(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 102742260 B
(45)授权公告日 2015.07.15

(21)申请号 201180009337.3

(22)申请日 2011.02.05

(30)优先权数据
12/704,047 2010.02.11 US

(85)PCT国际申请进入国家阶段日
2012.08.13

(86)PCT国际申请的申请数据
PCT/US2011/023839 2011.02.05

(87)PCT国际申请的公布数据
WO2011/100174 EN 2011.08.18

(73)专利权人 微软技术许可有限责任公司
地址 美国华盛顿州

(72)发明人 吴勇军 N.博里索夫 赵伟东
S.萨瓦尼 N.瑟姆普迪

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 刘鹏 汪扬

(51)Int.Cl.
*H04N 5/235*(2006.01)

(56)对比文件
US 2007/0292047 A1,2007.12.20,说明书第【0080】、【0081】、【0126】段.
US 2007/0242900 A1,2007.10.18,说明书第【0045】、【0080】、【0081、【0126】段.
US 2007/0242900 A1,2007.10.18,说明书第【0045】、【0080】、【0081、【0126】段.

审查员 张维克

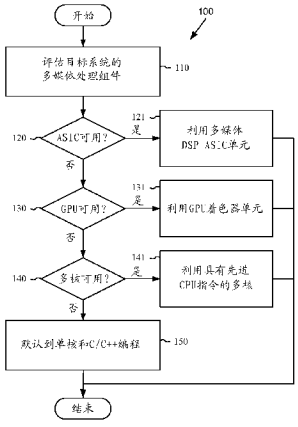
权利要求书3页 说明书11页 附图10页

(54)发明名称

通用平台视频图像稳定化

(57)摘要

通过评估可用的多媒体数字信号处理组件，并且选择可用的组件以根据层级结构用于视频稳定化性能以处理视频稳定化的部分，视频图像稳定化在用于计算设备的通用平台上提供了更好的性能。视频稳定化具有改善的运动矢量估计，其从视频帧的下采样的分辨率版本开始，根据金字塔块结构关系采用精细运动矢量搜索。通过执行用于处理局部运动矢量的随机采样一致性办法和用于运动矢量可靠性的选择判据，视频稳定化还改善了全局运动变换估计。通过利用误差检测和校正的实时一阶段或离线二阶段时间平滑，视频稳定化平滑地实现手晃动的去除。




CN 102742260 B

1. 一种数字视频稳定化的方法，该方法包括在计算设备上执行以下步骤：

评估所述计算设备以确定所述计算设备上各种多媒体数字信号处理组件中任何一个的可用性；

从所述评估的结果，按照至少包括该数字视频稳定化的性能的判据，根据对多媒体 DSP 组件排序的层级结构，确定哪一个可用的多媒体 DSP 组件用于数字视频稳定化的至少一部分，所述数字视频稳定化至少包括：局部运动矢量估计以产生估计出的局部运动矢量；处理所述估计出的局部运动矢量以估计表示抖动视频运动的全局运动变换；以及基于全局运动变换进行图像扭曲以补偿抖动视频运动；以及

对于所述数字视频稳定化的所述至少一部分，利用所确定出的可用多媒体 DSP 组件在所述计算设备上执行视频片段的数字视频稳定化，其中所述执行数字视频稳定化包括：基于应用于所述视频片段的至少一个帧的全局运动变换扭曲所述视频片段的所述至少一个帧。

2. 如权利要求 1 所述的方法，其中，所述层级结构依次包括多媒体 DSP ASIC 单元、GPU 着色器单元、多核 CPU 和单核 CPU。

3. 如权利要求 1 所述的方法，其中，当所述评估确定 GPU 着色器在所述计算设备上可用时，则所述数字视频稳定化将 GPU 着色器用于局部运动矢量估计。

4. 如权利要求 3 所述的方法，其中，当所述评估确定 GPU 着色器在所述计算设备上可用时，则所述数字视频稳定化进一步将 GPU 着色器用于所述图像扭曲。

5. 一种数字视频稳定化的方法，该方法包括在计算设备上执行以下步骤：

针对视频片段的场景的多个帧中的帧，执行局部运动矢量估计，其中，对于所述帧，局部运动矢量估计包括：

对所述帧和与所述帧相关联的基准帧下采样一次或更多次，以便以多个分辨率形成所述帧的多个版本，其中，所述帧的最低分辨率版本具有在所述帧的每个更高分辨率版本中对应位置处作为金字塔结构与相继更大数目的块相关联的多个块；

对于所述帧的最低分辨率版本，使用相关联的最低分辨率版本基准帧的初始运动矢量估计处理来估计所述最低分辨率版本帧的块的运动矢量；

对于所述帧的每个相继更高分辨率的版本，根据金字塔结构，从针对在前更低的分辨率版本帧的相关联块估计出的运动矢量开始，使用精细搜索估计所述相继更高分辨率版本帧的块的运动矢量；以及

产生针对所述帧的原始分辨率版本的块的所估计出的局部运动矢量；

处理估计出的所述帧的局部运动矢量，用于估计表示抖动视频运动的全局运动变换，包括：

在产生的估计出的所述帧的局部运动矢量中，选择运动矢量的集合以用于全局运动变换的估计，其中，所述选择运动矢量的集合包括以下项中的一项或多项：

排除图像边界上的运动矢量；

排除具有高运动补偿残差的运动矢量；以及

排除具有低图像内容变化的块的运动矢量；以及

基于针对所述帧估计出的全局运动变换，在所述帧上执行图像扭曲，以补偿抖动视频运动。

6. 如权利要求 5 所述的方法，其中，所述处理估计出的所述帧的局部运动矢量以估计全局运动变换，使用随机采样一致性来处理所述帧的局部运动矢量以去除异常运动矢量。

7. 如权利要求 5 所述的方法，其中，所述处理估计出的所述帧的局部运动矢量包括：

将从处理所述帧的局部运动矢量估计出的全局运动变换与下限和上限进行比较；

如果估计出的全局运动变换超过所述上限，则重置所述全局运动变换；

如果估计出的全局运动变换超过所述下限但没有超过所述上限，则将全局运动变换限制到所述下限；以及

否则，使用估计出的全局运动变换代表抖动视频运动。

8. 如权利要求 7 所述的方法，进一步包括：

作为视频片段的场景中帧上的第一遍，处理帧以估计视频片段的场景中的帧的全局运动变换；

确定视频片段的场景中的帧的全局运动变换的参数的概率分布；

基于概率分布确定所述下限和所述上限；以及

在处理视频片段的场景中的帧的第二遍中，将所述下限和上限应用于全局运动变换的估计。

9. 如权利要求 5 所述的方法，其中，所述处理估计出的所述帧的局部运动矢量包括：

基于相似性运动模型，使用随机采样一致性和最小均方误差估计全局运动变换的参数，来处理所述帧的运动矢量；以及

将估计出的全局运动变换的参数与关于所述参数的下限和上限比较；

如果估计出的全局运动变换的参数中的任何一个超过其上限，则将全局运动变换重置；

如果估计出的全局运动变换的参数中的任何一个超过下限但没有超过上限，则将估计出的全局运动变换的参数限制到其下限；以及

否则，使用估计出的全局运动变换代表抖动视频运动。

10. 如权利要求 5 所述的方法，其中，所述处理估计出的所述帧的局部运动矢量包括：使用高斯滤波器将时间平滑应用于所估计出的全局运动变换；以及

其中，所述处理估计出的所述帧的局部运动矢量包括：使用视频片段的场景中的帧上受约束的全局优化，将时间平滑应用于所估计出的全局运动变换。

11. 一种用于处理视频以将视频图像稳定化应用于所述视频的数字视频处理设备，所述数字视频处理设备包括：

用于评估数字视频处理装置的至少一个数字信号处理组件的构件；

用于按照包括至少数字视频稳定化的性能的判据，根据用于对数字信号处理组件排序的层级结构，确定所述至少一个数字信号处理组件中的哪个用于数字视频稳定化的至少一个部分的构件；以及

用于部分地通过处理视频的帧以估计帧的局部运动矢量、处理估计出的局部运动矢量以用于估计表示抖动视频运动的全局运动变换以及基于全局运动变换扭曲帧以补偿抖动视频运动来在视频上执行数字视频稳定化的构件，其中，所述执行数字视频稳定化执行通用平台视频稳定化库程序中的编程功能，以将确定出的数字信号处理组件用于所述数字视频稳定化的所述至少一部分。

12. 如权利要求 11 所述的数字视频处理设备，其中，所述处理帧以估计帧的局部运动矢量包括使得所确定出的至少一个数字信号处理组件执行以下行动：

对所述帧和与所述帧相关联的基准帧下采样一次或更多次，以便以多个分辨率形成所述帧的多个版本，其中，所述帧的最低分辨率版本具有在所述帧的每个更高分辨率版本中对应位置处作为金字塔结构与相继更大数目的块相关联的多个块；

对于所述帧的最低分辨率版本，使用相关联的最低分辨率版本基准帧的初始运动矢量估计处理来估计所述最低分辨率版本帧的块的运动矢量；

对于所述帧的每个相继更高分辨率的版本，根据金字塔结构，从针对在前更低分辨率版本帧的相关联块估计出的运动矢量开始，使用精细搜索估计所述相继更高分辨率版本帧的块的运动矢量；以及

产生针对所述帧的原始分辨率版本的块的所估计出的局部运动矢量。

13. 如权利要求 12 所述的数字视频处理设备，其中，所述处理估计出的局部运动矢量以用于估计表示抖动视频运动的全局运动变换包括：基于相似性运动模型，使用随机采样一致性和最小均方误差估计全局运动变换的参数，来处理估计出的所述帧的局部运动矢量。

14. 如权利要求 13 所述的数字视频处理设备，其中，所述处理估计出的局部运动矢量以用于估计表示抖动视频运动的全局运动变换包括：

将估计出的全局运动变换的参数与关于下限和上限比较；

如果估计出的全局运动变换的参数中的任何一个超过其上限，则将全局运动变换重置；

如果估计出的全局运动变换的参数中的任何一个超过下限但没有超过上限，则将估计出的全局运动变换的参数限制到其下限；以及

否则，使用估计出的全局运动变换代表抖动视频运动。

# 通用平台视频图像稳定化

## 背景技术

[0001] 将摄像机能力并入到多用途、小型且轻量型的手持式电子设备（如，移动电话、袖珍相机和媒体播放器）越来越常见。不同于更大的专用用途的视频摄录机，这种多用途的手持式设备通常缺少任何机械或光学机制以降低由于晃动或不稳定的手所引起的抖动视频运动。由于其更轻的重量和广角透镜的典型应用，多用途手持式设备可能更加易于受到来自手晃动的抖动视频运动的影响。进一步，随着这些并不昂贵的多用途手持式设备的可用性向业余爱好者和专业摄影师群体以外的拍摄视频的人群散布，消费者视频越来越常见地是由关于如何拍摄高质量视频的具有很少训练或经验的用户产生的。通过因特网经由电子邮件、博客、视频共享网络站点和社交联网站点在消费者之间共享短视频剪辑也日益流行。尤其对于并非是摄像专业人士或爱好者的那些用户，经常在没有任何剪辑或其它处理的情况下将视频剪辑上传至网络站点（有时直接从视频捕获设备上传至网络站点）。出于这些各种各样的原因，在网络上共享的视频剪辑的质量经常十分差，并且视频剪辑常常遭受手晃动引起的抖动视频运动。

[0002] 数字视频图像稳定化是可应用于视频以校正来自手晃动的抖动视频运动的数字信号处理技术。在一个示例性实施方案中，该技术包括：计算视频序列的每个图像相对于其先前图像的宏块的局部运动估计；处理宏块的局部运动矢量以产生由于抖动引起的全局运动的估计；然后，在与所估计出的抖动运动相反的方向上，通过数字平移或扭曲图像来补偿抖动视频运动。

[0003] 已知的数字视频图像稳定化的一个缺陷是，该技术在计算上十分密集。当将视频上传至博客、视频共享网络站点或社交联网站点时，可以从在处理能力上不同的设备上传视频。而且，临时用户可能对于将其视频快速发布至视频共享或社交联网站点的即时性更加感兴趣，使得视频的任何耗时处理都是不期望的。可以例如通过蜂窝网络从多用途手持式设备直接上传视频。然而，多用途手持式设备（如，移动电话）经常具有有限的处理资源，或者必须与该多用途设备的其它能力共享处理资源。可替代地，手持式设备也可以连接至 PC、膝上型计算机、上网本或者具有用于将视频上传至视频共享或社交联网站点的因特网连接性的类似设备。然而，这些因特网连接的 PC 在处理资源方面差别也很大。此外，为了即时性的利益，发布期间经由因特网连接的设备的任何处理延迟可能都是不期望的。

[0004] 出于这些原因，跨越各种可用的计算平台有效操作的视频图像稳定化技术将是有用和期望的。

## 发明内容

[0005] 概括地说，详细的描述总体上涉及用于视频序列的图像稳定化的数字视频处理。

[0006] 下面的具体实施方式部分涉及在通用平台上实施视频稳定化的系统和方法，其有效地利用可用的多媒体数字信号处理组件以提供更好的性能。例如，在可执行的库中编程通用平台上的视频稳定化，该可执行库提供编程功能以利用可以被提供在具有通用平台的计算设备上的各种多媒体数字信号处理组件中任何一个。通过对于计算设备上的任何可用

多媒体数字信号处理组件估计其运行于其上的计算设备，视频稳定化的实施在通用平台上操作。然后，在通用平台上的视频稳定化的实施按照至少包括数字视频稳定化的性能的判据，根据对多媒体数字信号处理组件排序的层级结构，确定哪一个或哪些可用的多媒体数字信号处理组件用于处理视频稳定化的一个或多个部分。实施方案然后处理视频图像稳定化的视频，包括利用所确定出的多媒体数字信号处理组件在视频图像稳定化的处理的那些部分上执行可执行的库函数。

[0007] 在下面的具体实施方式部分中给出的一些实施方案中，视频图像稳定化使用强调视频帧的相邻块之间的运动矢量的一致性的局部运动矢量估计的形式。在这种局部运动矢量估计中，帧和其基准帧被下采样一次或更多次，以便以多个分辨率形成帧和基准帧的多个版本。帧的最低分辨率版本具有在帧的每个更高分辨率版本中对应位置处作为金字塔结构与相继更大数目的块相关联的多个块。运动矢量估计过程开始于帧的最低分辨率版本，使用相关联的最低分辨率版本基准帧的全搜索估计最低分辨率版本帧的块的运动矢量。然后，对于帧的每个相继更高的分辨率版本，根据金字塔结构，从针对在前更低分辨率版本帧的相关联块估计出的运动矢量开始，使用精细搜索估计更高分辨率版本帧的块的运动矢量。在对于每个更高分辨率版本重复直至该帧的全分辨率版本后，这产生全分辨率版本帧的块的所估计局部运动矢量，其更好地强调相邻块之间一致性。局部运动矢量的这种更加一致性的估计更好地强调了帧的全局运动。

[0008] 在下面的具体实施方式部分中给出的某些实施方案中，视频图像稳定化包括处理运动矢量以用于更加可靠地估计视频中帧的全局运动的方式。这些包括：诸如通过消除画面边界上的运动矢量、消除产生高运动补偿残差的运动矢量和具有低方差的块的运动矢量，选择更加可靠地反映帧全局运动的运动矢量的集合。视频图像稳定化也可以包括通过使用随机采样一致性办法估计全局运动变换的参数来检测和移除运动矢量异常值(outlier)。视频图像稳定化还可以基于类似性运动模型估计全局运动变换，该模型可更好对来自手晃动的抖动视频运动建模。

[0009] 在下面的具体实施方式部分中给出的某些实施方案中，视频图像稳定化还可以经历与抖动视频运动对应的全局运动变换的估计，以限制用于全局运动估计中误差的检测和校正的阈值。视频图像稳定化将所估计出的全局动作转换的参数与两个界限进行比较。如果超过了上限，则全局运动变换的估计被重置。如果参数超过下限，则估计出的全局运动变换的参数被限制到其下限。否则，不改变地使用所估计出的全局运动变换。

[0010] 在下面的具体实施方式部分给出的某些实施方案中，视频图像稳定化执行针对帧估计出的全局运动变换的时间平滑。

[0011] 下面的具体实施方式部分给出了可在实时回放或代码转换场景下采用的视频图像稳定化的变型，以及适用于离线视频稳定化的变型。在实时模式中，可以使用局部运动矢量的单遍(pass)处理执行视频图像稳定化，以产生帧的全局动作的估计。在离线模式下，视频图像稳定化可以在两遍中执行运动矢量处理。例如，第一遍运动矢量处理可以用于聚集关于运动矢量的统计信息，如跨视频片段的场景的帧的运动矢量的概率分布。这种统计信息可用于自适应地确定被适用于全局运动变换估计的界限。也可以在全局运动变换的时间平滑中使用来自第一遍的信息。例如，该信息可以用于确定对于时间平滑的受约束全局优化的约束限制。另外，离线模式下的第一遍可以用于检测视频片段中的场景改变。

[0012] 此发明内容部分被提供用来以简化的形式介绍下面在具体实施方式部分中进一步描述的构思的选择。此发明内容部分并非旨在标识所要求保护主题的关键特征或必不可少的特征，也并非旨在用作辅助确定所要求保护主题的范围。从参照附图进行的实施例的以下详细描述中，将使得本发明的另外特征和优点明显。

## 附图说明

[0013] 图 1）是通用平台数字视频图像稳定化技术根据计算性能层级最有效地利用可用的多媒体数字信号处理组件的方法的流程图。

[0014] 图 2）是三部分数字视频图像稳定化处理的流程图，其根据图 1）的层级利用处理利用目标计算设备的可用多媒体数字信号处理组件。

[0015] 图 3）是图示用于基于金字塔块的局部运动估计的视频帧的下采样的示图。

[0016] 图 4）、5）、6） 和 7） 是图示基于金字塔块的局部运动估计的示例视频帧的处理的示图。

[0017] 图 8）是图示全局运动变换模型的示图，其包括用于数字视频图像稳定化的全局运动变换的相似性模型。

[0018] 图 9）是图示视频帧的运动矢量的单遍处理以估计由于手晃动引起的抖动运动的全局运动变换、比如以用在实时模式数字视频图像稳定化中的流程图。

[0019] 图 10）是图示视频帧的运动矢量的两遍处理以估计由于手晃动引起的抖动运动的全局运动变换比如以用在离线模式数字视频图像稳定化中的流程图。

[0020] 图 11）是可以实践数字视频图像稳定化技术的计算环境的框图。

[0021] 图 12）和 13）是可以实践数字视频图像稳定化技术的云计算网络环境的框图。

## 具体实施方式

[0022] 下面的详细描述涉及用以提供数字视频图像稳定化的系统和技术，更具体地，涉及在有效地利用可用的多媒体数字信号处理 (DSP) 硬件组件的同时可在通用计算硬件平台上操作的数字视频图像稳定化技术。可以跨越多种硬件设备实施数字视频图像稳定化技术，所述硬件设备包括手持式和便携式计算设备、摄像机、移动电话、娱乐控制台（例如，视频游戏控制台和电视机顶盒）、各种网络客户端计算机（例如，个人计算机、膝上型计算机、上网本和平板计算机）以及服务器计算机。数字视频图像稳定化技术也可以用在多种的用途和计算情境中，包括专用视频捕获设备上的视频处理、独立计算机上的视频处理、网络客户端计算机上的视频处理和服务器计算机上的视频处理。进一步，可以在多个计算设备上（如，在客户端 / 服务器、网络“云”服务或对等计算布置等等中）并行地或协作地执行数字视频图像稳定化技术的各种部分。因此，应当认识到，可以在多种不同的电子和计算设备上实现所述技术，包括终端用户消费者操作的设备以及服务器计算机两者，其可以提供所述技术作为提供给消费者的服务的一部分。

[0023] 用于数字视频图像稳定化的通用平台上的多媒体 DSP组件的层级选择

[0024] 视频的处理，更具体地，这里描述的数字视频图像稳定化技术的处理，不可避免地在计算上是密集的。另一方面，可以在多种计算设备上实践数字视频图像稳定化技术，所述计算设备的多媒体处理硬件组件的配置可以相对于彼此而十分不同。这里描述的数字视频

图像稳定化技术实现更好的潜在性能和质量的一种方式是：所述技术评估其所运行于的目标计算设备，并且根据一种考虑了性能、质量、功耗、一致性（conformance）和鲁棒性中一个或多个方面的方式而构造的层级来进行选择，以利用可用的多媒体处理组件。以此方式，数字视频图像稳定化技术被开发用于通用操作平台，然后适配为最佳地利用其所运行于其上的实际目标计算硬件的多媒体处理能力。

[0025] 在一个示例实施方案中，在要在计算设备（下面更详细地描述）上运行的可执行程序（如，动态链接库文件（DLL）或应用程序可执行文件）中实施下面描述的数字视频图像稳定化技术。当实施为 DLL 或其它的可执行库文件时，可执行程序将数字视频图像稳定化实现为库函数的集合，并且提供程序化的接口用于应用程序进行由可执行程序实施的数字视频图像稳定化功能的程序化用途。可执行程序运行在通用平台或计算环境上，也就是说，可执行程序可以运行在多种计算机和计算设备上，这些设备可以包括不同的多媒体数字信号处理（DSP）组件。

[0026] 参照图 1），数字视频图像稳定化库程序包括：使用在目标计算设备（程序运行在其上）上潜在地可用的各种多媒体 DSP 组件编程以执行数字视频图像稳定化的视频处理。然后，经由多媒体利用层级处理 100，程序适配为最佳地利用多媒体 DSP 组件。在此处理 100 的第一动作 110 中，库程序评估它已经安装在其上并且正在其上运行的目标计算系统或设备的多媒体 DSP 组件。库程序可以通过如下这样执行此评估：通过对用于查询系统信息的操作系统服务做出调用、通过检查注册表（registry）或其它系统数据库中记录的系统信息、通过调用与多媒体 DSP 组件相关联的编程接口、通过尝试访问多媒体 DSP 组件、或通过其它类似的方式，以访问系统信息和 / 或查询硬件组件的存在。

[0027] 基于动作 110 中评估的结果，库程序然后根据处理层级，从被确定为在目标计算设备上可用的多媒体 DSP 组件中进行选取。在例示的示例中，库程序首先选取利用被适配用于目标计算设备中的多媒体 DSP 的任何可用的 ASIC 单元，如动作 120-121 中所示。当前商业上可用的这样的 ASIC 单元的示例包括来自因特尔的视频卡（被称为 Larrabee）和来自 Nvidia 的视频卡（名称为 Tesla）。如果在目标设备上没有这种 ASIC 单元是可用的，则库程序接下来利用目标设备上的具有着色器能力的任何可用的图形处理单元，如动作 130-131 中所示。如果不存在 ASIC 或 GPU 着色器组件，则库程序选取利用具有先进多媒体指令集（例如，SSEx 指令）的任何可用的多核处理器或中央处理单元。最后，如果这些更高性能多媒体 DSP 组件都在目标设备上不可用时，则库程序返回到使用 C/C++ 编程执行单核 CPU 上的数字视频图像稳定化技术的视频处理，如动作 150 中所示。在其它可替换的实施方案中，可以基于针对然后可以在通用平台的商业上可用的计算设备硬件中可用的多媒体 DSP 组件的上述性能、质量、功耗、一致性和鲁棒性方面的考虑，使用多媒体 DSP 组件的其它层级。与所例示的层级多媒体 DSP 利用处理 100 中相比，这些可替换的实施方案可包括利用更少或附加的多媒体 DSP 组件。

[0028] 通用平台上的数字视频图像稳定化

[0029] 如图 2）中所例示的，数字视频图像稳定化技术执行主题视频的处理，其通常包括三个部分：局部运动估计 210、用于估计与来自手晃动的视频抖动对应的全局运动变换的运动矢量处理 220、以及用于补偿视频抖动运动的图像扭曲 230。局部运动估计 210 和图像扭曲 230 在计算上是密集的，并且最期望使用来自根据图 1）中所示且在上面讨论的层级

DSP 利用而选取的目标计算设备上的任何可用多媒体 DSP 组件的选取来完成。例如，尤其适用于利用 GPU 着色器的处理的局部运动估计 210 的形式可以在库程序中实现，并且在其可用的多媒体 DPS 组件之中，当目标计算设备具有 GPU 着色器能力时，用于完成数字视频图像稳定化的局部运动估计部分。另一方面，库程序可包括这样的功能：当确定评估的目标计算设备在其可用的多媒体 DSP 组件之中拥有这种能力时，使用 D3D API 调用执行图像扭曲 230。在示例实施方案中，在 CPU 上完成处理局部运动矢量以估计全局运动变换。

[0030] 在 GPGPU上基于金字塔块的 ME

[0031] 在一个示例实施方案中，用于数字视频图像稳定化的库程序包括：使用图 3)-7) 中例示的基于金字塔块的运动估计技术（其尤其非常适用于 GPU 着色器中的处理）进行编程以实施局部运动估计 210（当通过上面讨论的过程 100 确定在目标计算设备上可用时）。相比于针对全分辨率视频帧的块单独估计局部运动，基于金字塔块的运动估计技术还强调视频帧的全局运动，其正好适用于为数字视频图像稳定化估计视频帧的总体全局运动。

[0032] 基于金字塔块的运动估计针对视频的每个帧执行相对于基准帧的运动估计。一般而言，基准帧是视频序列中的连续帧（例如，直接前一帧或后一帧）。可替代地，可以将每个第三帧用作视频的下两个帧的基准，或者可以做出基准帧的其它选择。对于运动估计，将视频帧分为块的规则网格，并且所述运动估计计算视频帧的每个块相对于基准帧中紧密匹配的图像内容的运动矢量或位移。

[0033] 基于金字塔块的运动估计开始于使用计算设备的 GPU 上的下采样着色器按一半地重复下采样视频帧和其基准帧两者的分辨率。在图 3) 所示的示例中，按一半从头到尾三次下采样视频帧和基准帧，以产生全分辨率、1/2 分辨率、1/4 分辨率和 1/8 分辨率的视频帧和基准帧的版本。可以固定次数地在分辨率方面下采样视频帧和基准帧，或者可以根据可用图像缓冲空间和计算设备的处理资源，依据视频帧的尺寸和原始分辨率，可变次数地下采样视频帧和基准帧。可替代地，相比于图 3) 中所示的示例，可以更少或者更多次地下采样视频帧和基准帧。应当理解，图 3) 中例示的示例只是例示性的，并且数字视频图像稳定化技术处理的实际视频帧通常将具有比所示出的示例更大数目的像素和更高的原始分辨率。

[0034] 对于视频帧的每个分辨率，基于金字塔块的运动估计将视频帧分为块的网格。在每个分辨率处，使用同一尺寸的块。优选地，使用相对大的块尺寸，如 16×16 或 8×8 像素。以此方式，在下一更高分辨率图像中的对应位置，将最低分辨率图像的每个块分割为同一块尺寸的 4 个块，其有效地提供了最低分辨率图像中的块到更高分辨率图像中对应位置处那些块的金字塔结构或层级关系。为了例示的目的，在图 3) 中的粗线边界内，示出了与 1/8 分辨率图像的左上块处于金字塔结构关系的、视频帧的更高分辨率版本（即，1/1、1/2 和 1/4 分辨率）中的块。

[0035] 接下来，基于金字塔块的运动估计基于最低分辨率图像中的每个块相对于全基准帧的绝对差之和 (SAD) 执行全搜索，以发现基准帧中匹配图像内容的位置。这产生最低分辨率视频帧中每个块的局部运动矢量，对应于其相对于基准帧中匹配图像内容的位移。如图 4) 中所示，仅在视频帧和基准帧的最低分辨率版本执行这种全搜索。

[0036] 如图 5)-7) 中所例示的，基于金字塔块的运动估计然后在视频帧的每个相继更高分辨率版本相对于基准帧执行精细运动估计搜索，直到原始分辨率视频帧。如之前提到的，

在视频帧的下个更高分辨率版本的对应位置处，更低分辨率图像的每个块分裂成四个块。对于金字塔结构中处于视频帧的下一更高分辨率版本的四个块的每一个，将经由运动估计搜索针对更低分辨率图像中的块计算出的运动矢量用作对于精细运动估计搜索的开始点。使用更低分辨率视频帧中的块到下一更高分辨率视频帧中的四个对应块的金字塔结构关系的这种精细搜索强调相邻块之间运动估计计算中的一致性。由于针对视频帧的每个相继更高分辨率版本重复精细搜索，因此对于下面讨论的全局运动变换估计所基于的原始分辨率图像中的块，这种强调的领域中的一致性提供了局部运动矢量的更加一致的集合。

[0037] 一旦 GPU 着色器以原始视频帧分辨率完成局部运动估计，连同差结果之和与对于原始分辨率视频帧的每个块的方差一起，将运动矢量复制到 CPU 以用于全局运动变换估计。

[0038] 用于全局运动变换估计的运动矢量处理

[0039] 再次参照图 2)，在计算设备的中央处理单元 (CPU) 上执行在一个示例实施方案中用以估计数字视频图像稳定化处理 200 的全局运动变换部分 220 的运动矢量处理。数字视频图像稳定化技术可以使用运动矢量处理的单遍模式，其适用于视频回放、视频流或传输期间的实时处理。可替代地，可以采用运动矢量处理的两遍模式，其以增加的计算时间为代价提供更好的视频稳定化质量。因此，两遍模式可能更加适用于未经历实时视频回放的时间约束并且对于其而言期望高视频稳定化质量的离线视频处理场景。

[0040] 图 8) 例示可用于估计来自手晃动的抖动视频运动的各种全局运动变换模型，包括平移相似性 (translation similarity)、欧几里得、投影和仿射运动变换模型。在这里提出的示例实施方案中，运动矢量处理使用带有平移、旋转和缩放参数的相似性全局变换模型，如下列等式 (1) 所示：

[0041]
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s\cos\beta & s\sin\beta & tx \\ -s\sin\beta & s\cos\beta & ty \\ 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad (1)$$

[0042] 其中，x 和 y 是原始位置，x’和 y’是变换的位置，而 s， β， $t_x$， $t_y$分别是缩放、旋转和平移参数。

[0043] 一般而言，大多用户将会仅体验到对于平移、旋转和缩放的晃动动作效果。于是，相似性运动变换模型很好地适合视频稳定化的应用以对相机操作者的手晃动建模。然而，全局运动变换估计的可替换实施方案可以使用对来自于手晃动的视频抖动的附加或更少方面进行建模的其它模型。

[0044] 实时模式下的一遍运动矢量处理

[0045] 图 9) 例示用于估计视频帧的全局运动变换的一遍运动矢量处理 900 的示例实施方案，其适用于实时处理模式。将运动矢量处理应用于视频序列中场景的每个视频帧的运动矢量，如经由上述基于金字塔块的运动估计所产生的那些。示例实施方案在动作 910 开始于接收视频帧的运动矢量的信息。另外，示例运动矢量处理 900 还接收相对于基准帧中运动矢量位置处的其匹配块的那个的每个块的残差值的信息，其是在运动矢量估计期间计算出的绝对差之和 (SAD) 值。进一步，示例运动矢量处理 900 接收视频帧中原始块的图像内容中的方差的值，其也可以是在运动矢量估计期间计算出的。

[0046] 在动作 920 中，一遍运动矢量处理 900 选择更加可靠的运动矢量的集合，基于其

进行全局运动变换估计。在一个示例实施方案中，一遍运动矢量处理 900 基于三个规则选择运动矢量。首先，图像边界上的运动矢量可能不是可靠的，因此未被选择。例如，在从帧到帧的相机运动后，边界块处的图像内容可能部分地丢失。第二，块的运动补偿后的大残差误差（例如，块的 SAD 值是大的）可能指示不可靠的运动估计。因此，具有大运动补偿残差的块的运动矢量（如在块的运动估计期间计算出的大 SAD 值所指示的）也没有被选择。第三，具有低初始方差的块的运动估计可能不是可靠的，因此也不选择这种块的运动矢量。在可替换的实施方案中，处理 900 可以将运动矢量的集合的选择基于更少或额外的可靠性判据。

[0047] 在动作 930 中，运动矢量处理 900 基于在动作 920 中选择出的运动矢量，计算每个帧的全局运动变换的估计。在一个示例实施方案中，针对上面讨论的相似性全局运动模型使用公式 (1) 估计全局运动变换。可替代地，可以基于另一运动模型估计全局运动变换。为了估计全局运动变换，运动矢量处理 900 使用随机采样一致性 (RANSAC) 方法检测并估计为异常值 (outlier) 的运动矢量。更具体地，在 RANSAC 法中，随机地在动作 920 中选择出的运动矢量的集合中选取两个运动矢量。然后，运动矢量处理使用最小均方误差 (LMSE) 拟合求解等式 (1) 以为两个随机选择的运动矢量确定 s， β， $t_x$， $t_y$参数的值。然后，RANSAC 方法确定也与全局运动变换公式参数的这种解一致的多个其它运动矢量。RANSAC 方法重复其它成对的运动矢量的随机选择，直到求出与所述解一致的最大组的运动矢量。这消除了被视为异常值的或者与对于全局运动变换的参数的一致性 LMSE 解不一致的运动矢量。

[0048] 在动作 940-944 中，运动矢量处理通过使得用于全局运动变换的参数的估计经历缩放、旋转和平移运动中每一个的两个界限，执行误差检测和校正。在流程图中，将两个界限表示为下限 ($T_{1x}$) 和上限 ($T_{2x}$)，其中，$T_{1x} < T_{2X}$，并且 x 表示缩放、旋转或平移。通过关于视频的代表性的大采样的实验，可以在经验上得到这两个界限，所述视频呈现抖动视频运动，并且表示统计概率作为代表抖动视频运动的全局运动。如动作 940-944 中所示，如果全局运动变换参数 (s， β， $t_x$， $t_y$) 小于其下限，则使用如动作 930 中估计出的全局运动变换。否则，如果用于缩放、旋转和平移的全局运动变换超过下限但是仍然保持小于上限，则运动矢量处理将全局运动变换限制到下限，如动作 943 处所示。如果在动作 930 处估计出的全局运动变换超过上限，则将全局运动变换重置。这些动作由此旨在检测将不会是来自晃动手的抖动视频运动的特征的过量全局运动，并且然后从过度补偿校正。

[0049] 运动矢量处理 900 的最后动作 950 将时间平滑应用于全局运动变换。在一个示例实施方案中，对于场景的视频帧的序列的全局运动变换估计由 31 抽头的高斯滤波器滤波，如下面的等式 (2) 所示。

[0050]

$$W_t' = C_{t-15} \times (W_{t-1}W_{t-2}...W_{t-15})^{-1} + C_{t-14} \times (W_{t-1}W_{t-2}...W_{t-14})^{-1} + .... + C_t \times I + ... + C_{t+14}(W_{t+1}W_{t+2}...W_{t+14}) + C_{t+15}(W_{t+1}W_{t+2}...W_{t+15}) \quad (2)$$

[0051] 符号 $(W_{t-1}W_{t-2}\cdots W_{t-15})^{-1}$表示矩阵的逆。在此等式 (2) 中，$W_t'$ 是在时间 t 处平滑之后的全局运动变换，而 $W_{t_i}$是在时间 (t-i) 处平滑之前的全局运动变换。值 $C_{t_i}$是高斯滤波器系数，而 I 是单位矩阵。在实时视频回放或代码转换情境中，这种尺寸的滤波器带来了 15 个画面的延迟。可替换的实施方案可以使用更长或更短的滤波器尺寸，并且可替代地，可以

使用包括自适应滤波器的其它类型滤波器。

[0052] 离线模式下的两遍运动矢量处理

[0053] 可替代地，可以使用两遍运动矢量处理（如，图 10 中所示的示例两遍运动矢量处理 1000）完成全局运动变换的估计，其提供更好质量的视频图像稳定化，但是以大大增加的计算时间为代价。两遍处理通常也给用户带来了更少的便利，因为用户被迫等待，直到视频稳定化处理结束为止以浏览结果。由于其增大的计算负荷，两遍运动矢量处理通常更适用于作为离线模式运行。运动矢量处理的两遍配置允许误差检测与校正部分的更好优化以及运动矢量处理的时间平滑部分。在第一遍中，运动矢量处理可以聚集额外的信息以用于处理的这些部分中，其然后可以应用于第二遍运动矢量处理中。

[0054] 如图 10）中动作 1005 中所例示的，两遍运动矢量处理 1000 在第一遍期间针对视频片段执行场景改变检测。可以使用传统上已知的方法执行场景改变检测。然后，在动作 1010 中，诸如通过应用与图 9）的单遍运动矢量处理 900 的动作 910、920 和 930 相同的动作，两遍运动矢量处理 1000 针对用于第一遍的场景的每个视频帧，产生全局运动变换的估计。在第一遍中，仅聚集全局运动变换的这些估计来产生统计信息，以便更好地优化第二遍中的全局运动变换估计。因此，第一遍中的全局运动估计不直接用在视频图像稳定化处理 200 的图像扭曲 230 中去补偿抖动视频运动。而是，在动作 1020 处，两遍运动矢量处理 1000 跨域视频片段的每个场景上的所有帧，计算最小和最大值和全局运动变换的平移、旋转和缩放参数的概率分布（例如，$s_{min}$，$s_{max}$，$\beta_{min}$，$\beta_{max}$，$t_{xmin}$ $t_{xmax}$，$t_{ymin}$，$t_{ymax}$，$p(s)$，$p(\beta)$，$p(t_x)$ 和 $p(t_y)$）。可替代地，可以在整体上而非每一单独场景对于视频片段聚集概率统计。

[0055] 基于对于每个场景的参数的这些经验概率分布，在动作 1030 处，两遍运动矢量处理 1000 针对全局运动变换参数（s，$\beta$，$t_x$，$t_y$）的每一个确定每个场景中的下限和上限（$T_{1x}$和 $T_{2X}$）。例如，可以将下限和上限选取为对应于某个概率阈值 $p_1$和 $p_2$。在一个示例中，概率阈值 $p_1$和 $p_2$分别为 95% 和 98%。换言之，针对场景在第一遍中观察到的参数的值的 95% 都处于界限 $T_{1x}$内；而观察到的参数值的 98% 都处于上限 $T_{2X}$以下。在可替代的实施方案中，两遍运动矢量处理可以基于其它概率阈值或某些其它基于经验概率的判据来确定界限。另外，可以通过硬界限界定下限和上限。相比于图 9）中的单遍运动矢量处理 900，基于视频的经验概率统计的参数界限（$T_{1x}$和 $T_{2X}$）的这种设置允许应用于全局运动变换的误差校正和检测自适应于视频的特定内容。

[0056] 在动作 1030 处，两遍运动矢量处理 1000 然后执行运动矢量处理的第二遍。对于第二遍，处理 1000 执行如单遍运动矢量处理 900 的动作 910、920、930 和 940-944 中所做的处理。对于此第二遍的动作 940-944 中应用的界限（$T_{1x}$和 $T_{2X}$），该两遍处理使用基于在第一遍中聚集的视频的概率统计所确定出的界限。由于这些界限适用于主题视频的内容，因此两遍运动矢量处理 1000 应当使用这些自适应界限执行更好的质量误差检测和校正。

[0057] 通过也在视频片段上全局地执行时间平滑，两遍运动矢量处理也可以取得更好质量的数字视频图像稳定化（相比于单遍运动矢量处理）。在两遍运动矢量处理 1000 的一个示例实施方案中，处理使用受约束的全局优化执行时间平滑，而不是单遍运动矢量处理的滑动窗办法。这种约束的全局优化的一个示例包括：从平均数目 N 的视频帧的全局运动变换确定对于过平滑的全局运动变换的约束，作为时间平滑的上限，其中 N 可以是范围 [10，15] 中的用于过平滑（over-smoothing）的值。然后，在过平滑约束下进行凸优化。在一个

示例实施方案中，目标函数例如，可以将一个目标函数指定为

[0058]

$$T = \arg\min_{W_f}\left(\sum_{i=1}^{N}\sum_{j=i-N}^{i+N} C_i \times D(W_f, W_{oj})\right) + p\left(\sum_{i=1}^{N} D(W_f, W_{oi})\right) \quad (3)$$

[0059] 在目标函数等式 (3) 中，第一项意味着最终的平滑变换 $W_f$ 应当与原始变换和其相邻变换 $W_{ni}$ 类似，其中在窗口中具有不同的加权系数 $C_i$。第二项意味着最终的平滑变换 $W_f$ 应当与过平滑的 $W_{oi}$ 类似，作为具有权重 p 的约束。值 D 是距离测量度量值，其可以是具有 $W_1$ 和 $W_{oi}$ 的四个图像角落的变换的坐标之间的距离。凸优化可应用于最小化目标函数。

[0060] 利用全局运动变换的图像扭曲

[0061] 再次参照图 2)，数字视频图像稳定化处理 200 最后基于使用单遍或两遍运动矢量处理 220 所产生的全局运动变换执行图像扭曲 230。在 GPU 中进行图像扭曲，可以简单地使用 GPU D3D API 和 / 或定制的双三次顶点 (bi-cubic vertex ) 着色器以从由于抖动视频运动引起的全局运动变换的估计应用相反方向上的图像扭曲。为了在视频帧的边界提供足够的视频内容以允许图像扭曲，在一个示例实施方案中通过边界处的大约 τ % 裁切视频帧，其对于一遍处理模式通常可以为 10%，而针对二遍处理模式中的不同场景可以为可变 ( 自适应 ) 的百分比。这允许通过图像扭曲在视频帧的内容内移动显示端口，其然后形成图像稳定化的视频以用于输出。

[0062] 如从运动矢量处理中的动作 944 可以理解的，当全局运动超过可能只由于晃动手所引起的抖动视频运动的量时，估计的全局运动变换被简单地重置。以此方式，对于超过来自于手晃动的抖动的界限特性的全局视频运动，不应用图像扭曲。而是在对于手晃动没有任何补偿的情况下利用其实际运动播放视频帧。

[0063] 示例计算环境

[0064] 图 11) 图示所描述的实施例、技术和技术方案可以被实施的合适的计算环境 11)00 的一般性示例。例如，计算环境 11)00 可以是图 2) 的数字视频图像稳定化处理 200 运行于其上的计算设备 ( 例如，计算机服务器、个人计算机、膝上型计算机或其它计算设备 ) 之一。

[0065] 计算环境 11)00 并非旨在暗示关于技术的使用或功能的范围的任何限制，这是由于该技术可以在多种多样的通用或专用计算环境下实施。例如，可以用其它的计算机系统配置实施所公开的技术，包括手持式设备、多处理器系统、基于微处理器的或者可编程的消费型电子设备、网络 PC、迷你计算机、大型计算机等。也可以在分布式计算环境中实践所公开的技术，在所述环境中，由通过通信网络链接的远程处理设备执行任务。在分布式计算环境中，程序模块可以位于本地和远程存储器储存设备两者中。

[0066] 参照图 11)，计算环境 11)00 包括至少一个中央处理单元 11)10 和存储器 11)20。在图 11) 中，这种最基本的配置 11)30 包括在虚线内。中央处理单元 11)10 执行计算机可执行指令，并且可以是真实或虚拟的处理器。在多处理系统中，多个处理单元执行计算机可执行指令以增大处理能力，并且这样，多个处理器可以同时运行。存储器 11)20 可以是易失性存储器 ( 如，寄存器、缓存、RAM)、非易失性存储器 ( 例如，ROM、EEPROM、闪存等 )、或者两者的某种组合。存储器 11)20 存储软件 11)80，其例如可实施在此描述的技术。

[0067] 除了中央处理单元 11)10 之外，计算环境可包括其它的处理资源，如数字信号处

理 DSP 或多媒体组件 11)15。DSP 组件 11)15 可以是有利地用于结合图 1) 在上面讨论的通用平台库实施方案的数字视频图像稳定化处理的任何资源。例如，DSP 组件可包括多媒体 DSP ASIC 单元、GPU 着色器单元、多核 CPU、用于 CPU 的高级多媒体指令集，等等。

[0068] 计算环境可以具有另外的特征。例如，计算环境 11)00 包括储存器 11)40、一个或多个输入设备 11)50、一个或多个输出设备 11)60 和一个或多个通信连接 11)70。诸如总线、控制器或网络之类的互连机构（未示出）互连计算环境 11)00 的各组件。典型地，操作系统软件（未示出）提供用于在计算环境 11)00 中执行的其它软件的操作环境，并且协调计算环境 11)00 的各组件的活动。

[0069] 储存器 11)40 可以是可拆除的或者不可拆除的，并且包括磁盘、磁带或盒式磁带、CD-ROM、CD-RW、DVD、或者可用于储存信息并且可以在计算环境 11)00 内访问的任何其它介质。储存器 11)40 储存用于软件 11)80 的指令，其可以实施这里描述的技术。

[0070] 输入设备 11)50 可以是向计算环境 11)00 提供输入的触摸输入设备（如，键盘、辅助键盘、鼠标、笔或滚动球）、声音输入设备、扫描设备或另一设备。对于音频，输入设备 11)50 可以是接受以模拟或数字形式输入的音频的声卡或类似设备、或者将音频采样提供给计算环境 11)00 的 CD-ROM 读取器。输出设备 11)60 可以是显示器、打印机、扬声器、CD- 写入器、或者从计算环境 11)00 提供输出的另一设备。

[0071] 通信连接 11)70 使得能够经由通信介质（例如，连接网络）与另一计算实体通信。通信介质以调制的数据信号传递诸如计算机可执行指令、压缩的图形信息或其它数据之类的信息。

[0072] 计算机可读取介质是在计算环境 11)00 内可从其访问数据和处理器指令的任何可用介质。通过示例而非限制的方式，在所例示的计算环境 11)00 内，计算机可读介质包括存储器 11)20 和 / 或储存器 11)40。如应当容易理解的，术语计算机可读存储介质包括用于储存数据和程序指令的介质，如存储器 11)20 和 / 或储存器 11)40，和单独的没有调制的数据信号。

[0073] 示例云计算网络环境

[0074] 图 12) 和 13) 例示适当的用于云计算的联网环境 1200 的一般性示例，其中可以实践上述数字视频图像稳定化。

[0075] 在示例云计算网络环境 12)00 中，通过网络可访问的计算和储存资源的集合（称为云 12)10)，提供用于视频共享、储存或分发的各种类型的计算服务（例如，视频共享或社交联网站点）。例如，云 12)10 可以包括服务器计算设备（其可以位于中心或分散的位置）的集合，所述计算设备将基于云的服务提供给经由网络（如，因特网）连接的各种类型的用户和设备。

[0076] 在示例环境 12)00 中，云 12)10 为用户计算设备提供服务（如，除其他示例之外的视频储存、视频共享或社交联网服务）。可以通过云计算服务提供商或者通过在线服务的其它提供商，将服务提供在云 1210 中。例如，基于云的服务可以包括视频储存设备、视频共享站点、社交联网站点、或者经由其分发用户源的视频以供连接设备 1320A-N 上的其他浏览的其他服务。

[0077] 用户可以使用各种移动视频捕获设备来记录视频，如便携摄像机、具有视频模式的数字相机、移动电话和手持式计算设备。用户可以直接地（例如，使用电信网络的数据传

输服务）或者通过首先将视频首先传输至本地计算机 1230（如，膝上型计算机、个人计算机或其它网络连接计算设备）将视频上传至云 1210 上的服务。

[0078] 如图 13）中所示，可以稍后将视频从基于云的视频储存器或共享站点下载、流动和 / 或另外回放到其它连接的计算设备，所述其它连接的计算设备可以具有多种屏幕显示尺寸因子 1320A-N。连接设备 1320A 表示具有中间尺寸显示屏幕的设备，诸如在个人计算机、膝上型计算机、平板或其它类似网络连接设备上可用的。

[0079] 连接设备 1320B 表示具有形式因素设计为高便携的显示屏幕（例如，小尺寸屏幕）的设备。例如，连接设备 1320B 可以是移动电话、智能电话、个人数字助理等。

[0080] 连接设备 1320N 表示具有大浏览屏幕的连接设备。例如，连接设备 1320N 可以是电视屏幕（例如，智能电视）或者将视频输出提供给电视机或视频投影仪的另一设备（例如，机顶盒或游戏控制台）、或具有类似视频显示输出的其它设备。

[0081] 在所例示的云 - 计算网络环境 1200 中，可以在视频共享、存储和分发的各阶段，并且依据所期望的使用情境通过各种各样的所描绘的设备来实施和执行数字视频图像稳定化。在一个示例情境中，数字视频图像稳定化以软件方式实施在本地计算机 1230 上，并且在当将视频最初传输至本地计算机或当上传至基于云的服务时加以应用。在另一情境中，数字视频图像稳定化实施在云中，并且随着其被上传并存储在云中时被应用于视频。在另一情境中，数字视频图像稳定化通过云计算服务实施，并且在当视频被播放、传输至或分发至另一连接设备或服务时加以应用。在又一情境中，数字视频图像稳定化通过云计算服务实施，并且在当代码转换视频以用于以适当的分辨率呈现或者以对于要在其上浏览的连接设备而言适当的传输带宽流动时加以应用。在其他情境中，可以在回放时在连接设备上执行数字视频图像稳定化。

[0082] 示例替代方案和组合

[0083] 这里描述的任何方法都可以经由一个或多个计算机可读介质（例如，储存器或其它的有形介质）来执行，所述计算机可读介质包括（例如，具有或存储）用于执行（例如，使得计算设备执行）这样的方法的计算机可执行指令。操作可以是完全自动的、半自动的或者涉及手动干预。

[0084] 已经在详细的说明书和附图中描述和例示了我们的创新的原理，要认识到，在不脱离这种原理的情况下，各种实施例可以在布置和细节上予以修改。应当理解，这里描述的程序、过程或方法并不涉及或者限于任何特定类型的计算环境，除非另外指出。各种类型的通用或专用计算环境可以与根据这里描述的教导的操作一起使用或者执行根据这里描述的教导的操作。以软件示出的实施例的要素可以以硬件加以实施，并且反之亦然。

[0085] 鉴于我们的发明的原理可以应用到的许多可能的实施例，作为我们的发明，我们请求保护可能出现在下列权利要求和其等效物的精神和范围内的所有这样的实施例。

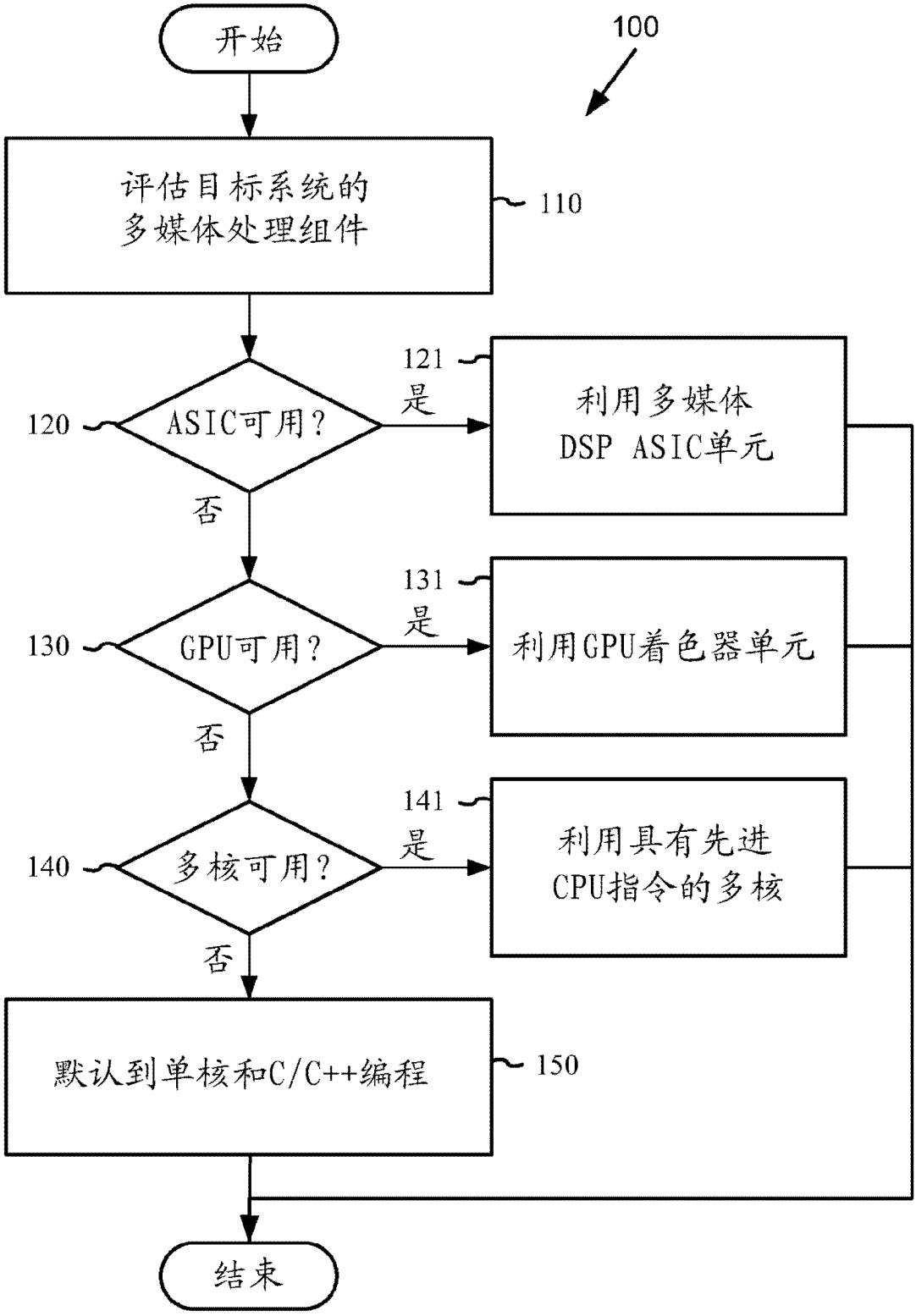

开始
100
评估目标系统的
多媒体处理组件
110
120
ASIC可用？
121
是
利用多媒体
DSP ASIC单元
否
130
GPU可用？
131
是
利用GPU着色器单元
否
140
多核可用？
141
是
利用具有先进
CPU指令的多核
否
默认到单核和C/C++编程
150
结束

图 1

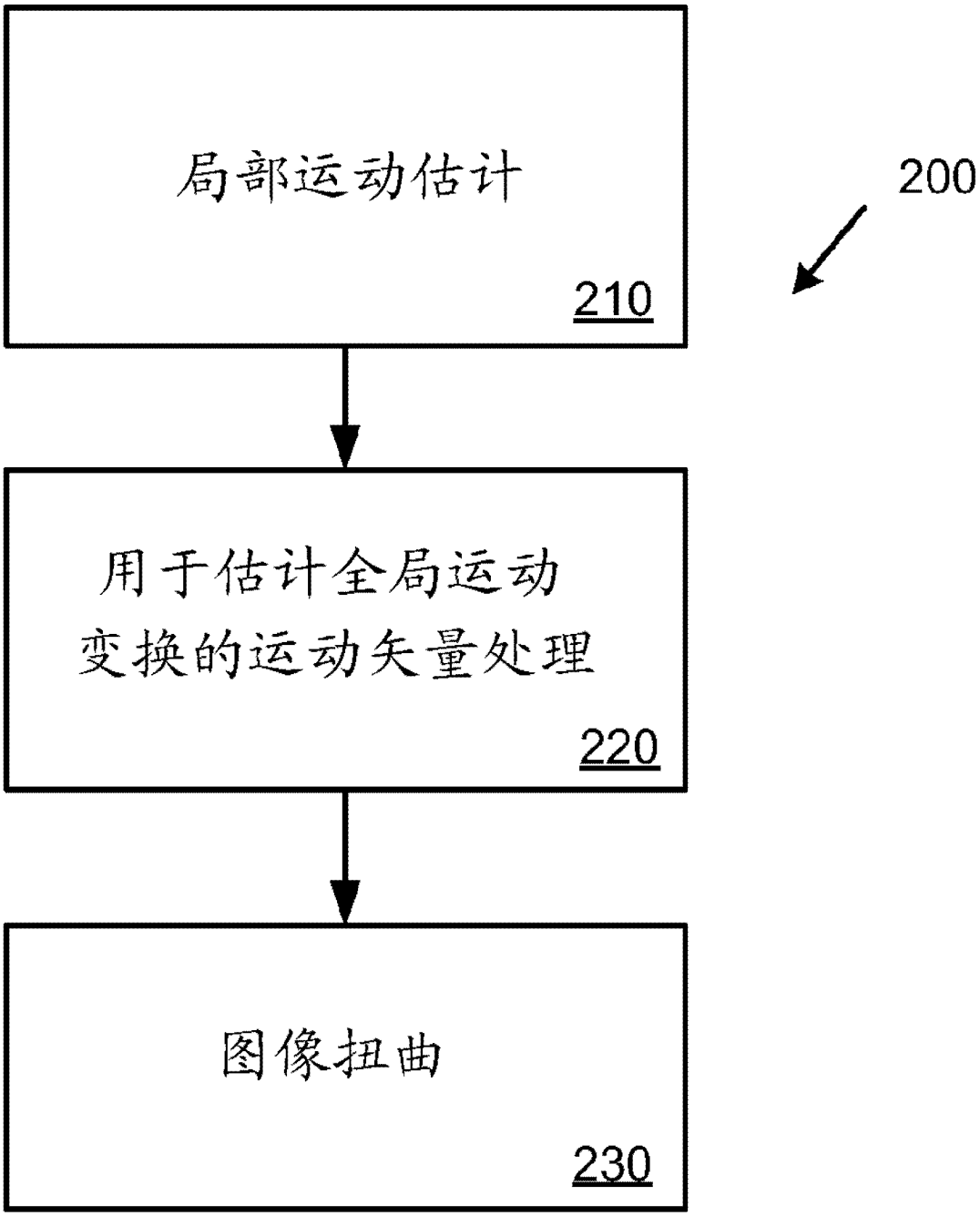



图 2

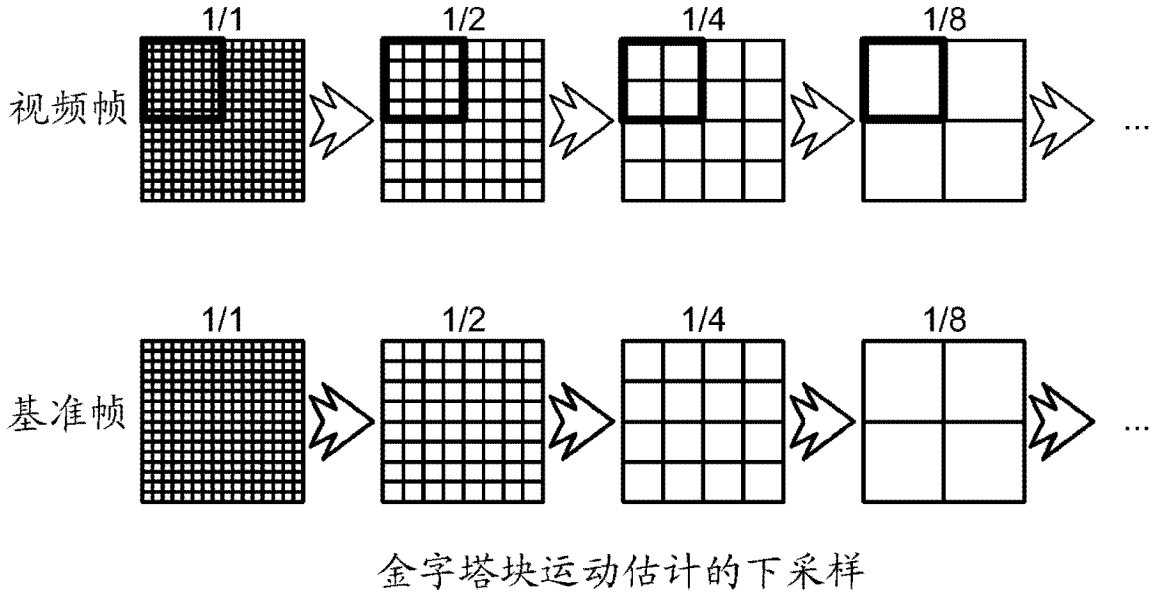

1/1
1/2
1/4
1/8
视频帧
...
1/1
1/2
1/4
1/8
基准帧
...
金字塔块运动估计的下采样

图 3

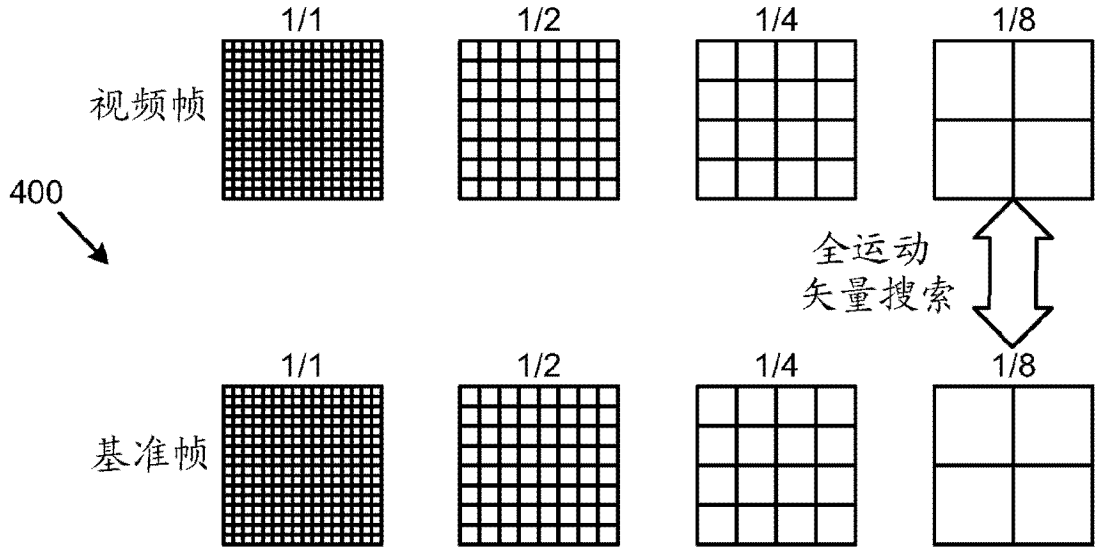

1/1
1/2
1/4
1/8
视频帧
400
全运动
矢量搜索
1/1
1/2
1/4
1/8
基准帧

图 4

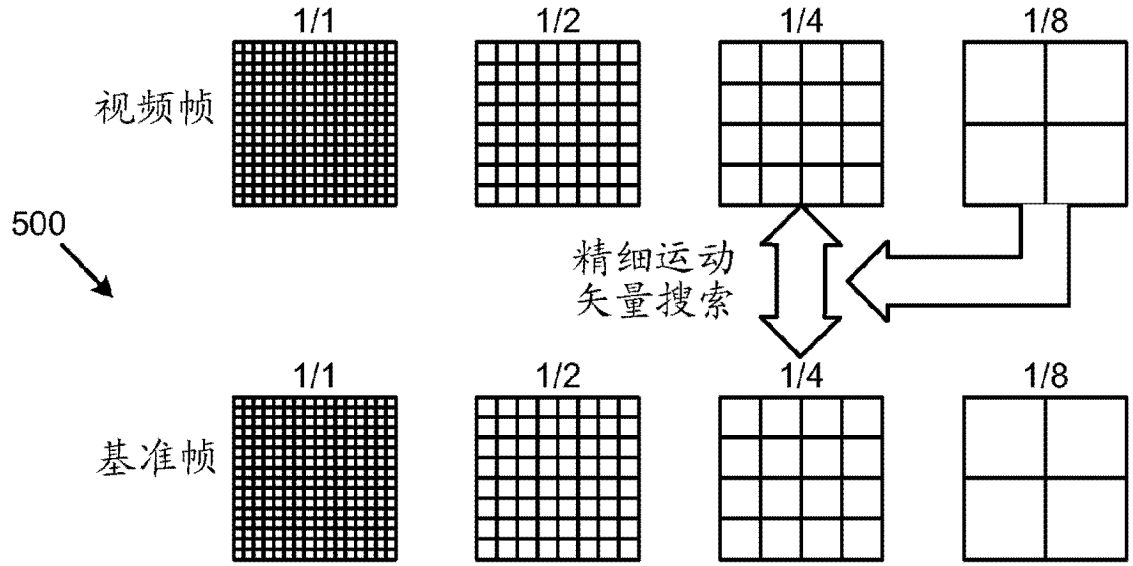

1/1
1/2
1/4
1/8
视频帧
500
精细运动
矢量搜索
1/1
1/2
1/4
1/8
基准帧

图 5

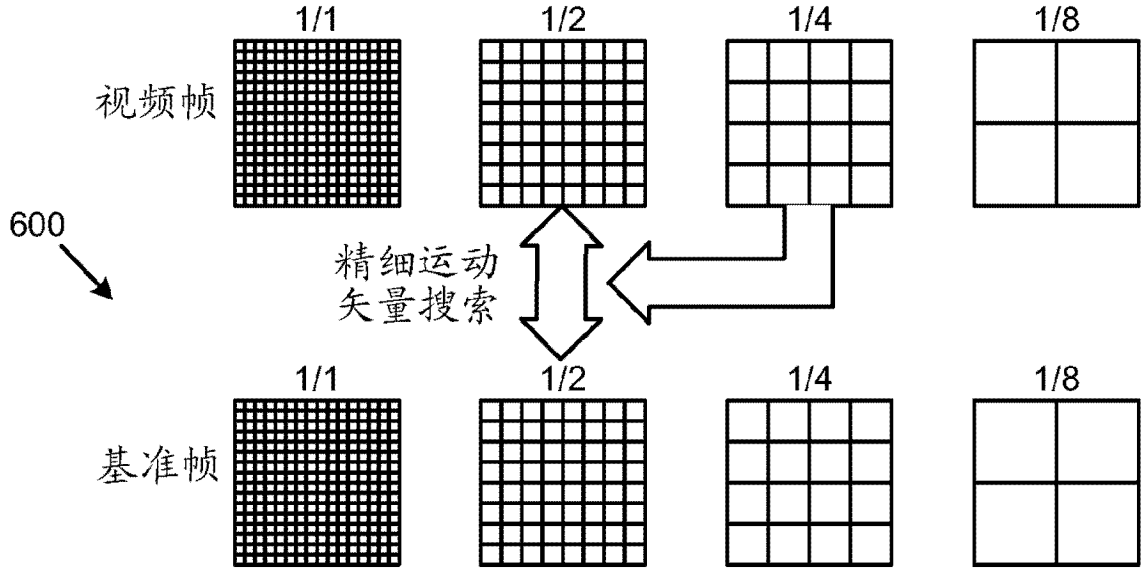

1/1
1/2
1/4
1/8
视频帧
600
精细运动
矢量搜索
1/1
1/2
1/4
1/8
基准帧

图 6

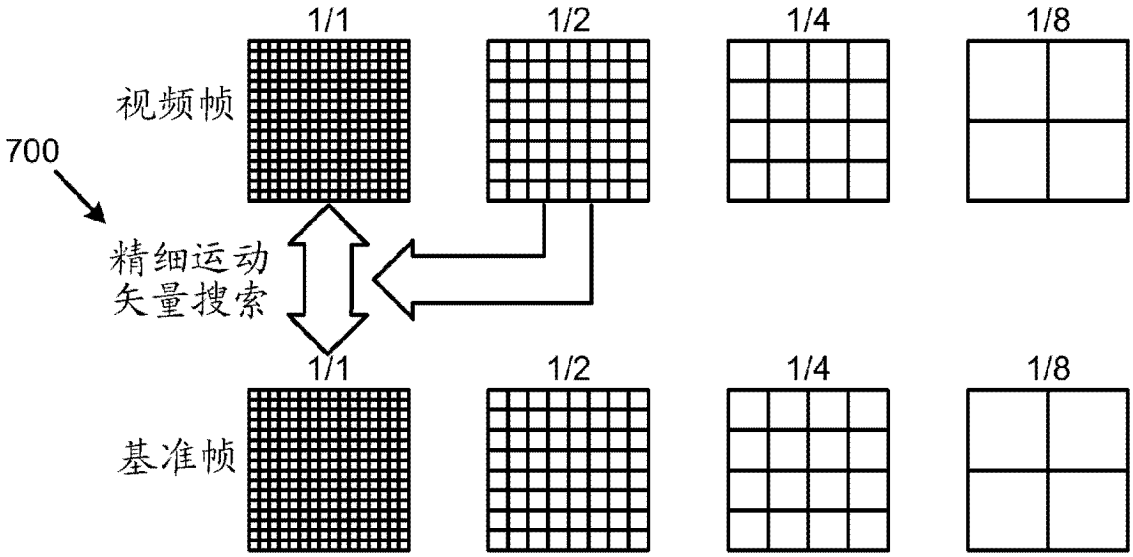

1/1
1/2
1/4
1/8
视频帧
700
精细运动
矢量搜索
1/1
1/2
1/4
1/8
基准帧

图 7

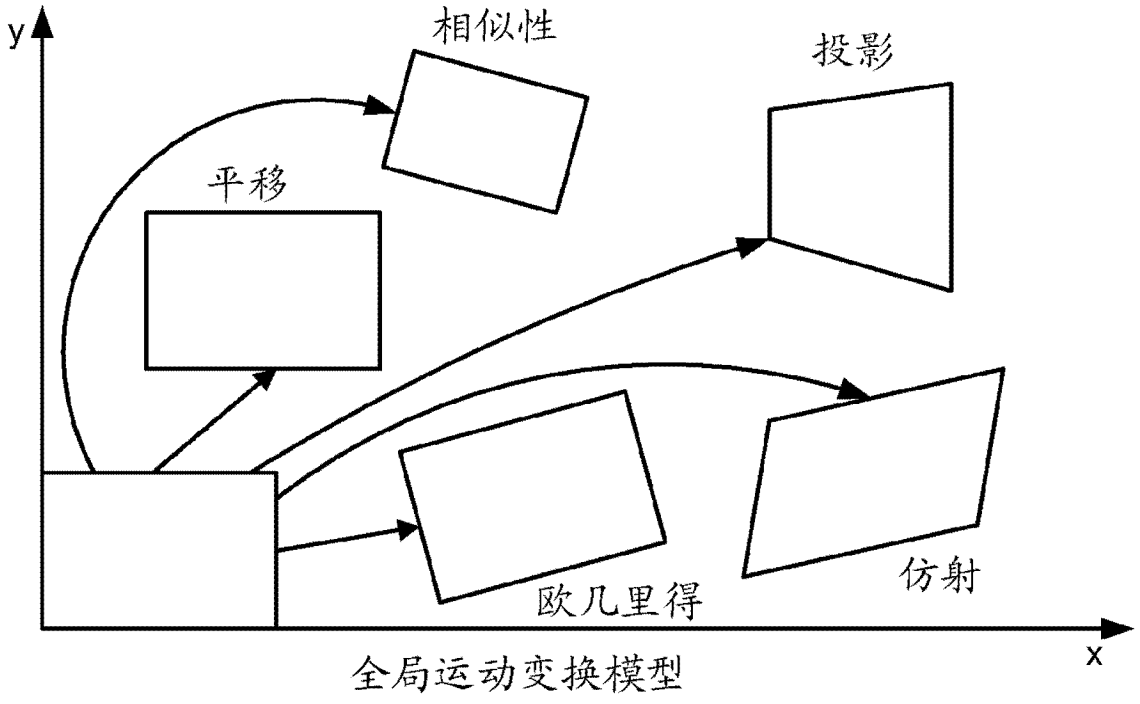

y
相似性
投影
平移
欧几里得
仿射
x
全局运动变换模型

图 8

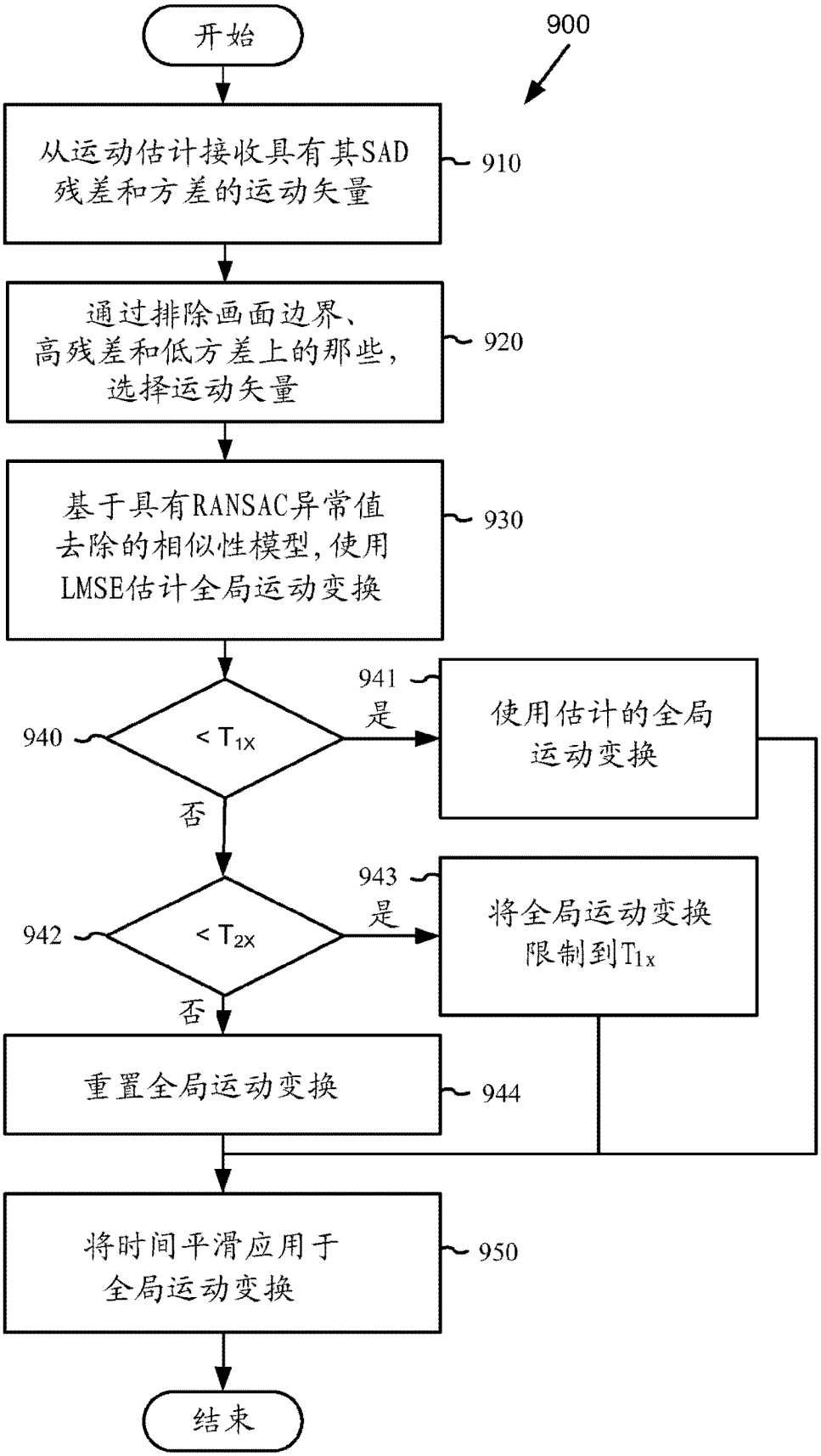

开始
900
从运动估计接收具有其SAD残差和方差的运动矢量
910
通过排除画面边界、高残差和低方差上的那些，选择运动矢量
920
基于具有RANSAC异常值去除的相似性模型，使用LMSE估计全局运动变换
930
940
< T1x
941
是
使用估计的全局运动变换
否
942
< T2x
943
是
将全局运动变换限制到T1x
否
重置全局运动变换
944
将时间平滑应用于全局运动变换
950
结束

图 9

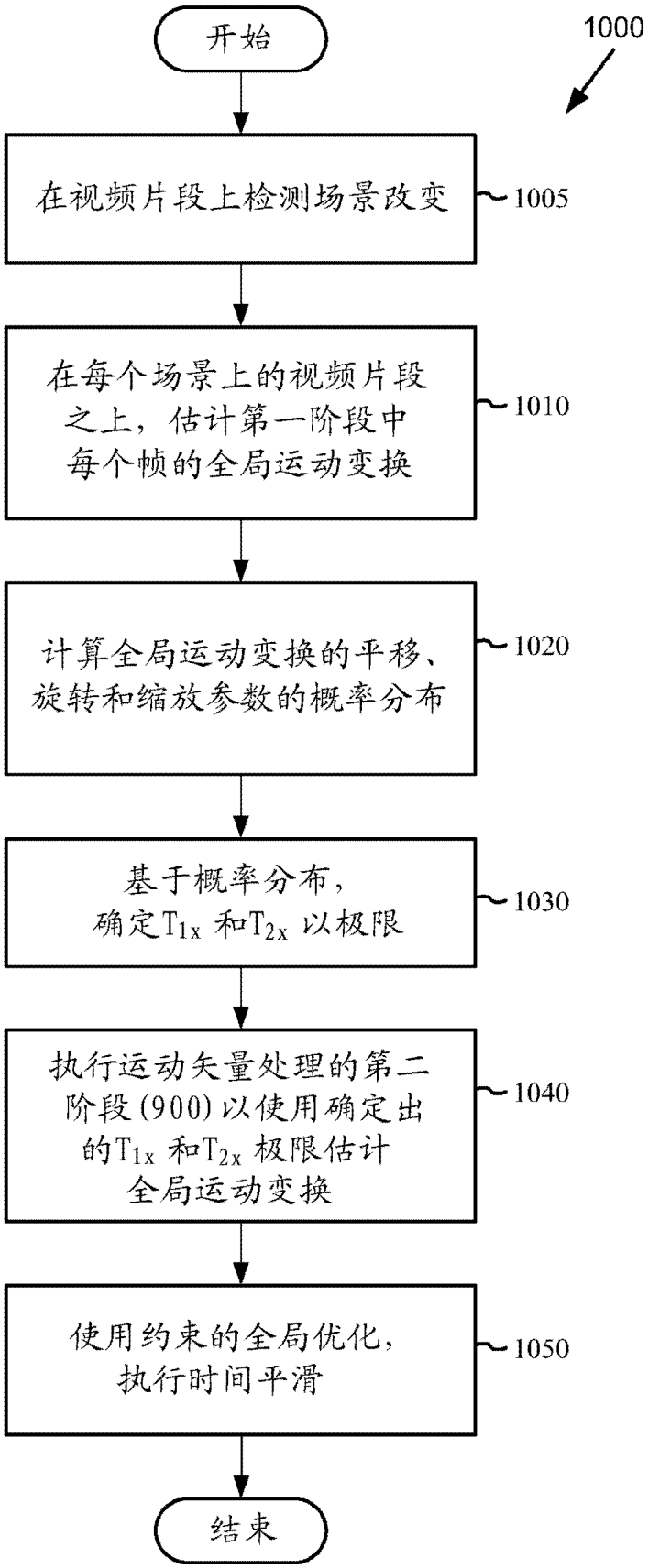

开始
1000
在视频片段上检测场景改变
1005
在每个场景上的视频片段之上，估计第一阶段中每个帧的全局运动变换
1010
计算全局运动变换的平移、旋转和缩放参数的概率分布
1020
基于概率分布，确定$T_{1x}$和$T_{2x}$以极限
1030
执行运动矢量处理的第二阶段(900)以使用确定出的$T_{1x}$和$T_{2x}$极限估计全局运动变换
1040
使用约束的全局优化，执行时间平滑
1050
结束

图 10

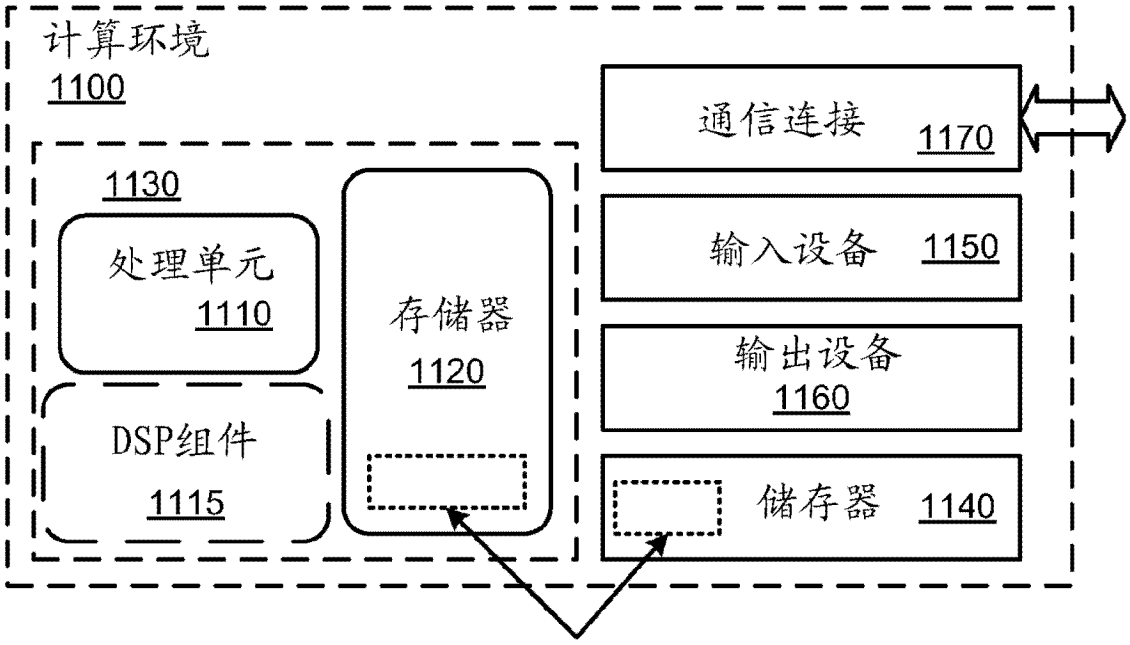

计算环境
1100
通信连接
1170
1130
处理单元
1110
存储器
1120
输入设备
1150
输出设备
1160
DSP组件
1115
储存器
1140
用于所描述的技术的软件1180

图 11

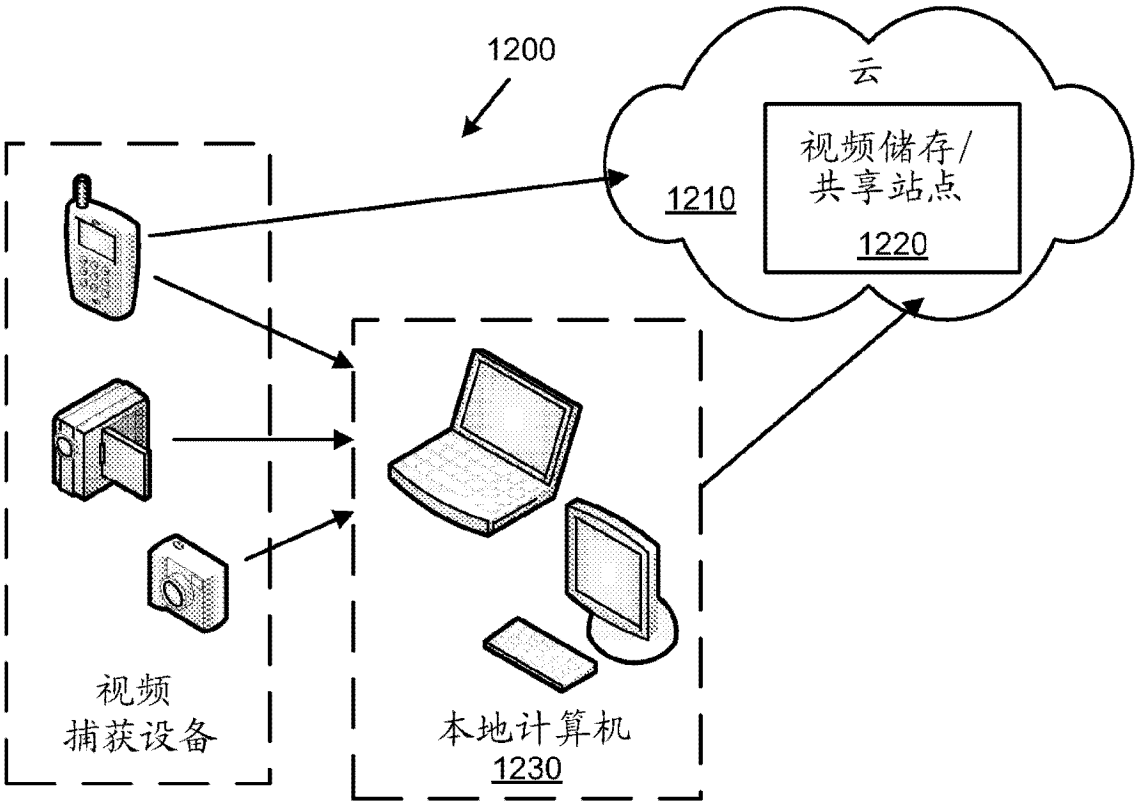



图 12

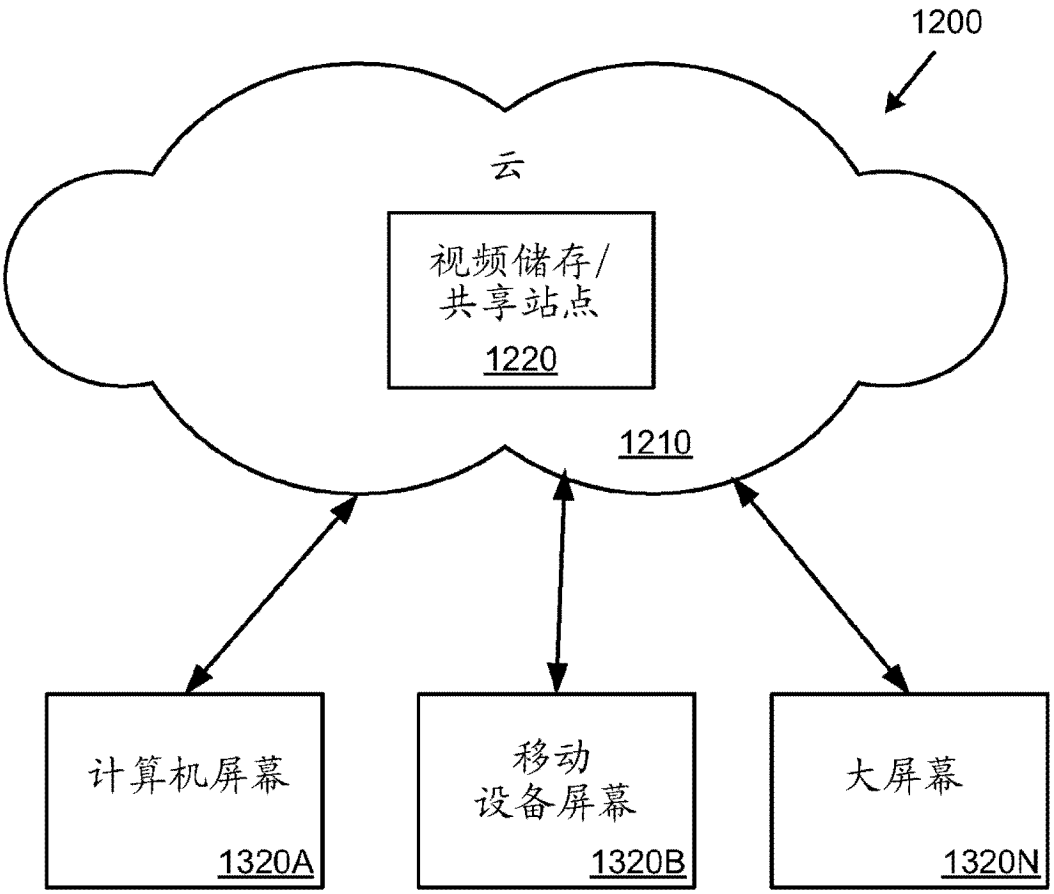

1200
云
视频储存/
共享站点
1220
1210
计算机屏幕
1320A
移动
设备屏幕
1320B
大屏幕
1320N

图 13