符执行成对匹配过程

[0066] 现在参考图4,其以框图形式示出了视频特征提取器300的一个示例实施例。该示例实施例依赖于基于SIFT的特征识别和提取以及基于光流的运动估计,但是将理解,其它实施方式可以使用其它特征提取技术和其它运动估计技术。

[0067] 示例性视频特征提取器300接收视频序列作为输入。分段过程302将视频序列划分为分段并识别每个分段内的关键帧。然后特征提取器(诸如SIFT提取器304)识别关键帧中的特征并提取它们的描述符。匹配过程306用来对来自分段的两个关键帧的特征描述符进

行配对。同时,运动估计过程308确定由分段过程302识别的两个关键帧之间的运动场。然后,运动场用于评估运动一致性检查过程310中与运动场一致的特征描述符对。如果一对特征描述符偏离运动场,则该对可以作为有效的特征描述符对被丢弃。从这个意义上而言,运动一致性检查过程310过滤通过匹配过程306产生的特征描述符对的集合,以输出与该分段的运动场一致的特征描述符对的子集。

[0068] 然后,所输出的特征描述符子集可以被存储在诸如本地数据库或其它数据结构的存储器中,或者如该示例中所指示的,可以由描述符编码器312进行编码以产生视频的特征描述符的经编码的比特流。

[0069] 现在将参考图5,其以流程图形式示出了提取视频的特征描述符的一个示例过程400。在该示例性过程400中,如由操作402所示,视频序列首先被划分成分段。如上所述,可以使用场景变化检测和其它这类分析来将该序列划分成分段。在一些实施例中,每个分段都包含每分段预定数量的画面/帧。

[0070] 在操作404中,针对每个分段识别关键帧。在一个示例实施方式中,关键帧(或“关键画面”)是分段的第一帧/画面和分段的最后一帧/画面。在另一示例实施方式中,分段的关键帧是分段的第一帧/画面和序列中下一分段(连续分段)的第一帧/画面。在下面的描述中,分段的两个关键帧可分别被指定为 KF_1 和 KF_2 。

[0071] 在操作406中,从关键帧中提取特征描述符。提取过程可以包括SIFT或任何其它这类图像特征描述符提取过程。然后,在操作408中,找到匹配描述符对。在该示例中,针对第一关键帧中的每个描述符,系统查找第二关键帧中是否存在匹配描述符。匹配可以包括测试第二关键帧中的描述符以确定特定搜索窗内的该关键帧中是否存在与第一关键帧中的描述符足够相似的描述符。在一些情况下,在作为多维向量的描述符的情况下,可以根据欧氏距离来测量相似性。在示例实施方式中,通过以下操作可以找到描述符对:通过针对从第一关键帧提取的每个描述符基于它们之间的最小欧氏距离来识别来自第二关键帧的描述符。在另一个示例实施方式中,应用比值测试来确定是否存在匹配。例如,比值测试可能涉及:

[0072] 1. 对于来自 KF_1 的每个描述符(d_i),找到该描述符(d_i)与 KF_2 中的全部描述符之间的欧氏距离,并且仅保留最小的两个距离。

[0073] 2. 计算第一最小距离(到 d_i 的第一最接近描述符)与第二最小距离(到 d_i 的第二最接近描述符)的比值:

$$[0074] \quad r = \frac{dist_1}{dist_2}$$

[0075] 其中 $dist_1$ 和 $dist_2$ 分别是第一最小距离和第二最小距离。

[0076] 3. 如果 $r < 0.8$,则来自 KF_1 的描述符 d_i 与 KF_2 中的(在之前的步骤中查找到的)第一最接近描述符匹配。否则, d_i 在 KF_2 中没有匹配。换句话说,对于将与来自第一关键帧的描述符进行匹配的第二关键帧中的候选描述符,它必须比第二关键帧中的下一个最接近候选描述符匹配得好至少20%(按照欧氏距离)。

[0077] 应该理解,可以通过调整被指定为匹配的比值结果来调整对是否存在匹配的确定的灵敏度。在上述示例中,如果比值 < 0.8 ,则识别出匹配描述符,但是,根据系统在识别匹配方面的选择性,也可以使用其它预设的最大值,例如0.6或0.7。在一些情况下,可以通过经

验测试来选择最大比值。

[0078] 其中针对第一关键帧中的每个描述符寻找第二关键帧中的匹配描述符的示例匹配过程导致第一关键帧中的多于一个描述符可能与第二关键帧中的相同描述符匹配,如操作410所示。在存在涉及与来自第二关键帧的相同描述符的匹配的多个对的情况下,在操作412中,系统确定涉及来自第二关键帧的相同描述符的最佳匹配并丢弃其它匹配。最佳匹配可以通过多种方式进行评估。在一个示例中,确定每个匹配的质量并保留其中的最佳匹配。最佳匹配可能依赖于第一关键帧中的描述符与第二关键帧中的描述符之间的欧几里德差异的比较。在一个示例中,基于操作408的比值测试给每个对赋予评分。例如,可以基于以下表达式对每个对进行评分,其中r是操作408中比值测试的结果:

$$[0079] \quad \beta = \cos\left(\frac{r\pi}{2}\right)$$

[0080] 来自 KF_1 的具有与来自 KF_2 的描述符的最高评分 β 的描述符被声明为最佳匹配。因此,产生唯一的描述符对,即来自 KF_2 的描述符仅与来自 KF_1 的一个描述符匹配。

[0081] 在操作414中,可以应用几何一致性测试来验证描述符对。没有通过几何一致性测试的那些对从针对分段的描述符对的集合中消除。在2014年的N14961“Test Model 12: compact descriptors for visual search”中提供了几何一致性测试的示例,其内容通过引用并入本文。该测试检查所述描述符对是否可能代表两个帧中的相同区域。在查找到M个匹配对的初始集合 $S_{\text{初始-匹配}} = \{(d^{KF1}, d^{KF2})_1, \dots, (d^{KF1}, d^{KF2})_M\}$ 之后,可以构造称为Z的对数距离比(LDR)的矩阵。Z中的每个条目代表以下比值:

$$[0082] \quad z_{ij} = \log \frac{\|x_i - x_j\|}{\|y_i - y_j\|}$$

[0083] 在以上表达式中, x_i 和 y_i 是 $S_{\text{初始-匹配}}$ 中的第i个匹配对的坐标,并且 x_j 和 y_j 是同一集合 $S_{\text{初始-匹配}}$ 中的第j个匹配对的坐标。可以构造矩阵Z中的值的直方图并使用例如皮尔逊拟合优度检验来针对已知的离群概率分布进行检查。

[0084] 在操作416中,系统可以确定从第一关键帧到第二关键帧的运动场,并且可以针对与所确定的运动场的一致性来评估描述符对的集合中的每一对。丢弃与运动场过于不一致的对,得到可以被指定为针对该视频分段的持续性描述符的描述符对的子集。然后,在操作418中输出持续性描述符的子集,特别是每一对中的第一描述符。将理解的是,将第一关键帧的描述符作为特定分段的持续性描述符输出。

[0085] 图6以流程图形式示出了与运动一致性测试有关的操作416的一个示例实施方式。一般而言,运动一致性测试涉及评估单独描述符对(根据它们在帧中的相对位置)通常是否符合帧之间的运动场。在该示例中,系统使用光流算法来确定该分段的运动场,如操作416-1所示。在1981年的International Joint Conference on Artificial Intelligence中的B.D.Lucas和T.Kanade的“An Iterative Image Registration Technique with an Application to Stereo Vision”中描述了光流算法的一个示例,但也可以使用其它这类算法。

[0086] 在操作416-2中,系统以第一关键帧中的描述符开始,并使用运动场精确定位第二关键帧中的特征位置的估计位置。如操作416-3所示,围绕第二关键帧中的估计位置建立搜索窗。然后,在操作416-4中,系统确定来自该对的匹配描述符是否落入搜索窗内。如果是,

那么它通常符合运动场。如果否,那么它偏离运动场太远,并且在操作416-5中,它被丢弃。这可以被认为是对描述符对的集合进行过滤以去除与运动场不匹配的那些描述符。如操作416-6所示,针对为该分段识别的每一个描述符对执行该操作。

[0087] 如上所述,可以设置标准以评估特征描述符提取方案的成对匹配能力。一个示例使用真正(TP)和伪正(FP)比值。它们由以下两个等式给出:

$$[0088] \quad TPR = TP / (TP + FN)$$

$$[0089] \quad FPR = FP / (FP + TN)$$

[0090] 其中:

[0091] TP是真正的数量。当被测试的两个分段被声明为匹配而事实上它们彼此匹配时,出现真正。

[0092] FN是伪负的数量。当被测试的两个分段被声明为不匹配而它们彼此匹配时,出现伪负。

[0093] FP是伪正的数量。当被测试的两个分段被声明为匹配而它们不匹配时,出现伪正。

[0094] TN是真负的数量。当被测试的两个分段被声明为不匹配而它们彼此不匹配时,出现真负。

[0095] 目前描述的解决方案的实施例在不对视频特征提取器的输出处的每秒平均描述符长度施加任何限制(对输出比特率没有上限)的情况下实现良好的TPR和FNR。假设在解码器处可以完美地恢复近似的持续性集合,即使用无损编码方案来实施实验。记录的TPR和FPR在测试中分别为99%和小于1%。

[0096] 现在参考图7,图7示出了视频特征提取器500的示例实施例的简化框图。编码器500包括处理器502、存储器504和视频特征描述符提取应用506。应用506可以包括存储在存储器504中并包含指令的计算机程序或应用,所述指令在被执行时使得处理器502执行诸如本文所述的操作。例如,应用506可以根据本文描述的过程来识别和提取多个持续性特征描述符集合。在一些实施方式中,应用506可以对持续性特征描述符集合进行编码,并将其输出为编码比特流。应当理解的是,应用506可以被存储在诸如光盘、闪存设备、随机存取存储器、硬盘驱动器等非瞬时性计算机可读介质上。当指令被执行时,处理器502执行指令中指定的操作和功能,以便作为实施所描述过程的专用处理器来操作。在一些示例中,这类处理器可被称为“处理器电路”或“处理器电路系统”。

[0097] 将理解的是,根据本申请的视频特征提取器可以在许多计算设备中实施,包括但不限于服务器、适当编程的通用计算机、音频/视频编码设备以及移动设备。视频特征提取器可以通过包含用于配置一个或多个处理器以执行本文所描述的功能的指令的软件来实施。软件指令可以存储在任何合适的非瞬时性计算机可读存储器上,包括CD、RAM、ROM、闪存等。

[0098] 将了解的是,可以使用标准计算机编程技术和语言来实现本文描述的视频特征提取器以及实现用于配置视频特征提取器的所描述的方法/过程的模块、例程、过程、线程或其它软件组件。本申请不限于特定的处理器、计算机语言、计算机编程约定、数据结构、其它这类实施方式细节。本领域技术人员将认识到,所描述的过程可以被实施为存储在易失性或非易失性存储器中的计算机可执行代码的一部分,作为专用集成电路(ASIC)等的一部分。

[0099] 可以对所描述的实施例进行某些改编和修改。因此,上面讨论的实施例被认为是说明性的而不是限制性的。

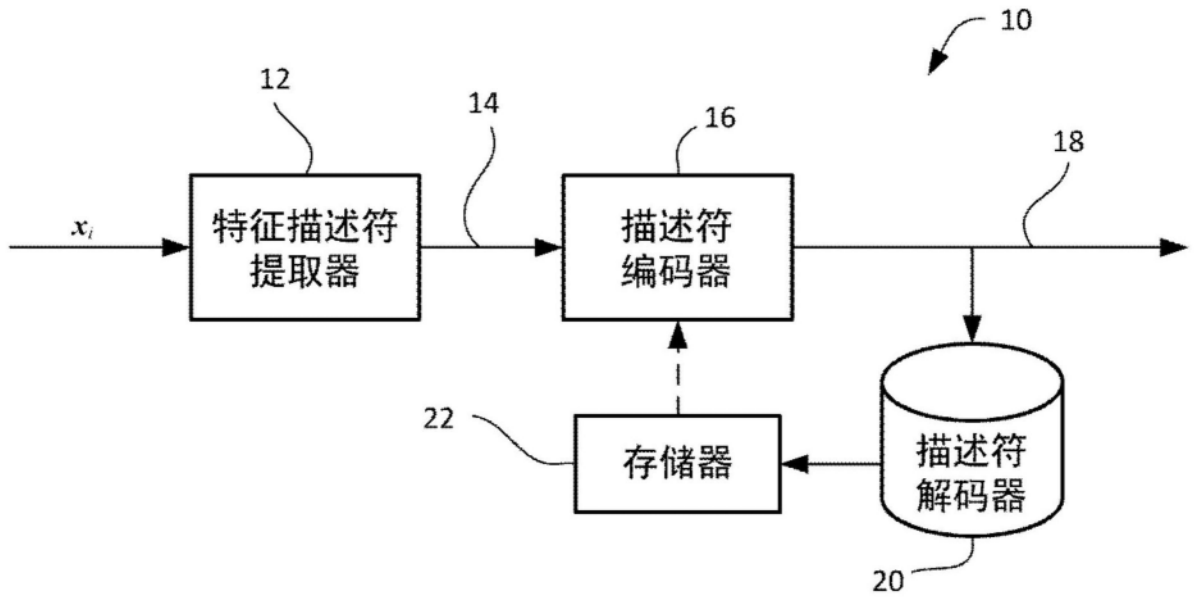


图1

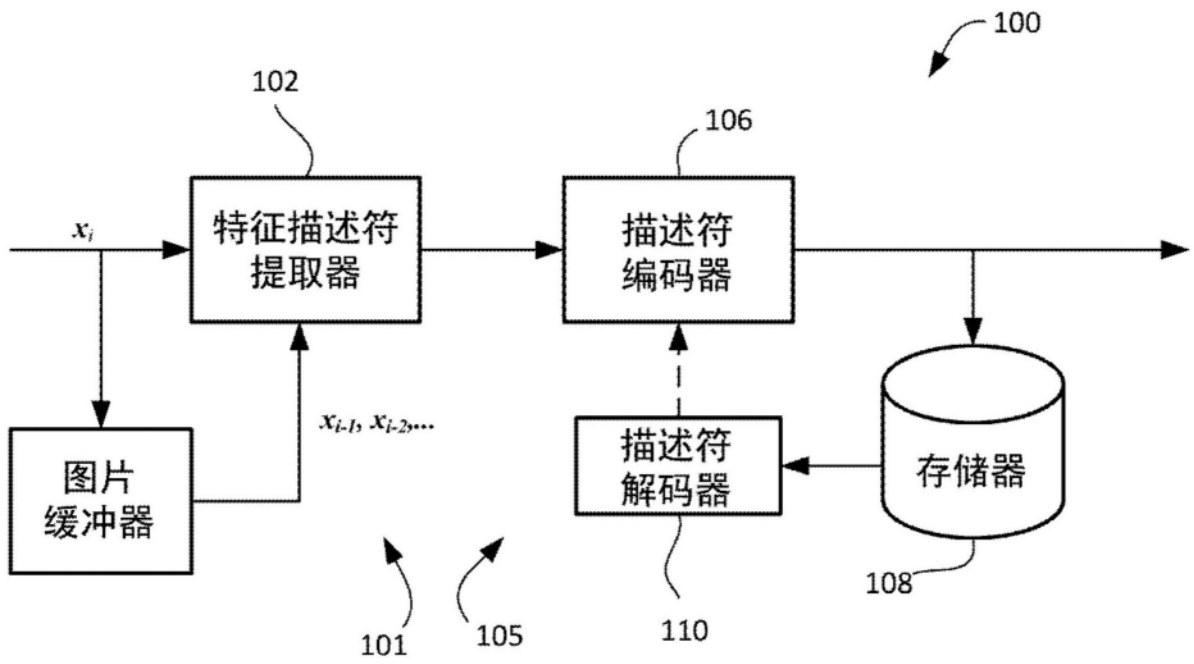


图2

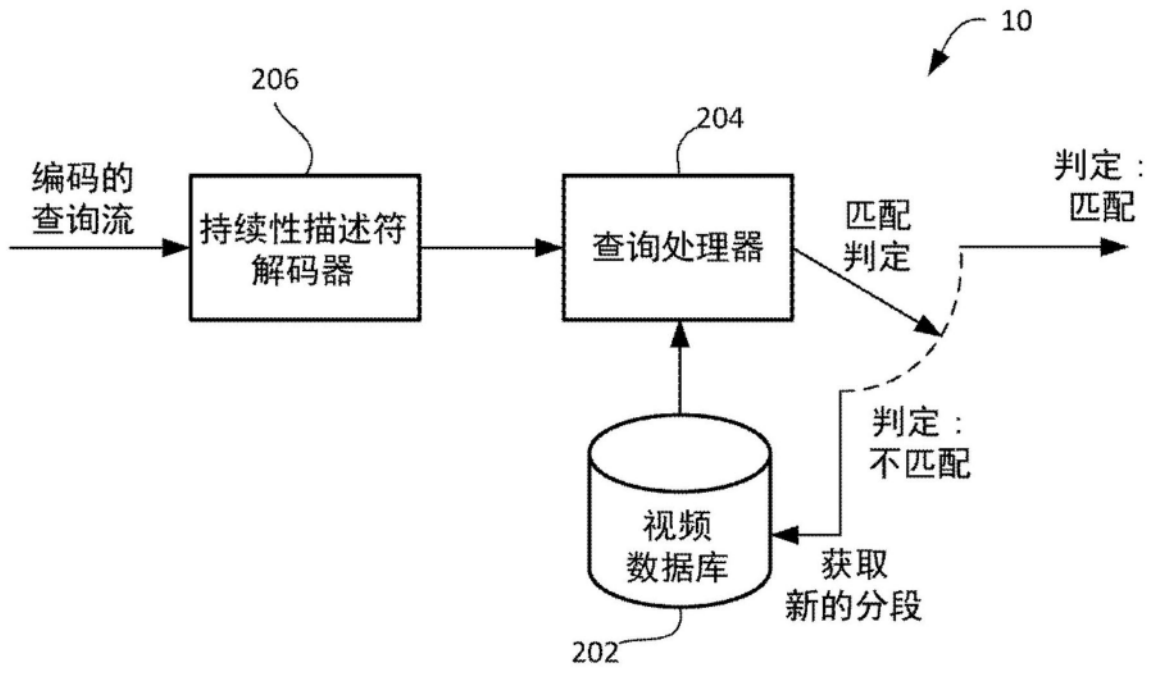


图3

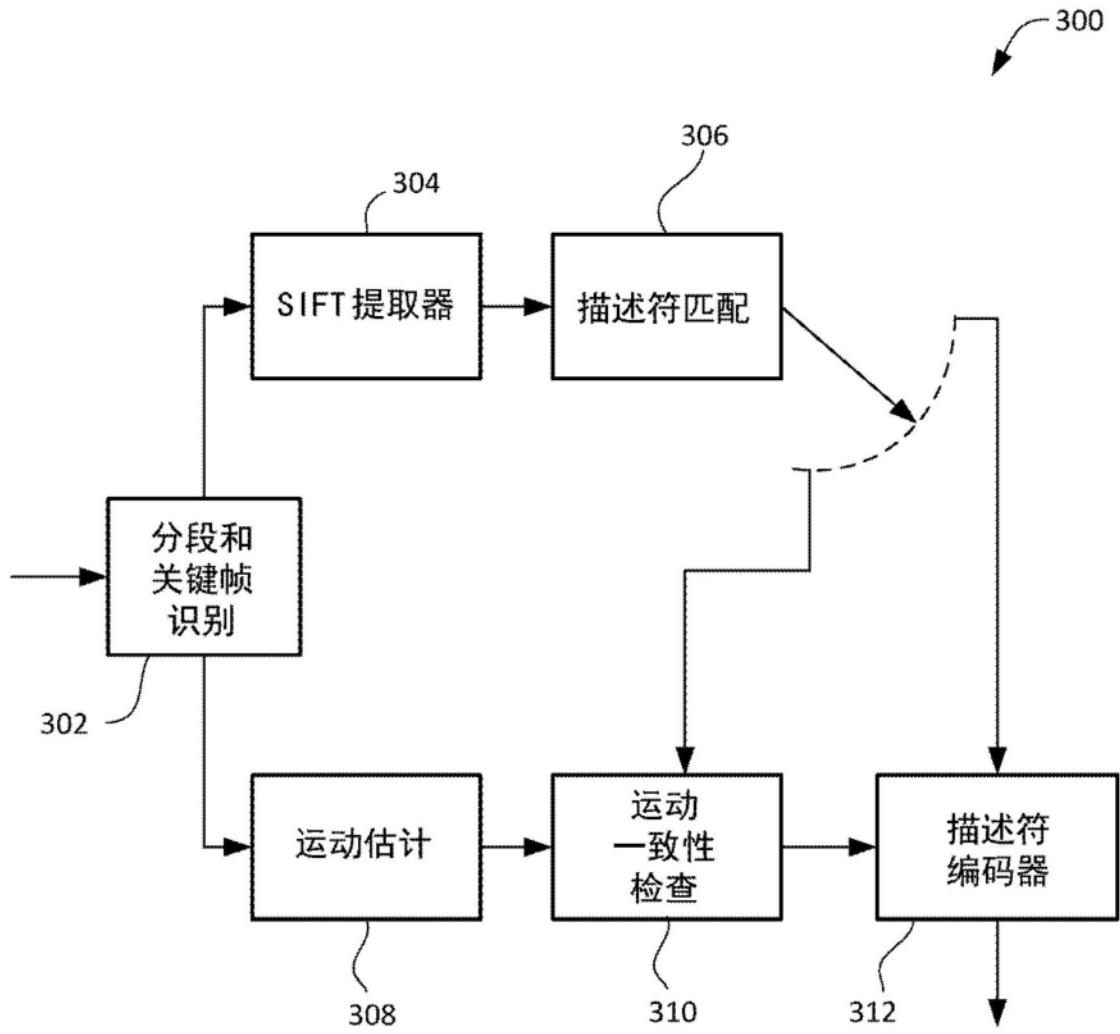


图4

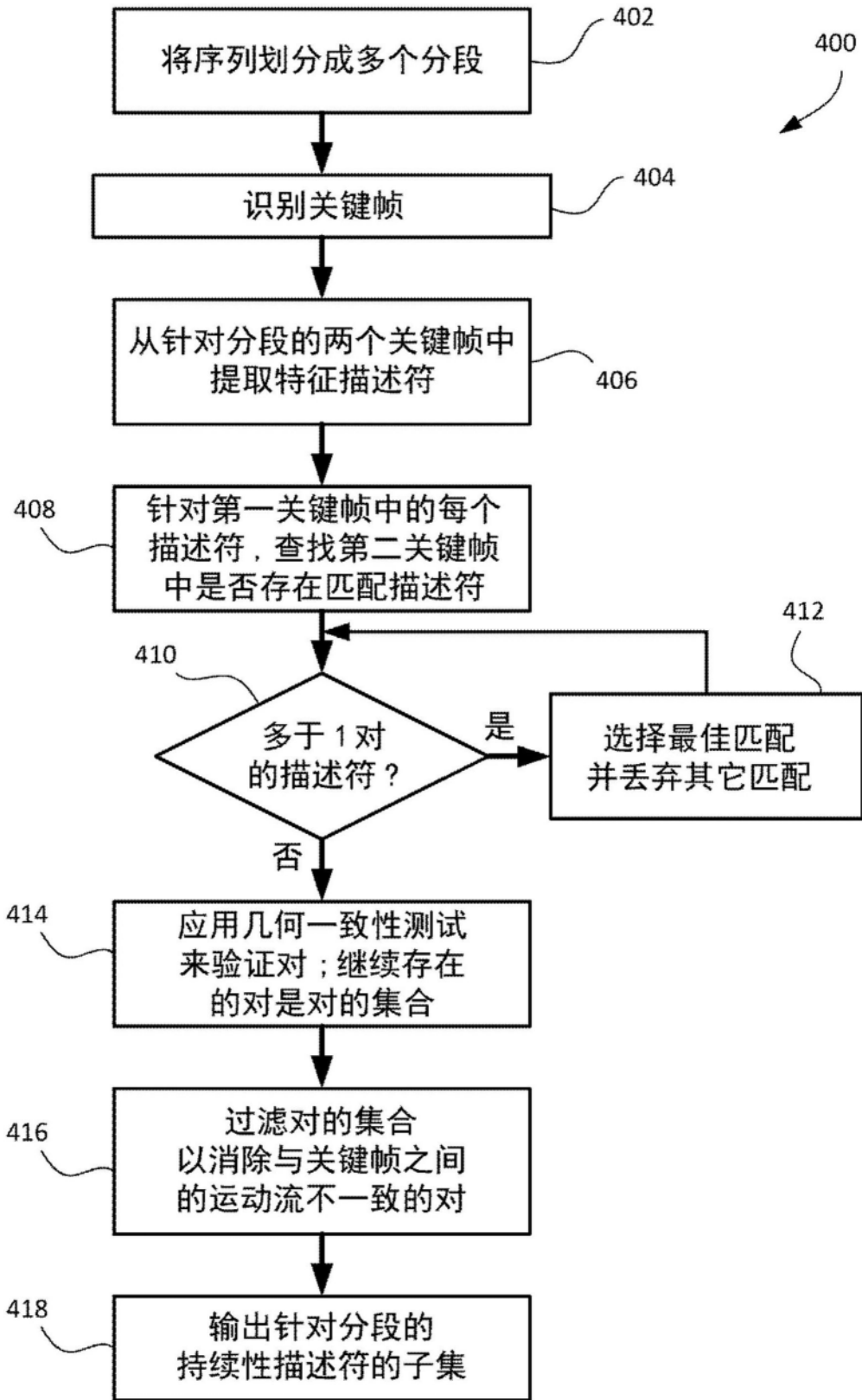


图5

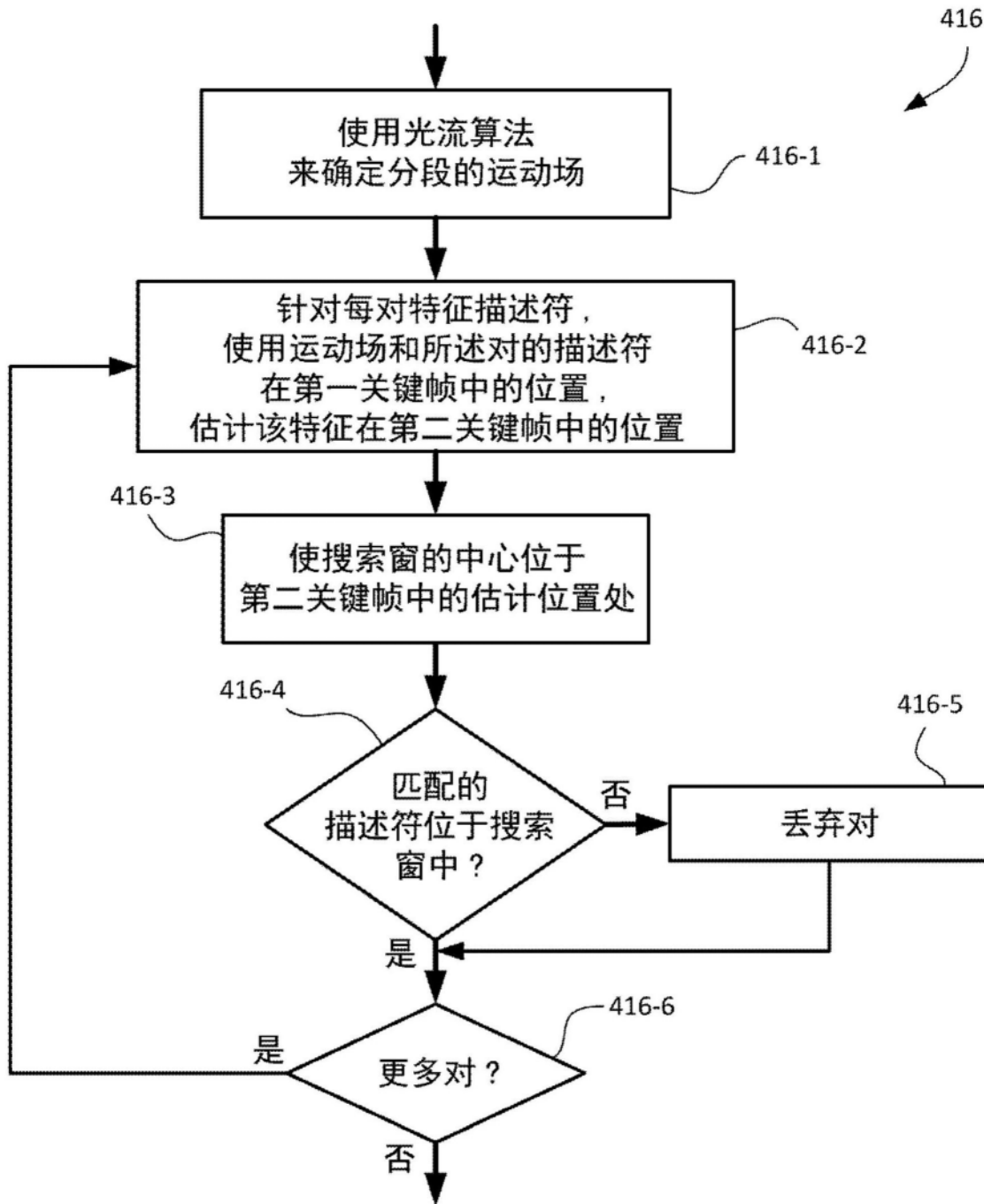


图6

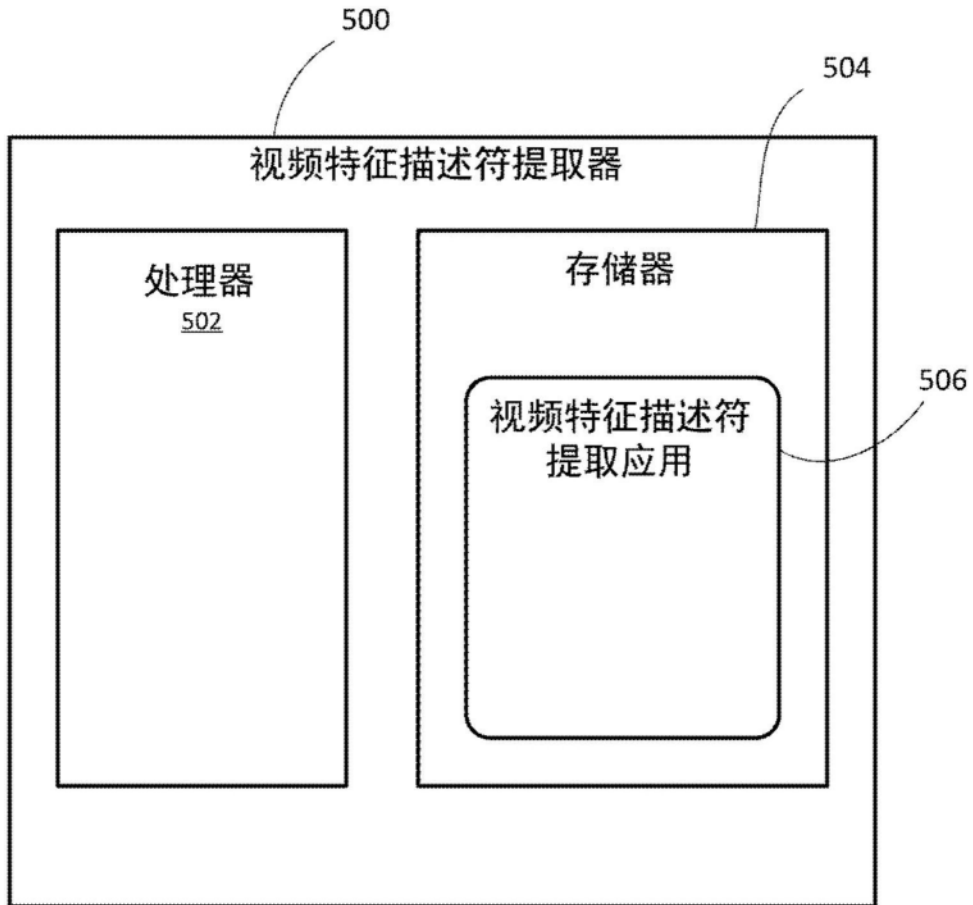


图7

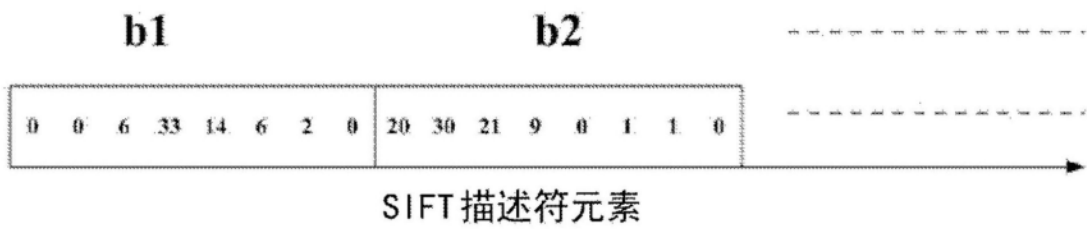
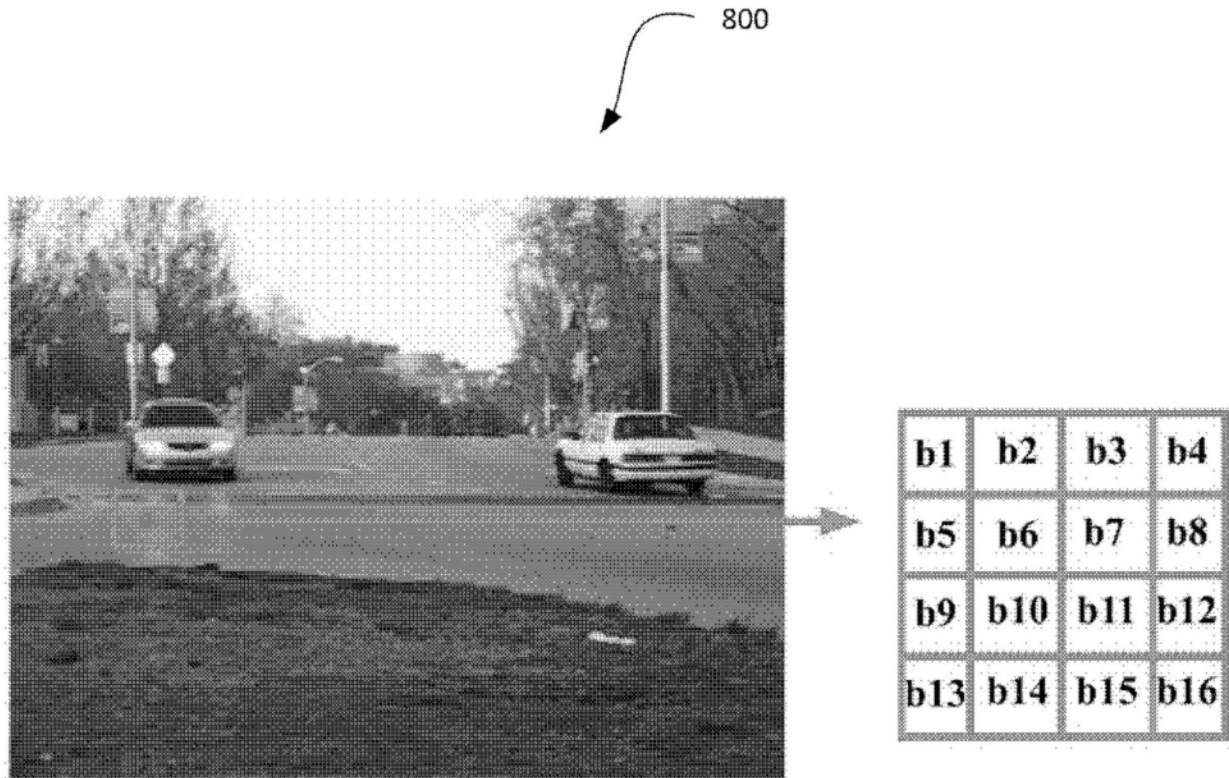


图8

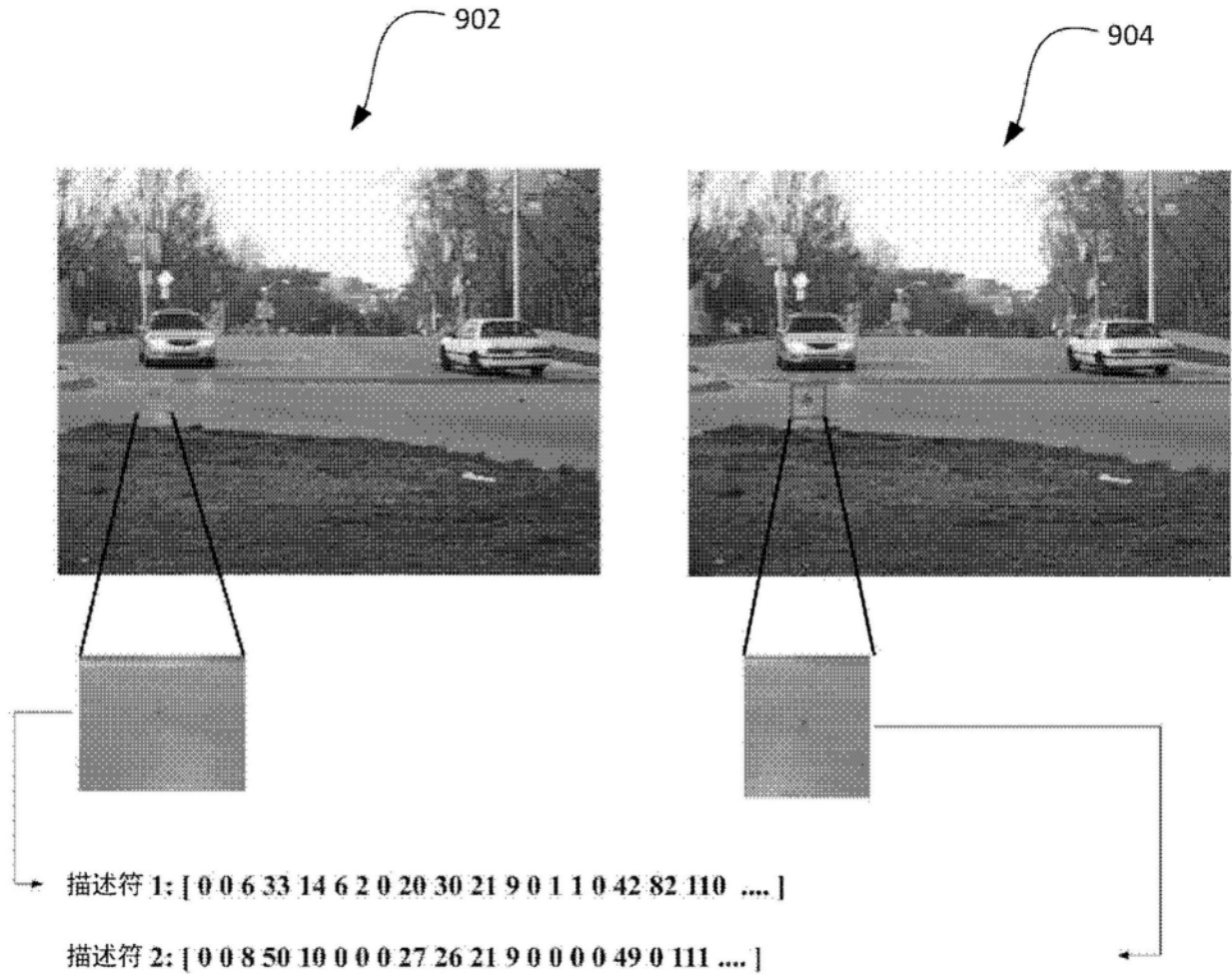


图9