

(12) 发明专利

(10) 授权公告号 CN 101366027 B

(45) 授权公告日 2013. 03. 20

(21) 申请号 200680048754. 8

代理人 王漪 郑霞

(22) 申请日 2006. 11. 15

(51) Int. Cl.

(30) 优先权数据

G06F 17/30 (2006. 01)

60/736, 313 2005. 11. 15 US

60/759, 044 2006. 01. 17 US

(56) 对比文件

WO 0178050 A2, 2001. 10. 18,

CN 1444398 A, 2003. 09. 24,

CN 1459198 A, 2003. 11. 26,

(85) PCT申请进入国家阶段日

2008. 06. 23

(86) PCT申请的申请数据

PCT/IL2006/001320 2006. 11. 15

审查员 田志刚

(87) PCT申请的公布数据

W02007/057893 EN 2007. 05. 24

(73) 专利权人 耶路撒冷希伯来大学伊森姆研究
发展公司

地址 以色列耶路撒冷

(72) 发明人 施缪尔·潘莱格

亚历山大·莱弗-阿佳

(74) 专利代理机构 北京安信方达知识产权代理
有限公司 11262

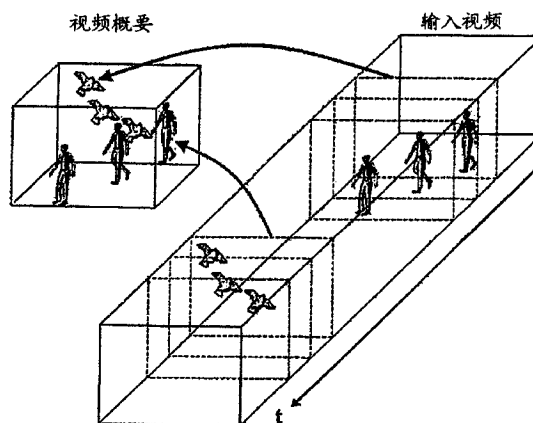
权利要求书 2 页 说明书 15 页 附图 8 页

(54) 发明名称

用于产生视频概要的方法和系统

(57) 摘要

一种计算机实现的方法和系统将第一动态场景的视频帧的第一序列转换到描述第二动态场景的至少两个视频帧的第二序列。在第一序列中获得显示具有位于相应的 x、y 坐标的多个像素的至少一个物体的运动的视频帧的子集, 以及从子集中选择在第一动态场景中显示至少一个物体的非空间重叠外观的部分。所述部分从至少 3 个不同的输入帧复制到第二序列的至少两个连续的帧而不改变像素在物体中的相应的 x、y 坐标, 并且使得第二序列的至少一个帧包括出现在第一序列中的不同帧处的至少两个部分。



1. 一种用于通过将视频摄像机拍摄的第一动态场景的视频帧的源视频序列转换为描述第二动态场景的视频帧的较短概要序列来产生概要视频的方法,所述方法包括:

获得所述源视频序列中的视频帧的子集,所述视频帧的子集显示至少两个物体的运动;

所述方法的特征在于:

从所述子集选择在所述第一动态场景中显示所述至少两个物体的外观的部分,所述部分是选择的概要物体;

为每个选择的概要物体确定用于开始其在所述概要视频中的显示的相应的显示时间;以及

通过在其相应的预定显示时间显示每个选择的概要物体而不改变所述概要物体在所述第一动态场景中的空间位置而产生具有比所述源视频序列更短的持续时间的概要视频,使得所述概要视频中的至少一个所述帧包括出现在所述源视频序列中的不同帧处的至少两个概要物体,

其中所述选择的概要物体中的至少两个以与所述源视频序列中的它们对应的源物体出现的时间顺序不同的时间顺序被显示在所述概要视频中。

2. 如权利要求 1 所述的方法,其包括将所述概要物体与背景缝合成无缝视频。

3. 如权利要求 1 或 2 所述的方法,其中所述源视频序列由相对于在固定位置的轴旋转的摄像机拍摄,并且所述方法包括使所述部分的至少两个在复制到所述概要视频以前空间地弯曲。

4. 如权利要求 1 或 2 所述的方法,其中所述源视频序列由在固定位置的静态摄像机拍摄。

5. 如权利要求 1 或 2 所述的方法,其中在所述源视频序列中同时出现的两个事件在不同的时间显示在所述概要视频中。

6. 如权利要求 1 或 2 所述的方法,其用于以下各项中的任一项:用于监视的概要视频;增加电影的活动密度;视频索引。

7. 如权利要求 6 所述的方法,其包括对在所述概要视频中的每个物体维持指向所述源视频序列中的对应物体的指针。

8. 如权利要求 1 或 2 所述的方法,其包括通过下列步骤预先对准所述源视频序列,以便给出对准的源视频序列:

(a) 计算在所述源视频序列中的帧之间的图像运动参数;

(b) 使在所述源视频序列中的视频帧弯曲,以便在所述第一动态场景的静止物体在所述对准的源视频序列中是静止的。

9. 一种用于将第一动态场景的视频帧的源视频序列转换到描述第二动态场景的至少两个视频帧的概要视频的系统 (10),所述系统包括:

第一存储器 (12),其用于存储在所述源视频序列中的视频帧的子集,所述视频帧的子集显示至少两个物体的运动;

所述系统的特征在于:

选择单元 (13),其耦合到所述第一存储器 (12),用于从所述子集选择在所述第一动态场景中显示所述至少两个物体的外观的部分,所述部分是选择的概要物体;

帧生成器 (14), 其用于为每个选择的概要物体确定用于开始其在所述概要视频中的显示的相应的显示时间并通过在其相应的预定显示时间显示每个选择的概要物体而不改变所述概要物体在所述第一动态场景中的空间位置而产生所述概要视频, 使得所述概要视频中的至少一个所述帧包括出现在所述源视频序列中的不同帧处的至少两个概要物体, 其中所述概要视频具有比所述源视频序列更短的持续时间,

第二存储器 (15), 其耦合到所述帧生成器, 用于存储所述概要视频的帧, 以及

显示装置 (16), 其耦合到所述第二存储器 (15), 以显示所述第二动态场景,

其中所述选择的概要物体中的至少两个以与所述源视频序列中的它们对应的源物体出现的时间顺序不同的时间顺序被显示在所述概要视频中。

10. 如权利要求 9 所述的系统, 其中所述帧生成器 (14) 包括用于使所述部分的至少两个在复制到所述概要视频之前空间地弯曲的弯曲单元 (17)。

用于产生视频概要的方法和系统

[0001] 相关申请

[0002] 本申请是于 2006 年 11 月 2 日递交的 USSN 10/556,601 (Peleg 等人)“Method and system for spatio-temporal video warping”的部分继续申请并且对应于在 2006 年 5 月 11 日公开的 W02006/048875, 并且进一步要求于 2005 年 11 月 15 日递交的临时申请序列号 60/736,313 和于 2006 年 1 月 17 日递交的 60/759,044 的利益, 所有这些申请的内容在此通过引用被包括。

发明领域

[0003] 本发明一般涉及基于图像和视频的再现 (rendering), 其中新的图像和视频通过组合场景的多个原始图像的部分建立。具体地, 本发明涉及这种为了视频摘要或概要的目的的技术。

[0004] 已有技术

[0005] 以下列出了认为与本发明的背景有关的已有技术引用, 并且其内容在此通过引用被并入。另外的引用在上述的美国临时申请号 60/736,313 和 60/759,044 中被提到, 以及其内容在此通过引用被并入。引用的承认在此不应被推断为意味着这些引用以任何方式与在此公开的本发明的可专利性相关。每个引用由包括在方括号内的数字识别, 并且因此在整个说明书中由包括在方括号内的数字表示已有技术。

[0006] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In SIGGRAPH, pages 204-302, 2004.

[0007] A. Agarwala, K. C. Zheng, C. Pal, M. Agrawala, M. Cohen, B. Curless, D. Salesin, and R. Szeliski. Panoramic video textures. In SIGGRAPH, pages 821-827, 2005.

[0008] J. Assa, Y. Caspi, and D. Cohen-Or. Action synopsis: Pose selection and illustration. In SIGGRAPH, pages 667-676, 2005.

[0009] O. Boiman and M. Irani. Detecting irregularities in images and in video. In ICCV, pages 1:462-469, Beijing, 2005.

[0010] A. M. Ferman and A. M. Tekalp. Multiscale content extraction and representation for video indexing. Proc. Of SPIE, 3229:23-31, 1997.

[0011] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu. Efficient representations of video sequences and their applications. Signal Processing: Image Communication, 8(4):327-351, 1996.

[0012] C. Kim and J. Hwang. An integrated scheme for object-based video abstraction. In ACM Multimedia, pages 303-311, New York, 2000.

[0013] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. Science, 4598(13):671-680, 1983.

[0014] V. Kolmogorov and R. Zabih. What energy functions can be minimized via

graph cuts? In ECCV, pages 65–81, 2002.

[0015] Y. Li, T. Zhang, and D. Tretter. An overview of video abstraction techniques. Technical Report HPL-2001-191, HP Laboratory, 2001.

[0016] J. Oh, Q. Wen, J. Lee, and S. Hwang. Video abstraction. In S. Deb, editor, Video Data Management and Information Retrieval, pages 321–346. Idea Group Inc. and IRM Press, 2004.

[0017] C. Pal and N. Jojic. Interactive montages of sprites for indexing and summarizing security video. In Video Proceedings of CVPR05, page II:1192, 2005.

[0018] A. Pope, R. Kumar, H. Sawhney, and C. Wan. Video abstraction: Summarizing video content for retrieval and visualization. In signals, Systems and Computers, pages 915–919, 1998.

[0019] W02006/048875 Method and system for spatio-temporal video warping, pub. May 11, 2006 by S. Peleg, A. Rav-Acha and D. Lischinski. 其对应于 2005 年 11 月 2 日递交的 USSN 10/556, 601.

[0020] A. M. Smith and T. Kanade. Video skimming and characterization through the combination of image and language understanding. In CAIVD, pages 61–70, 1998.

[0021] A. Stefanidis, P. Partsinevelos, P. Agouris, and P. Doucette. Summarizing video datasets in the spatiotemporal domain. In DEXA Workshop, pages 906–912, 2000.

[0022] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In CVPR, pages 819–826, 2004.

[0023] X. Zhu, X. Wu, J. Fan, A. K. Elmagarmid, and W. G. Aref. Exploring video content structure for hierarchical summarization. Multimedia Syst., 10(2):98–115, 2004.

[0024] J. Barron, D. Fleet, S. Beauchemin and T. Burkitt. Performance of optical flow techniques. volume 92, pages 236–242.

[0025] V. Kwatra, A. **Schödl**, I. Essa, G. Turk and A. Bobick. Graph cut textures; image and video synthesis using graph cuts. In Siggraph, pages 227–286, July 2003.

[0026] C. Kim and J. Hwang, Fast and Automatic Video Object Segmentation and Tracking for Content-Based Applications, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, No. 2, February 2002, pp 122–129.

[0027] 美国专利 6,655,003.

[0028] 发明背景

[0029] 视频概要 (synopsis) (或摘要 (abstraction)) 是旨在实现视频浏览和检索的时间紧凑的表示。

[0030] 有两种用于视频概要的主要方法。在一种方法中, 一组显著的图像 (关键帧) 从原始视频序列中选择。所选择的关键帧是最好地表示视频的帧 [7, 18]。在另一种方法中, 选择了短视频序列的集合 [15]。第二种方法较不紧凑, 但是提供场景动态的更好的效果。

在关于视频摘要的综合调查中描述了那些方法（以及其它的方法）[10,11]。

[0031] 在上面的两种方法中，全部帧被用作基本构建块。一个不同的方法共同使用镶嵌图像以及用于视频索引的一些元数据（meta-data）[6,13,12]。在这种方法中，静态概要图像包括来自不同时间的物体。

[0032] 基于物体的方法也是已知的，其中物体从输入视频中提取 [7,5,16]。然而，这些方法使用用于识别显著的关键帧的物体检测，并且没有合并来自不同时间间隔的活动。

[0033] 使用叠加的最小切割（min-cut）来产生单个全景图像 [1] 以及使用叠加的最小切割来产生全景电影 [2] 的方法在本领域中也是已知的。在这两种方法中，具有指数复杂性（在输入帧的数量方面）的问题是近似的，并且因此它们更适合于少量的帧。本领域中的有关工作与使用最小切割来组合两段电影有关 [20]。

[0034] WO2006/048875 [14] 公开了一种用于在视频中操纵时间流的方法和系统。第一动态场景的视频帧的第一序列被转换为描述第二动态场景的视频帧的第二序列，以便在一方面，对于第一动态场景中的至少一个特征，视频帧的第一序列的相应部分以不同于视频帧的第一序列的周围部分的速率采样；并且所采样的部分被复制到第二序列的对应帧中。这允许改变动态场景中特征的时间同步。

发明内容

[0035] 根据本发明的第一个方面，提供了一种计算机实现的方法，所述方法 用于将第一动态场景的视频帧的第一序列转换为描述第二动态场景的至少两个视频帧的第二序列，所述方法包括：

[0036] (a) 获得所述第一序列中的视频帧的子集，所述视频帧的子集显示包括位于相应的 x、y 坐标的多个像素的至少一个物体的运动；

[0037] (b) 从所述子集选择在所述第一动态场景中显示至少一个物体的非空间重叠外观的部分；以及

[0038] (c) 将来自至少 3 个不同的输入帧的部分复制到所述第二序列的至少两个连续帧而不改变像素在所述物体中相应的 x、y 坐标，并且使得所述第二序列的至少一个帧包括出现在第一序列中的不同帧处的至少两个部分。

[0039] 根据本发明的第二个方面，提供了一种用于将第一动态场景的视频帧的第一序列转换到描述第二动态场景的至少 2 个视频帧的第二序列的系统，所述系统包括：

[0040] 第一存储器，其用于存储所述第一序列中的视频帧的子集，所述视频帧的子集显示包括位于相应的 x、y 坐标的多个像素的至少一个物体的运动。

[0041] 选择单元，其耦合到所述第一存储器，用于从所述子集中选择显示所述第一动态场景中至少一个物体的非空间重叠外观的部分，

[0042] 帧生成器，其用于将来自至少 3 个不同的输入帧的所述部分复制到所述第二序列的至少两个连续帧而不改变所述像素在所述物体中相应的 x、y 坐标，并且使得所述第二序列的至少一个所述帧包括出现在所述第一序列中的不同帧的至少两个部分，以及

[0043] 第二存储器，其用于存储所述第二序列的帧。

[0044] 根据第三个方面，本发明还包括可触知地体现描述动态场景的输出视频帧的序列的数据载体，所述输出视频帧的至少两个连续帧包括多个像素，所述多个像素具有相应的

x、y 坐标并且来源于来自至少 3 个不同的输入帧的物体的部分而不改变所述像素在所述物体中相应的 x、y 坐标，并且使得所述输出视频帧的至少一个包括出现在不同的输入帧处的至少两个部分。

[0045] 由本发明公开的动态视频概要在以下两个特性方面不同于上面回顾的以前的视频摘要方法：(i) 视频概要本身是表示场景的动态的视频。(ii) 为了尽可能多地减少时空冗余，在活动之间的相对定时可以改变。

[0046] 作为例子，考虑在图 1 被表示为时空体积的示意性视频片断。视频从在地面上行走的人开始，并且一段不活动的时间以后，鸟正在天上飞。不活动的帧在大多数视频摘要方法中被省略。视频概要通过同时播放人和鸟，是实质上更紧凑的。当在空间位置没有其它活动发生时，这通过将事件从其原始时间间隔转移到另一个时间间隔而产生图像区域的最佳使用。这种处理缓和了最初出现在 [14] 中的事件的时间顺序连贯性。

[0047] 本发明也介绍了一种使用对马尔可夫随机场 (Markov Random Field) [9] 的最优化来产生视频概要的低层次方法。

[0048] 本发明提供的选择之一是显示单个物体的多个动态外观的能力。这种效果是用在移动物体的传统视频概要 [6, 1] 中的“频闪 (stroboscopic)”图片的概括。介绍了用于完成此的两个不同的方案。在第一个方案中，物体在不同时刻的快照呈现在输出视频中，以便提供在从开始位置到结束位置的整个视频中物体进程的指示。在第二个方案中，物体没有限定的起始或结束位置而是随机地和不可预知地运动。在这种情况下，物体在不同时刻的快照再次呈现在输出视频中，但是这次给出了比实际存在的物体更多的增加的物体的印象。两种方案的共同点在于，在不同时刻从输出视频拍摄的多个快照以避免空间重叠的方式被复制到输出视频，且没有来自无助于感兴趣的物体的动态发展的输入视频数据的复制。

[0049] 在本发明的上下文和所附权利要求中，术语“视频”在其最广泛的术语方面是与“电影”同义的，仅仅假定它作为适合于后处理的计算机图像文件是可访问的，并且包括任何种类的电影文件，例如数字的、模拟的。摄像机优选地在固定的位置，这意味着它可旋转和缩放 - 但不经历平移运动，如在迄今为止提议的技术中进行的。与例如在美国专利 6, 665, 003 [22] 以及贯注于显示立体图像的其他引用中处理的静态场景相对比，与本发明有关的场景是动态的，上述立体图像的显示没有描绘动态场景，在动态场景中连续帧具有空间和时间的连续性。根据本发明的一个方面，我们将这种问题用公式表示为单个最小切割问题，其可以通过在图上找到最大流而在多项式时间中解决 [5]。

[0050] 为了描述本发明，将使用由我们称为“时空体积 (space-time volume)”的结构以建立动态全景视频。时空体积可通过沿着时间轴连续堆叠所有的帧由图像的输入序列构成。然而，应理解，只要涉及到实际的实现，就实际上没有必要例如通过在时间上实际堆叠动态源场景的 2D 帧来构造时空体积。更一般地，源帧是被单独地处理以构造目标帧，但是它将有助于涉及到时空体积的理解，好像它是物理构造而不是概念上的构造。

附图说明

[0051] 为了理解本发明并观察它在实践中如何实现，现在参考附图仅作为非限制性的例子描述优选的实施方式，其中：

[0052] 图 1 是图形表示,其显示了本发明的通过同时播放时间上转移的特征来产生紧凑的视频概要的方法;

[0053] 图 2a 和 2b 是示意性表示,其描述了根据本发明产生的视频概要;

[0054] 图 3a、3b 和 3c 是图形表示,其显示了根据本发明的时间重排的实施例;

[0055] 图 4 是图形表示,其显示了使用图 3b 中描述的动态频闪效应的视频概要的单个帧;

[0056] 图 5a、5b 和 5c 是图形表示,其显示了当短的概要可描述较长的序列且没有活动损失亦没有频闪效应时的实施例。

[0057] 图 6 是图形表示,其显示了根据本发明的全景视频概要的进一步实施例;

[0058] 图 7a、7b 和 7c 是图形表示,其显示了来自街道监视的视频概要的细节;

[0059] 图 8a 和 8b 是图形表示,其显示了来自围墙监视的视频概要的细节;

[0060] 图 9 是图形表示,其显示了根据本发明的进一步的实施方式的电影的增加了的活动密度;

[0061] 图 10 是用来产生图 9 所示的电影的过程的示意图;

[0062] 图 11 是方块图,其显示了根据本发明的系统的主要功能;以及

[0063] 图 12 是流程图,其显示了根据本发明执行的基本操作。

[0064] 具体实施方式

[0065] 1. 活动检测

[0066] 本发明假设每个输入像素以其“重要”程度被标注。虽然从现在起我们将使用“重要”程度作为活动程度,但是很明显,根据所需要的应用对“重要性”可使用任何其它测量标准。重要性(或活动)程度的评估是假设的并且本身不是本发明的特征。它可以使用各种用于检测不规则性[4, 17]、运动物体检测以及物体追踪的方法之一实现。可选地,它可基于识别算法,例如脸部检测。

[0067] 作为例子,可以选择简单的和通常使用的活动指示器,其中如果输入像素 $I(x, y, t)$ 时间中值在位置 (x, y) 的颜色差大于给定的阈值,那么输入像素 $I(x, y, t)$ 被标为“活动的”。活动的像素由特征函数定义:

$$[0068] \quad \chi(p) = \begin{cases} 1 & \text{if } p \text{ is active} \\ 0 & \text{otherwise,} \end{cases}$$

[0069] 为了从活动指示器中清除噪声,中值过滤器在继续概要处理前应用于 x 。

[0070] 尽管使用连续的活动测量是可能的,但是发明人专心于二进制情况。可以使用以下详细描述中的几乎所有等式以仅仅很小的变化[4, 17, 1]来使用连续的活动测量。

[0071] 我们描述了两种用于视频概要的计算的不同实施方式。一种方法(部分 2)使用图形表示以及使用图形切割的代价函数(cost function)的最优化。另一种方法(部分 3)使用物体分割和追踪。

[0072] 2. 通过能量最小化的视频概要

[0073] 使输入视频序列的 N 个帧被表示在 3D 时空体积 $I(x, y, t)$ 中,其中 (x, y) 是该像素的空间坐标,且 $1 \leq t \leq N$ 是帧数。

[0074] 我们希望产生具有以下性质的视频概要 $S(x, y, t)$:

[0075] ● 视频概要 S 应实质上短于原始视频 I 。

[0076] ●来自原始视频的最大量的“活动”应出现在视频概要中。

[0077] ●在视频概要中的物体运动应与其在原始视频中的运动相似。

[0078] ●视频概要应看起来良好,并且应避免明显的接缝或破碎的物体。

[0079] 具有以上性质的视频概要 S 用映射 M 生成,给概要 S 中的每个坐标 (x, y, t) 分配来自 I 的源像素的坐标。我们集中于像素的时间偏移,保持空间位置固定。因此,任何概要像素 $S(x, y, t)$ 可来自于输入像素 $I(x, y, M(x, y, t))$ 。时间偏移 M 通过解决能量最小化问题获得,其中代价函数由下式给出:

$$[0080] \quad E(M) = E_a(M) + aE_d(M), \quad (1)$$

[0081] 其中 $E_a(M)$ 表示活动损失,而 $E_d(M)$ 表示接缝两端的不连续性。活动损失是没有出现在视频概要 S 中的输入视频 I 中的活动像素的数量,

$$[0082] \quad E_a(M) = \sum_{(x,y,t) \in I} \chi(x, y, t) - \sum_{(x,y,t) \in S} \chi(x, y, M(x, y, t)) \quad (2)$$

[0083] 不连续的代价 E_d 被定义为在视频概要中的时空邻居和在输入视频中的对应邻居之间的接缝两端的颜色差之和(类似的公式可以在 [1] 中找到):

$$[0084] \quad E_d(M) = \sum_{(x,y,t) \in S} \sum_i \|S((x, y, t) + e_i) - I(x, y, M(x, y, t)) + e_i\|^2 \quad (3)$$

[0085] 其中 e_i 是表示 6 个时空邻居的 6 个单位矢量。

[0086] 图 2a 和 2b 是描述时空操作的示意性图示,该操作通过最小化代价函数来产生短视频概要,其中运动物体的运动由图中的“活动条”描述。上部分表示原始视频,而下部分表示视频概要。具体地,在图 2a 中,较短视频概要 S 通过包括大多数活动像素由输入视频 I 产生。为了确保平滑,当在 S 中的像素 A 对应于 I 中的像素 B 时,其“交叉边界(cross border)”邻居应该是类似的。找到最佳的 M 极小化 (3) 是非常大的最优化问题。在图 2b 中显示了近似的解决方案,其中在视频概要中的连续像素被限制为来自于连续的输入像素。

[0087] 注意,代价函数 $E(M)$ (等式 1) 对应于 3D 马尔可夫随机场 (MRF),其中每个节点对应于在输出电影的 3D 体积中的像素,并且可被赋以对应于输入帧的任何时间值。在节点上的权由活动代价确定,而在节点之间的边缘根据不连续的代价确定。代价函数因此可由类似于迭代的图像切割 [9] 的算法最小化。

[0088] 2.1. 使用 2D 图形的受限制的解决方案

[0089] 允许在视频概要中的每个像素来自于任何时间的等式 1 的最优化是大规模的问题。例如,被概括为 5 秒钟的视频概要的 3 分钟的输入视频导致具有大约 2^{25} 个节点的图像,每个节点具有 5400 个标签。

[0090] 对于在水平路径上运动的动态结构或物体的情况显示在 [2] 中,3DMRF 可通过将该问题简化为 1D 问题而被有效地解决。在这项工作中,我们处理以更一般的方式运动的物体,因此我们使用不同的限制。在视频概要 S 中的连续像素被限制为来自于输入视频 I 的连续像素。在这种限制下,3D 图形简化为 2D 图形,其中每个节点对应于在概要电影中的空间位置。每个节点 $M(x, y)$ 的标签确定显示在 S 的第一帧中的 I 中的帧数 t,如图 2b 所示。如果 $M(x_1, y_1) \neq M(x_2, y_2)$,则在 S 中的两个邻近位置 (x_1, y_1) 和 (x_2, y_2) 之间存在接缝,并且沿着接缝的不连续代价 $E_d(M)$ 是对 S 中的所有帧在此空间位置的颜色差的总和。

$$[0091] \quad E_d(M) = \sum_{x,y} \sum_i \sum_{t=1}^K \|\mathcal{S}((x,y,t)+e_i) - I((x,y,M(x,y)+t)+e_i)\|^2 \quad (4)$$

[0092] 其中 e_i 现在是描述 4 个空间邻居的 4 个单位矢量。

[0093] 每个节点的标签的数量是 $N-K$ ，其中 N 和 K 分别是在输入和输出视频中的帧数。每个像素的活动损失是：

$$[0094] \quad E_a(M) = \sum_{x,y} \left(\sum_{t=1}^N \chi(x,y,t) - \sum_{t=1}^K \chi(x,y,M(x,y)+t) \right)$$

[0095] 3. 基于物体的概要

[0096] 如较早描述的用于动态视频概要的低层次方法限于满足局部性质，例如避免明显的接缝。当物体可被检测到时，较高层次的基于物体的性质可被合并。例如，避免频闪效应需要在体积中检测和追踪每个物体。这部分描述了一种用于动态视频概要的基于物体的方法的实现。在文献（例如 [7, 5, 16]）中存在几种基于物体的视频摘要方法，并且这些方法全部使用检测的物体来选择显著帧 (significant frame)。与这些方法不同，为了更好地利用空间和时间，本发明在时间上转移物体，并且建立从来没有在输入序列中出现的新概要帧。

[0097] 在一个实施方式中，如上所述，通过比较每个像素与时间中值并且确定该差值的阈值来检测运动物体。这跟随有使用空间中值过滤器的噪声清除，以及跟随有将时空连接的部分分组在一起。应认识到，在文献中有许多可用在该任务中的物体检测和追踪的其他方法（例如，[7, 17, 21]）。物体检测和追踪的每个过程产生一组物体，其中每个物体 b 由其特征函数表示：

$$[0098] \quad \chi_b(x,y,t) = \begin{cases} 1 & \text{if } (x,y,t) \in b \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

[0099] 图 3a、3b 和 3c 是图形表示，其显示了根据本发明的时间重排的实施例。每个图的上部分表示原始视频，而下部分表示视频概要，其中运动物体的运动由图中的“活动条”表示。图 3a 显示了在不同时间记录、被转移到视频概要中的同一时间间隔的两个物体。图 3b 显示了在一段长的时间期间运动的单个物体分成具有较短时间间隔的片断，片断然后同时被播放，产生动态频闪效应。图 3c 显示了当物体体积分成片断时，物体的交叉不干扰概要。

[0100] 从每个物体中，通过选择物体出现的帧的子集而产生片断。这种片断可表示可选地在不同采样率下获得的时间间隔。

[0101] 视频概要 S 采用以下操作根据输入视频 I 构造：

[0102] (1) 从输入视频 I 中提取物体 $b_1 \dots b_r$ 。

[0103] (2) 从原始物体中选择一组非重叠片断 B 。

[0104] (3) 将时间偏移 M 应用到每个选定的片断，产生更短的视频概要，同时避免在物体之间的闭塞 (occlusion) 并实现无缝的缝合。这在图 1 和图 3a 至 3c 中被解释。图 4 是图形表示，其显示了使用如图 3b 所描述的动态频闪效应的视频概要的单个帧的实施例。

[0105] 上面的操作 (2) 和 (3) 是相互关联的，因为我们希望选择片断并且在时间上移动它们来获得短而无缝的视频概要。应认识到，上面的操作 (2) 和 (3) 不需要是完美的。当我们说“非重叠的片断”时，可允许小的重叠，而当我们说“避免闭塞”时，在时间偏移的物体之间的小的重叠可被允许但应被最小化，以便得到具有视觉吸引力的视频。

[0106] 在基于物体的表示中，在作为结果的概要中的像素可具有多个源（来自不同的物

体),并且因此我们增加后处理步骤,其中所有的物体被缝合在一起。背景图像通过获取序列的所有帧的像素中值而产生。所选定的物体然后可使用与在每个帧和中间图像的像素值之间的距离(在 RGB 空间中)成比例的权而被调和。该缝合机制类似于在 [6] 中使用的机制。

[0107] 我们将映射到单个概要像素 $(x, y, t) \in S$ 的所有像素的集合定义为 $sre(x, y, t)$, 并且我们将在物体(或片断) b 中的(活动)像素的数量表示为 $\#b = \sum_{x, y, t \in I} x_b(x, y, t)$ 。

[0108] 我们然后定义测量片断 B 的子集选择和时间偏移 M 的代价的能量函数。代价包括活动损失 E_a 、处罚长视频概要的对物体 E_o 和项 E_1 之间的闭塞的处罚 (penalty):

$$[0109] \quad E(M, B) = E_a + \alpha E_o + \beta E_1 \quad (6)$$

[0110] 其中

$$[0111] \quad E_a = \sum_b \#b - \sum_{b \in B} \#b \quad (7)$$

$$[0112] \quad E_o = \sum_{(x, y, t) \in S} Var\{src(x, y, t)\}$$

$$[0113] \quad E_1 = length(S)$$

[0114] 3. 1. 具有预定长度的视频概要

[0115] 我们现在描述由较长视频构造预定长度 K 的短视频概要的情况。在这个方案中,每个物体被分割成长度 K 的重叠和连续的片断。所有的片断被时间偏移以在 $t = 1$ 时刻开始,并且让我们决定哪个片断包括在视频概要中。很明显,在这个方案中,一些物体可不出现在视频概要中。

[0116] 我们首先定义在所有片断对之间的闭塞代价。使 b_i 和 b_j 是具有出现时间 t_i 和 t_j 的两个片断,并且使每个片断的支持由其特征函数 x (如在等式 5 中)表示。

[0117] 在这两个片断之间的代价定义为在被偏移到时刻 $t = 1$ 之后的两个片断之间的颜色差的总和。

$$[0118] \quad v(b_i, b_j) = \sum_{(x, y, t) \in S} (I(x, y, t + t_i) - I(x, y, t + t_j))^2 \cdot \chi_{b_i}(x, t, t + t_i) \cdot \chi_{b_j}(x, t, t + t_j) \quad (8)$$

[0120] 对于视频概要,我们选择最小化等式 6 中的代价的片断 B 的部分集合,其中 E_1 现在是常数 K ,并且闭塞代价由下式给出:

$$[0121] \quad E_o(B) = \sum_{i, j \in B} v(b_i, b_j) \quad (9)$$

[0122] 为了避免显示同一时空像素两次(这是可接受的但是浪费的),我们对在原始电影中交叉的片断 b_i 和 b_j 设定 $v(b_i, b_j) = \infty$ 。另外,如果频闪效应是不期望有的,它可通过对从同一个物体采样的所有 b_i 和 b_j 设定 $v(b_i, b_j) = \infty$ 而被避免。

[0123] 模拟退火 (simulated annealing) [8] 用来最小化能量函数。每个状态描述了包括在概要中的片断的子集,并且邻近的状态设置成片断被去除、增加或用其他片断代替的集合。

[0124] 在片断选择后,长度 K 的概要电影通过将所有偏移的片断粘贴到一起而构建。在图 4 中给出了使用这种方法的来自视频概要的一个帧的实施例。

[0125] 3.2. 无损视频概要

[0126] 对于一些应用,例如视频监控,我们可能更喜欢较长的视频概要,但其中保证所有的活动出现。在这种情况下,目的不是如在前面的部分中实现的那样选择一组物体片断,而更确切地是找到物体片断的紧凑的时间重排。

[0127] 再次,我们采用模拟退火来最小化能量。在这种情况下,状态对应于所有片断的一组时间偏移,并且两个状态被定义为邻居,如果其时间偏移仅仅对单个片断不同。有两个在这种情况下应该注意的问题:

[0128] ●出现在第一帧或最后一帧中的物体片断应该在视频概要中如此保留;(否则它们可能突然出现或消失)。我们注意每个状态将通过相应地固定所有这些物体的时间偏移来满足这个限制。

[0129] ●输入视频的时间排列一般是能量函数的局部最小化,并且因此不是对开始退火处理的优先选择。我们用更短的视频开始我们的模拟退火,其中所有的物体重叠。

[0130] 图 5a、5b 和 5c 是图形表示,其显示了当短概要可描述较长的序列而没有活动损失以及没有频闪效应时这种方法的实施例。3 个物体可被时间偏移以同时播放。具体地,图 5a 描述了原始视频(顶部)和视频概要(底部)的示意性时空图示。图 5b 描述了来自原始视频的 3 个帧;如从图 5a 中看到的,在原始视频中,每个人单独地出现,但在视频概要中,所有 3 个人可一起出现。图 5c 描述了来自同时显示所有 3 个人的视频概要的一个帧。

[0131] 4. 全景视频概要

[0132] 当视频摄像机扫描场景时,通过使用全景镶嵌可除去很多冗余。然而,现有方法构建了单个全景图像,其中失去了场景动态。有限的动态可由频闪图像 [6, 1, 3] 表示,其中运动物体显示在沿着其路径的几个位置。

[0133] 全景视频概要可通过同时显示在场景的不同时间不同区域发生的行为而产生。可得到相当大的浓缩,因为对每个物体的活动的持续时间被限制在它被摄像机观察的时间。特别的情况是当摄像机追踪例如图 6 所示的奔跑的雌狮等的物体时。当摄像机追踪奔跑的雌狮时,视频概要是背景的全景镶嵌,并且前景包括奔跑的雌狮的几个动态复制。在这种情况下,仅仅通过允许频闪效应可得到短的视频概要。

[0134] 构建全景视频概要以类似于普通视频概要的方式完成,具有使所有的帧对准一些参考帧的初步阶段。在对准之后,物体的图像坐标从全局坐标系统中取得,所述全局坐标系统可以是输入图像之一的坐标系统。

[0135] 为了能够处理视频,即使在运动物体的分割不完美时,我们处罚了闭塞而不是完全阻止它们。该闭塞处罚实现了物体的时间布置中的灵活性,即使在分割不完美时,并且物体的像素可包括一些背景。

[0136] 可以增加另外的项,其将视频概要的时间排序朝着输入视频的排序偏置。

[0137] 由于大量的可能性,对所有可能的片断选集 B 以及时间偏移 M 最小化上面的能量是非常消耗的。然而,该问题可通过限制解决方案而显著地按比例减小。两个受限的方案在以下部分中描述。

[0138] 5. 监视实施例

[0139] 对于视频概要的有趣的应用可以是获取存储的监视视频。当在视频中检查某些事件变得必要时,其可使用视频概要非常快地完成。

[0140] 如上所述,图 5 显示了视频概要在将所有活动浓缩至短时间段而不损失任何活动方面的能力的实施例。这使用从监控咖啡站的摄像机中收集的视频来完成。从真实的监视摄像机给出了两个另外的实施例。图 7a、7b 和 7c 是图形表示,其显示了来自街道监视的视频概要的细节。图 7a 显示了来自原始视频 (22 秒) 的典型帧。图 7b 描述了来自显示浓缩活动的视频概要电影 (2 秒) 的帧。图 7c 描述了来自显示甚至更浓缩活动的更短的视频概要 (0.7 秒) 的帧。在这些图中显示的图像来源于由观察城市街道的摄像机拍摄的视频,行人偶尔越过视场。他们中的许多可以被收集到非常浓缩的概要中。

[0141] 图 8a 和 8b 是图形表示,其显示了来自围墙监视的视频概要的细节。在围墙附近有非常少的活动,并且有时我们可以看到朝着栅栏爬行的士兵。视频概要显示了同时爬行和步行的士兵的所有情况,或可选地通过频闪地播放视频而使得视频概要更短。

[0142] 6. 通过视频概要的视频索引

[0143] 视频概要可用于视频索引,为使用者提供有效和直观的连接以访问视频中的行为。这可通过将指针与每个概要像素相关联而实现,指针指向原始视频中对应物体的出现。在视频概要中,视频的信息被投影到“活动的空间”中,其中只有活动重要,不管其时间情境(尽管我们仍然保留空间情境)。因为活动被集中在短的时间段内,在视频中的具体活动可以被容易地获取。

[0144] 从前面的描述中应该清楚,当视频摄像机扫描动态场景时,在输入视频中区域变得可见的绝对“时序时间”不是场景动态的一部分。在每个区域的可见时间段期间的“局部时间”对于描述场景中的动态更相关,并在构建动态镶嵌时应被保留。上述的实施方式呈现了本发明的第一个方面。根据第二个方面,我们现在将显示如何建立无缝全景镶嵌,其中在图像之间的缝合尽可能多地避免了来自场景中物体的切断部分,即使当这些物体可能运动时。

[0145] 7. 利用 3D 最小切割建立全景图像

[0146] 使 I_1, \dots, I_N 是输入序列的帧。我们假设使用一种现有的方法使序列对准单个参考帧。为了简单,我们假设所有帧在对准之后具有相同尺寸(在摄像机的视场之外的像素将被标记为无效的)。还假设摄像机是顺时针摇摄的。(不同的运动可以按类似的方式处理)。

[0147] 使 $P(x, y)$ 是构建的全景图像。对于在 P 中的每个像素 (x, y) ,我们需要选择帧 $M(x, y)$,该像素从其取得。(也就是说,如果 $M(x, y) = k$,那么 $P(x, y) = I_k(x, y)$ 。)明显地,在摄像机是顺时针摇摄的假设下,左列必须从第一帧取得,而右列必须从最后一帧取得。(可选择其它的边界条件来产生具有更小视场的全景图像)。

[0148] 我们的目的是产生无缝的全景图像。为了这么做,我们将试图避免在物体内的缝合,特别是它们在运动时。我们使用接缝分数(seam score),类似于由 [1] 使用的分数,而不是(近似地)解决 NP-hard 问题,我们将找到对限制更多的问题的最佳解决方案。

[0149] 8. 将问题用公式表示为能量最小化问题

[0150] 与以前公式的主要不同是我们的缝合代价,由以下公式定义:

[0151]

$$E_{stitch}(x, y, x', y') = \sum_{k=\min M}^{\max M-1} \frac{1}{2} \|I_k(x, y) - I_{k+1}(x, y)\|^2 + \frac{1}{2} \|I_k(x', y') - I_{k+1}(x', y')\|^2 \quad (10)$$

[0152] 其中：

[0153] $\min M = \min(M(x, y), M(x', y'))$

[0154] $\max M = \max(M(x, y), M(x', y'))$

[0155] 此代价在假设帧的分配是连续的情况下是合理的，这意味着如果 (x, y) 和 (x', y') 是邻近的像素，则它们的源帧 $M(x, y)$ 和 $M(x', y')$ 是接近的。此代价的主要优点是它允许我们将该问题作为对图的最小切割问题来解决。

[0156] 我们将最小化的能量函数是：

$$\begin{aligned}
 [0157] \quad E(M) = & \sum_{(x,y)} \sum_{(x',y') \in N(x,y)} E_{stitch}(x, y, x', y') \\
 [0158] \quad & + \sum_{(x,y)} (1 - Valid(x, y, M(x, y))) \cdot D \quad (11)
 \end{aligned}$$

[0159] 其中：

[0160] $N(x, y)$ 是在 (x, y) 附近的像素。

[0161] $E(x, y, x', y')$ 是对每个邻近像素的缝合代价，如等式 1 中所述的。

[0162] $Valid(x, y, k)$ 是 $1 \Leftrightarrow I_k(x, y)$ 是有效像素（即，在摄像机的视场内）。

[0163] D 是非常大的数字（代表无限）。

[0164] 9. 建立单个全景

[0165] 我们接下来显示如何将 2D 多标签问题（其具有指数复杂性）变换为 3D 二进制问题（其具有多项式复杂性，并且实际上可被快速解决）。对于每个像素 (x, y) 以及输入帧 k ，我们定义二进制变量 $b(x, y, k)$ ，当且仅当 $M(x, y) \leq k$ 时等于 1（ $M(x, y)$ 是像素 (x, y) 的源帧）。显然， $b(x, y, N) = 1$ 。

[0166] 注意，对每个 $1 \leq k \leq N$ 给定 $b(x, y, k)$ ，我们可将 $M(x, y)$ 确定为 $b(x, y, k) = 1$ 的最小 k 。我们写出能量项，其最小化将给出无缝全景。对于每个邻近的像素 (x, y) 和 (x', y') 且对于每个 k ，我们对 $b(x, y, k) \neq b(x', y', k)$ 的赋值 (assignment) 增加误差项：

$$[0167] \quad ||I_k(x, y) - I_{k+1}(x, y)||^2 + ||I_k(x', y') - I_{k+1}(x', y')||^2$$

[0168] （该误差项是对称的）。

[0169] 我们还对 $b(x, y, k) = 1$ 但 $b(x, y, k+1) = 0$ 的赋值增加无限处罚。（因为 $M(x, y) \leq k$ 但 $M(x, y) > k$ 是不可能的）。

[0170] 最后，如果 $I_k(x, y)$ 是无效像素，则我们可通过将无限处罚赋予赋值：如果 $k > 1$ ， $b(x, y, k) = 1 \wedge b(x, y, k+1) = 0$ 或 $k = 1$ 的 $b(x, y, k) = 1$ ，来避免选择这个像素。（这些赋值暗指 $M(x, y) = k$ ）。

[0171] 上面的所有项都在 3D 栅格中的变量对上，并且因此我们可描述为对 3D 二进制 MRF 最小化能量函数，并且使用最小切割 [9] 在多项式时间中最小化它。

[0172] 10. 使用 4D 最小切割建立全景电影

[0173] 为了建立全景电影（长度为 L ），我们必须建立全景图像的序列。独立地构建每个全景图像是不合适的，因为没有加强时间连贯性。另一种方式是以作为第一帧的初始镶嵌图像开始，并且对连续的镶嵌图像，从用于前面镶嵌的连续帧中取出每个像素 $(M_1(x, y) =$

$M(x, y)+1$)。这个可能性类似于上面参考附图的图 2b 描述的可能性。

[0174] 根据本发明的第二个方面,我们改为使用不同的公式,其给予缝合从一个全景帧变化为另一个的机会,这对于成功地缝合运动物体非常重要。

[0175] 我们构建由以前描述的 3D 图的 L 个实例组成的 4D 图:

$$[0176] \quad b(x, y, k, l) = 1 \Leftrightarrow M_l(x, y) \leq k$$

[0177] 为了加强时间连贯性,我们对每个 $1 < L$ 的赋值 $b(x, y, N, l) = 1$ 给予无限处罚,以及对每个 $l > 1$ 的赋值 $b(x, y, 1, l) = 0$ 给予无限处罚。

[0178] 另外,对于每个 (x, y, k, l) ($1 \leq l \leq L-1, 1 \leq k \leq N-1$),我们对赋值 $b(x, y, k, l) = 1 \neq b(x, y, k+1, l+1)$,设定代价函数:

$$[0179] \quad E_{temp} = \frac{1}{2} \|I_k(x, y) - I_{k+1}(x, y)\|^2 + \frac{1}{2} \|I_{k+1}(x, y) - I_{k+2}(x, y)\|^2 \quad (12)$$

[0180] (对于 $k = N-1$,我们仅仅使用代价函数的左项)。该代价促进显示在作为结果的电影中的(时间)连续像素(除非例如这些像素在背景中)。

[0181] 该方法的变形是在连续帧处不将每个像素 (x, y) 连接到同一个像素,而是根据在像素 (u, v) 处的光流(optical flow)连接到对应的像素 $(x+u, y+v)$ 。可以例如在 [19] 中找到计算光流的合适的方法。使用光流更好地处理运动物体的情况。

[0182] 再次,我们可利用对 4D 图的最小切割来最小化能量函数,并且二进制解决方案定义了减少缝合问题的全景电影。

[0183] 11. 实践改进

[0184] 可能需要大量的存储器来保存 4D 图。我们因此使用既减少存储器要求又减少算法的运行时间的几个改进:

[0185] ●如上所述,能量可被最小化而没有明确地保存无效像素的顶点。顶点的数量因此减少到输入视频中像素的数量乘以输出视频中帧的数量。

[0186] ●不是解决输出视频中的每一帧,我们可只解决一组采样的输出帧,并且将缝合函数插入其间。此改进基于在场景中的运动不是非常大的假设。

[0187] ●我们可将每个像素限制为仅仅来自输入帧的部分组。这对从视频中取得的帧的序列特别有意义,其中在每对连续帧之间的运动非常小。在这种情况下,我们通过对每个像素采样这组源帧而不失去很多。但是以一致的方式采样源帧是合理的。例如,如果帧 k 是在第 1 个输出帧中的像素 (x, y) 的可能源,那么 $k+1$ 帧应是在第 $1+1$ 个输出帧中像素 (x, y) 的可能源帧。

[0188] ●我们使用多分辨率框架(例如在 [2] 中完成的),其中找到了对低分辨率图像(在模糊化和子采样后)的粗略的解决方案,并且该解决方案仅仅在边界内被改进。

[0189] 12. 使兴趣分数(interest score)与视频联合

[0190] 我们现在描述一种用于根据兴趣分数来组合视频的方法。有几种应用,例如建立具有更密集(或更稀疏)的活动的电影,或甚至按用户特定的方式控制场景。

[0191] 在 [14] 中描述动态全景可以认为是特别的情况,其中同一个电影的不同部分被组合来得到具有较大视场的电影:在这种情况下,我们根据在每个时间的每个像素的“可见度”定义了兴趣分数。更一般地,组合同一个电影的不同部分可以用在其它的情况中。例如,为了使电影中的活动更密集,我们可以将电影中动作发生的不同部分组合为具有大量

动作的新电影。以上参考图 1 至 8 描述的实施方式描述了最大化活动的特别情况，并且使用不同的方法。

[0192] 应被处理的两个问题是：

[0193] 1. 如何将电影组合为“好看的”电影。例如，我们想要避免缝合问题。

[0194] 2. 最大化兴趣分数。

[0195] 我们通过描述可被使用的不同分数来开始，并且然后描述用来组合电影的方案。

[0196] 可以用作电影的兴趣函数的一个主要特征是像素的“重要”程度。在我们的实施方式中，我们考虑像素中的“活动”来表示其重要性，但是重要性的其它测量标准也是适合的。活动程度的评估本身不是本发明的特征，并且可以使用如上面在第一部分中提到的（活动检测）多种方法之一来实现。

[0197] 13. 其它分数

[0198] 可以使用其它分数来组合电影：

[0199] ●可见度分数：当摄像机移动时，或如果我们试图填充视频中的洞时，存在不可见的像素。我们可处罚（不必具有无限分数）无效像素。以这种方式，我们可鼓励填充洞（或增加视场），但可能更喜欢不填充洞，或如果它导致劣质的缝合则使用较小的视场。

[0200] ●方向：活动测量可以用方向测量代替。例如，与垂直移动的区域相比，我们可能更喜欢水平移动的区域。

[0201] ●使用者指定：使用者可以指定喜欢的兴趣函数，例如颜色、质地等等。另外，使用者可用不同的分数手动指定区域（或时间片）。例如，通过画面具，其中 1 表示期望大量的活动，而 0 表示期望没有活动，使用者可控制在场景中的动态，即发生在指定的地点。

[0202] 14. 算法

[0203] 我们使用类似于由 [20] 使用的方法的具有以下变化的方法：

[0204] ●我们为要从一个电影或另一个选择的每个像素添加兴趣分数。该分数可使用来自每个电影的每个像素的边缘被添加到终端顶点（源和汇（sink）），并且在这些边缘中的权是兴趣分数。

[0205] ●我们（可选地）计算在每个连续的帧对之间的光流。然后，为了加强连贯性，我们可根据光流（ (x, y, t) 到 $(x+u(x, y), y+v(x, y), t+1)$ ），用邻居之间的边缘代替在时间邻居（ (x, y, t) 到 $(x, y, t+1)$ ）之间的边缘。这加强了在缝合的电影之间的过渡，因为它鼓励缝合跟随较不明显的流。

[0206] ●当决定组合电影的哪些部分（或哪些电影）时，我们应不仅考虑缝合代价，而且考虑兴趣分数。例如，当建立具有更密集的活动程度的电影时，我们选择最大化分数的一组电影 S：

$$[0207] \sum_{x,y,t} \bigcup_{b \in S} \chi_b(x,y,t)$$

[0208] 图 9b 是图形表示，将其此效应展示为增加的活动密度的电影，在图 9a 中显示来自该电影的原始帧。当多于两个的电影被组合时，我们使用迭代方法，其中在每次迭代中，新的电影被组合到作为结果的电影中。为了正确地这么做，我们应考虑由以前的迭代产生的旧的接缝和分数。尽管没有兴趣分数，该方案由 [20] 描述。图 9b 显示了来自作为结果的视频的一个采样帧。

[0209] 图 10 是过程的示意图。在这个实施例中,视频与本身的时间偏移的版本组合。该组合使用根据以上所述的标准的最小切割,即,最大化兴趣分数同时最小化缝合代价来实现。

[0210] 现在参考图 11,其显示了根据本发明的系统 10 的方块图,用于将由摄像机 11 拍摄的第一动态场景的视频帧的第一序列转换为描述第二动态场景的至少两个视频帧的第二序列。所述系统包括用于存储在第一序列中的视频帧的子集的第一存储器 12,该视频帧的子集显示包括位于相应的 x、y 坐标的多个像素的至少一个物体的运动。选择单元 13 耦合到第一存储器 12,用于从子集中选择在第一动态场景中显示至少一个物体的非空间重叠外观的部分。帧生成器 14 将来自至少 3 个不同的输入帧的部分复制到第二序列的至少两个连续帧而不改变在物体中的像素的相应的 x、y 坐标,并且使得第二序列的至少一个帧包括出现在第一序列中的不同帧处的至少两个部分。第二序列的帧存储在第二存储器 15 中,用于随后的处理或由显示单元 16 显示。帧生成器 14 可包括用于使至少两个部分在复制到第二序列之前空间地弯曲的弯曲单元 17。

[0211] 系统 10 可实际上由具有图形卡或工作站的适当编程的计算机以及适当的外围设备实现,如在本领域中都被公知的。

[0212] 在系统 10 中,所述至少 3 个不同的输入帧可在时间上连续。系统 10 还可包括耦合到第一存储器的可选的对准单元 18,用于预先对准视频帧的第一序列。在这种情况下,摄像机 11 将耦合到对准单元 18,以便在第一存储器 12 中存储预先对准的视频帧。对准单元 18 通过下列步骤操作:

[0213] 计算在第一序列中的帧之间的图像运动参数;

[0214] 使在第一序列中的视频帧弯曲,以便在第一动态场景中的静止物体在视频中是静止的。

[0215] 同样,系统 10 也可包括耦合到选择单元 13 的可选的时间片段生成器 19,用于通过“时间波前 (time front)”表面扫描对准的时空体积,并且产生时间片段的序列。

[0216] 这些可选的特征没有被详细地描述,因为在进行了引用的上述 W02006/048875 中完整地描述了它们以及术语“时间波前”和“时间片段”。

[0217] 为了完整性起见,图 12 是流程图,其显示了根据本发明由系统 10 执行的基本操作。

[0218] 15. 讨论

[0219] 视频概要被提议为一种用于将视频中的活动浓缩为非常短的时间段的方法。此浓缩的表示可实现对视频序列中的活动的有效访问。介绍了两种方法:一种方法使用低层次的图形最优化 (graph optimization),其中在视频概要中的每个像素是该图形中一个节点。这种方法具有直接从输入视频得到视频概要的好处,但是该解决方案的复杂性可能非常高。可选的方法是首先检测运动物体并且对检测的物体执行最优化。尽管在第二种方法中需要运动分割的预备步骤,但是它快得多,并且基于物体的限制是可能的。在作为结果的视频概要中的活动比在任何普通视频中的活动浓缩得更多,并且观看这种概要对于无经验的观看者可能看起来难以使用。但是当目的是在短时间内观察很多信息时,视频概要提供这个目的。应对得到动态频闪的可能性给予特别的注意。尽管允许视频概要的长度进一步减少,但是动态频闪可能需要使用者进一步适应。确实需要一些训练来实现表示较长活动

时间的单个物体的多个空间出现。尽管我们详细描述了对于动态视频概要的特别实现,但许多扩展是简单直接的。例如,活动指示器可以是连续的,而不是利用二进制“活动”指示器。连续的活动可扩展可用于建立视频概要的选择,例如通过基于物体的活动水平控制显示的物体的速度。视频概要也可以应用于由许多镜头组成的长电影。理论上,由于闭塞(或不连续)处罚,我们的算法不结合来自不同场景的部分。在这种情况下,用于单个镜头的简单背景模式必须用可调节的背景估算器代替。可应用在长电影中的另一方法是使用用于镜头边界检测的已有方法,并且对每个镜头单独地建立视频概要。

[0220] 也应理解,根据本发明的系统可以是适当编程的计算机。同样,本发明设想计算机可读的计算机程序用于执行本发明的方法。本发明还设想机器可读存储器,其可触知地包括由用于执行本发明的方法的机器可执行的指令的程序。

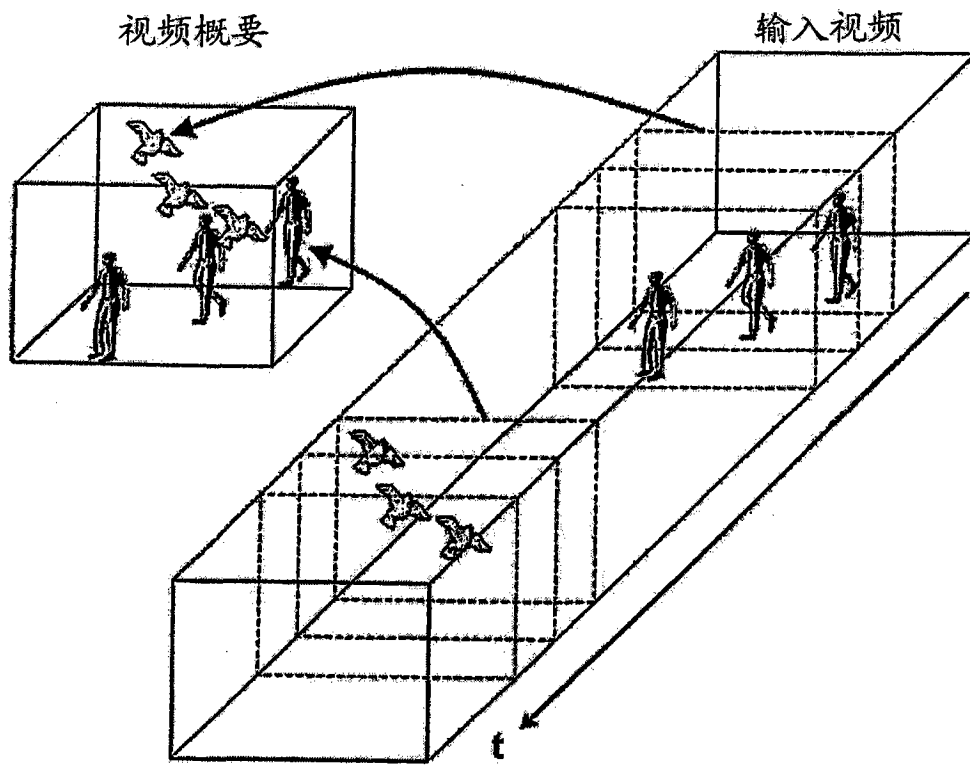


图 1

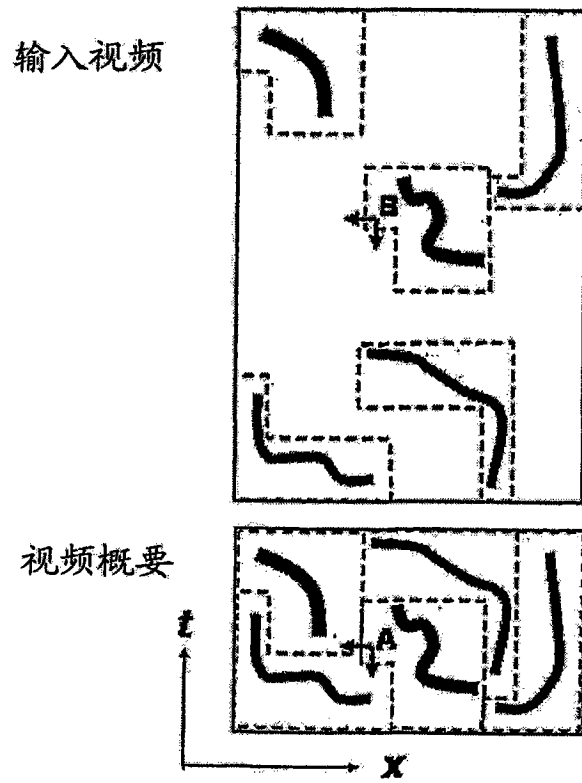


图 2a

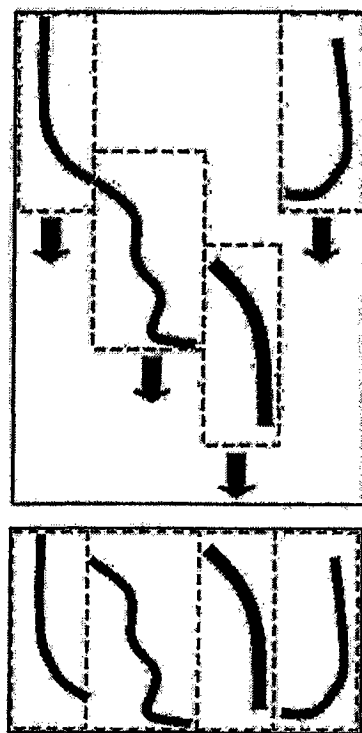


图 2b

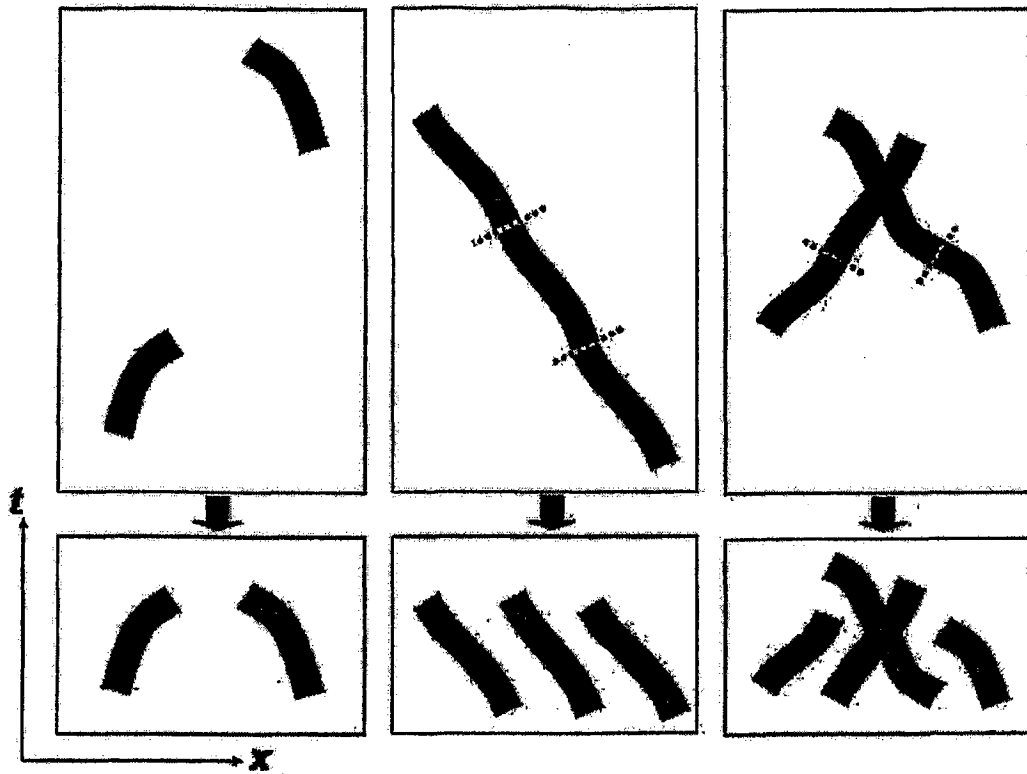


图 3a

图 3b

图 3c

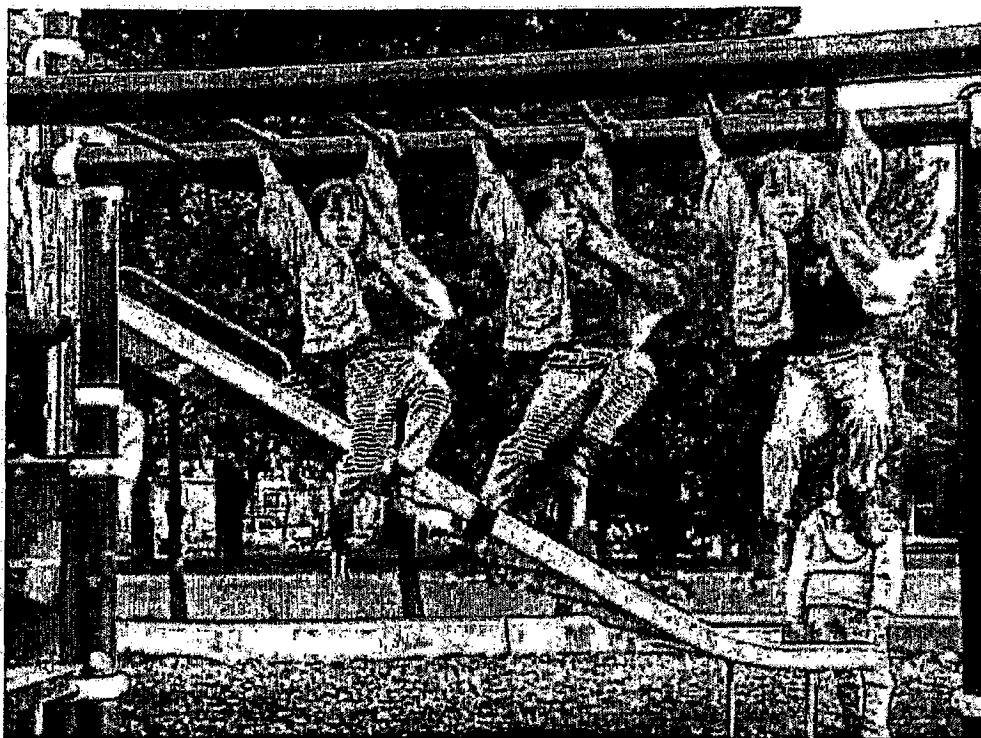


图 4

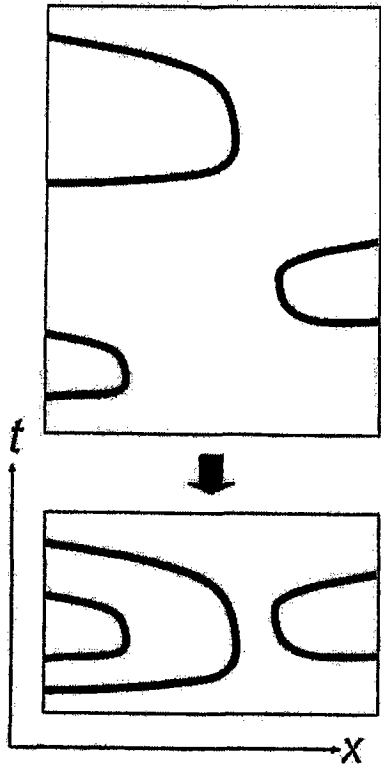


图 5a



图 5b



图 5c

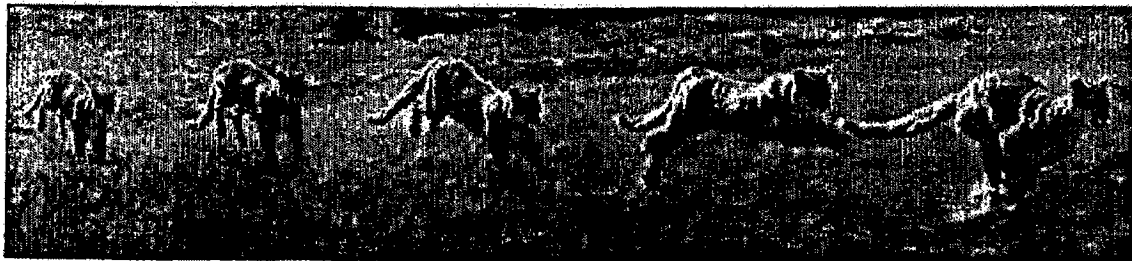


图 6



图 7a



图 7b



图 7c

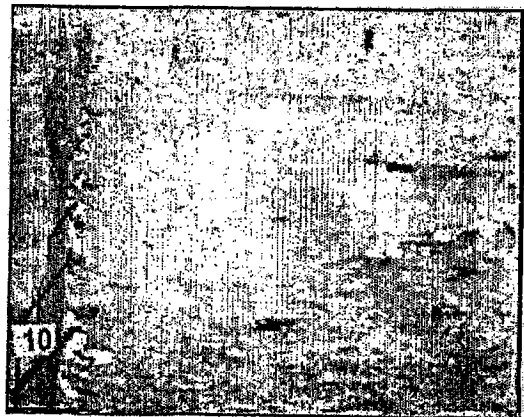


图 8a



图 8b



图 9a

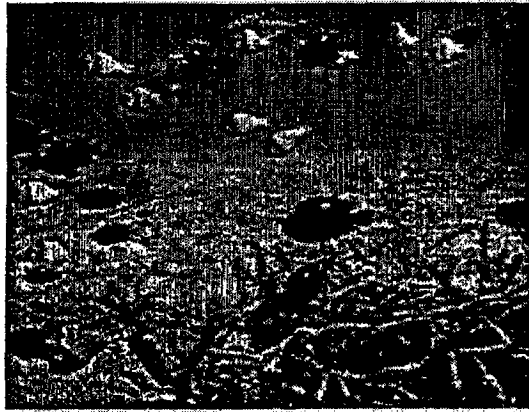


图 9b

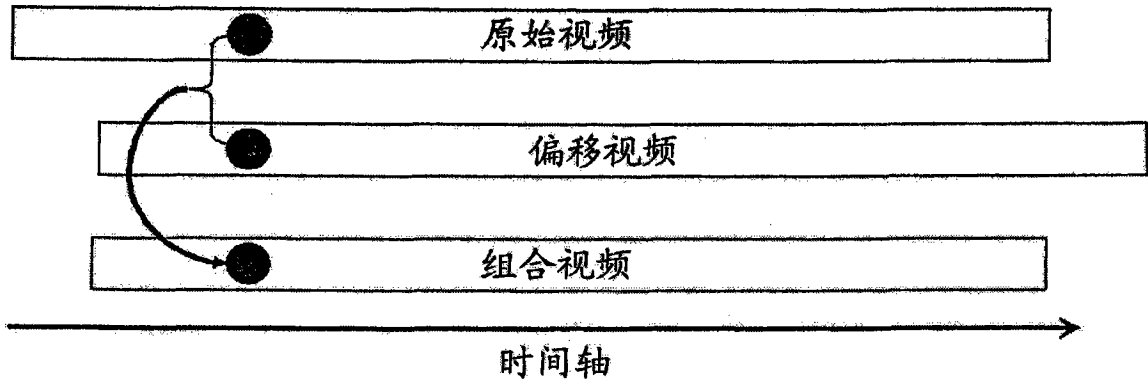


图 10

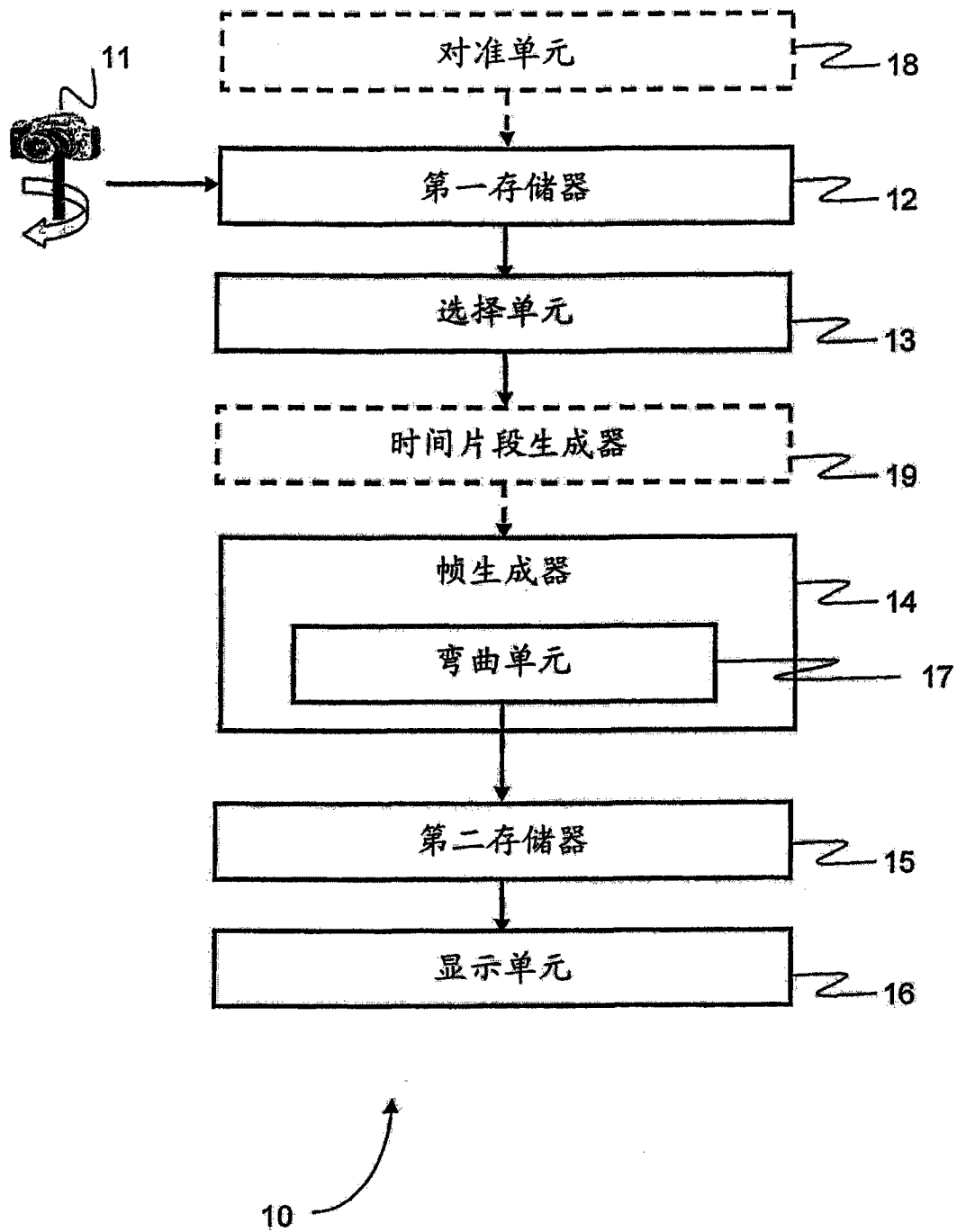


图 11

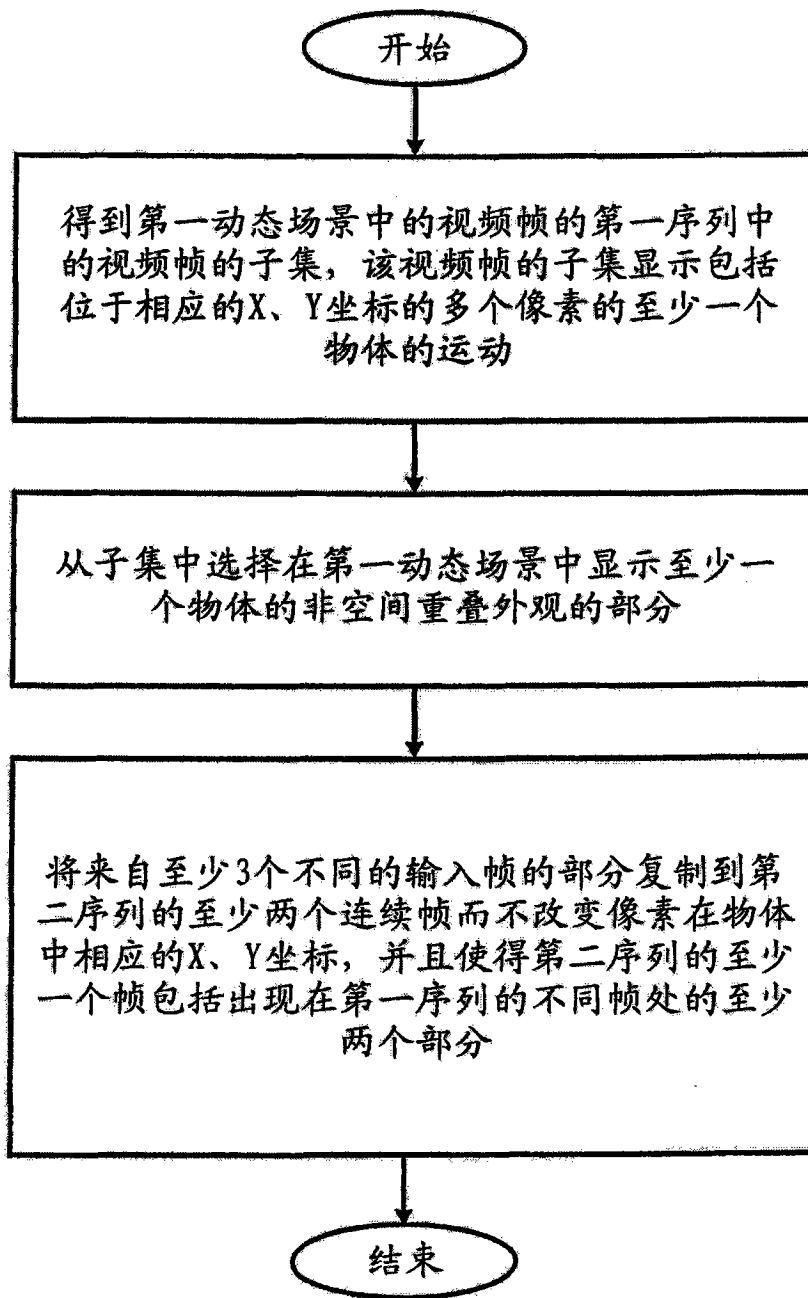


图 12