[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06K 9/46 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200680037581.X

[43] 公开日 2010 年 1 月 27 日

[11] 公开号 CN 101636748A

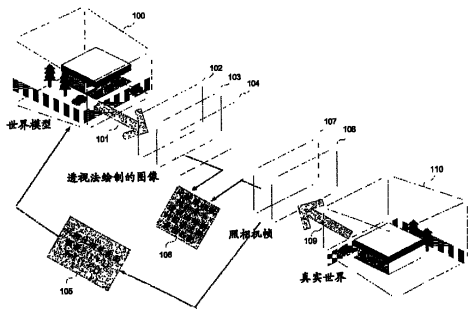
[22] 申请日 2006.9.12
[21] 申请号 200680037581.X
[30] 优先权
[32] 2005. 9.12 [33] US [31] 60/716,139
[86] 国际申请 PCT/IB2006/053244 2006.9.12
[87] 国际公布 WO2007/031947 英 2007.3.22
[85] 进入国家阶段日期 2008.4.9
[71] 申请人 卡洛斯·塔庞
地址 美国华盛顿州
[72] 发明人 卡洛斯·塔庞
[74] 专利代理机构 北京银龙知识产权代理有限公司
代理人 许 静
权利要求书 5 页 说明书 5 页 附图 2 页

[54] 发明名称

模型生成的图形图像到照相机帧的基于帧和象素的匹配

[57] 摘要

本发明允许在不进行实际场景分析和特征识别时进行相机位置的三角测量。 本发明利用在视觉领域的世界的先验的、正确的模型。 使用最新的图形处理单元在图形表面上绘制 3D 模型。 然后在图形表面上绘制的多个候选中搜索来自照相机的每个帧用于最佳匹配。 通过从一个帧到另一个帧计算照相机位置和视角的改变使得绘制的要比较的图像的计数是小的，然后使用这样计算的结果来限制下一个可能的位置和视角以绘制先验的世界模型。 本发明和现有技术相比较的主要优点是将真实世界映射到世界模型。



1. 一种用于实时跟踪校准照相机的位置和视角（自我运动）的方法，包括以下步骤：

创建存在照相机的世界的先验模型；

从照相机中获取每个原始的、未处理的视频帧；

对于每个视频帧，假设获取这些视频帧的一小组可能的位置和视角；

对于每个视频帧，使用图形处理器和来自先验模型的顶点数据来绘制图像，每个假设的位置和视角绘制一个图像；

对于每个视频帧，通过寻找对该视频帧的最佳匹配图像来选出最佳位置和视角。

2. 根据权利要求 1 所述的方法，其中使用低成本的图形处理器来绘制世界的先验模型已经在现有技术中实现，并且已在实际的图形计算机游戏中使用。

3. 根据权利要求 2 所述的方法，其中第一视频帧是来自已知的位置和视角。

4. 根据权利要求 3 所述的方法，其中视频帧和绘制的图像具有相同的分辨率；并且二者都被划分为矩形和正方形区域，这些区域重叠 0 或直到百分之百的象素。

5. 根据权利要求 4 所述的方法，其中通过从两个帧来计算照相机的最大可能运动矢量和视角来限制假设的一组位置和视角的计数，所述两个帧中的一个帧在另一个帧前面；

所述计算包括如下子步骤：

计算当前帧的每个区域的快速傅立叶变换，每个区域独立于另一个区域被处理；

获取得到的快速傅立叶变换矩阵的相位分量并且提出纯相位分量矩阵；在存储器中存储该相位分量矩阵；

利用来自当前的和在前的帧的相位分量矩阵，获取当前照相机帧的每个区域和在前的照相机帧的对应区域之间的相位差；

计算相位差矩阵的反快速傅立叶变换，得到相位相关面；

在每个区域中确定相位相关面的最大值的 2D 位置；所述 2D 位置形成了每个区域的 2D 光流矢量；

从所有区域的光流矢量计算照相机的最可能的 3D 运动矢量和视角。

6. 根据权利要求 5 所述的方法，其中来自 2D 光流矢量的最大可能 3D 运动矢量和视角的计算包括如下子步骤：

确定在参考的世界帧中运动的取向或方向，然后定义一条线，最可能的下一个位置沿着该线分布；

使用在前的速度计算来确定沿着取向线的候选的下一个位置；

从计算的候选位置周围的立方选择的点中选出多个最可能的位置；

使用梯度下降来在立方选择的点中选择最佳的下一个位置。

7. 根据权利要求 5 所述的方法，其中选择对每个视频帧的最佳匹配的绘制图像的方法包括如下子步骤：

在绘制的图像中计算每个区域的快速傅立叶变换；

获得每个区域的快速傅立叶变换矩阵的相位分量；

利用来自当前的帧的相位分量矩阵和来自绘制的图像的相位分量矩阵，获取当前照相机帧的每个区域和绘制的图像的相对应的区域之间的相位差；所述相位差形成相位相关矩阵；

计算照相机帧区域和绘制的图像区域之间的相位相关矩阵的反快速傅立叶变换，得到每个区域的相位相关面；

最佳匹配的绘制图像是具有平方(点积)的光流矢量的最小和的绘制图像，其中对于所有区域求和。

8. 根据权利要求 5 所述的方法，其中选择每个视频帧的最佳匹配绘制图像的方法包括如下子步骤；这在现有技术中称为“直接法”；

对于每个绘制的图像，获得绘制的图像和视频帧之间的每个象素的灰度级的差值；

对于每个区域计算所有所述差值的简单平方和；

选择的绘制的图像是具有与视频帧的平方差的最小和的绘制图像。

9. 根据权利要求 5 所述的方法，其中使用诸如 AutoCAD 的当前可用的

工具构建先验模型。

10. 根据权利要求 5 所述的方法，其中使用现有技术中的方法，通过从存在照相机的世界中预先获取的视频帧的图像处理来构建先验模型。

11. 根据权利要求 5 所述的方法，其中实时构建先验模型是和使用现有技术的方法的运动评估同时发生，但是和运动评估是分开的。

12. 一种用于实时跟踪照相机的位置和视角（自我运动）的设备，包括:

视频照相机及其帧缓存器，以固定帧速率更新帧缓存器的内容;

数字处理装置，用于从一个视频帧到另一个视频帧来计算光流，并且从这样的光流分析中假设多个试验照相机位置和视角;

世界的先验模型;

一个图形处理器或者多个图形处理器，能够在照相机更新帧缓存器的时间片断上成倍增加绘制世界模型;

多个图形表面或者存储绘制的表面的图像缓存器，每个绘制的表面对应于世界模型中的试验位置和视角;

数字处理装置，用于将每个绘制的图像和视频帧缓存器相比较，然后选择最佳匹配的绘制的图像，由此也确定照相机的最精确的瞬态位置和视角。

13. 根据权利要求 12 所述的设备，其中使用低成本的图形处理器来绘制世界的先验模型已经在现有技术中实现，并且已在实际的图形计算机游戏中使用。

14. 根据权利要求 13 所述的设备，其中初始化该设备使得从已知的位置、视角、速度和角速度来开始计算。

15. 根据权利要求 14 所述的设备，其中视频帧和绘制的图像具有相同的分辨率; 并且二者都被划分为矩形和正方形区域，这些区域重叠 0 直到百分之百的象素。

16. 根据权利要求 15 所述的设备，其中通过从两个帧中计算照相机的最大可能运动矢量和视角来限制假设的一组位置和视角的计数，所述两个帧中的一个帧在另一个帧前面; 所述计算包括如下步骤:

计算当前帧的每个区域的快速傅立叶变换，每个区域独立于另一个区域被处理;

获取得到的快速傅立叶变换矩阵的相位分量并且提出纯相位分量矩阵；在存储器中存储该相位分量矩阵；

利用来自当前的和在前的帧的相位分量矩阵，获取当前照相机帧的每个区域和在前的照相机帧的对应区域之间的相位差，所述相位差形成相位相关矩阵；

计算相位相关矩阵的反快速傅立叶变换，得到相位相关面；

在每个区域中确定相位相关面的最大值的 2D 位置；所述 2D 位置形成了每个区域的 2D 光流矢量；

从所有区域的光流矢量计算照相机的最可能的 3D 运动矢量和视角。

17. 根据权利要求 16 所述的设备，配置用于从光流矢量计算最大可能 3D 运动矢量和视角，使用如下计算：

确定在世界的参考帧中活动的取向或方向，然后定义一条线，最可能的下一个位置沿着该线分布；

使用原来的速度计算来确定沿着取向线的候选的下一个位置；

从计算的候选位置周围的立方选择的点中选出多个最可能的位置；

使用梯度下降来在立方选择的点中选择最佳的下一个位置。

18. 根据权利要求 16 所述的方法，配置用来选择对每个视频帧的最佳匹配的绘制的图像，使用如下计算：

在绘制的图像中计算每个区域的快速傅立叶变换；

获得每个区域的快速傅立叶变换矩阵的相位分量；

利用来自当前的帧的相位分量矩阵和来自绘制的图像的相位分量矩阵，获取当前照相机帧的每个区域和绘制的图像的相对应的区域之间的相位差；所述相位差形成相位相关矩阵；

计算照相机帧区域和绘制的图像区域之间的相位相关矩阵的反快速傅立叶变换变换，得到每个区域的相位相关面；

最佳匹配绘制图像是具有平方（点积）的光流矢量的最小和的绘制图像，其中对于所有区域求和。

19. 根据权利要求 16 所述的设备，配置用来使用下面的计算方式选择每个视频帧的最佳匹配绘制图像；这在现有技术中称为“直接法”；

对于每个绘制的图像，获得绘制的图像和视频帧之间的每个象素的灰度级的差值；

对于每个区域计算所有所述差值的简单平方和；

选择的绘制的图像是具有与视频帧平方差的最小和的绘制图像。

20. 根据权利要求 16 所述的设备，其中使用诸如 AutoCAD 的当前可用的工具构建先验模型。

21. 根据权利要求 16 所述的设备，其中使用现有技术中的方法，通过从存在照相机的世界中预先获取的视频帧的图像处理来构建先验模型。

22. 根据权利要求 16 所述的设备，其中实时构建先验模型是和使用现有技术的方法的运动评估同时发生，但是和运动评估是分开的。

23. 一种计算机程序产品，体现在权利要求 6、7、8、9、10 和 11 的任一个权利要求中的方法。

模型生成的图形图像到照相机帧的基于帧和象素的匹配

技术领域

本发明使用当前技术发展水平的计算机图形学来推进计算机视觉领域。

图形引擎，特别是那些实时使用的图形引擎，例如第一人称射击游戏，已经变得十分真实。本发明的基本概念是在图像处理中使用图形引擎：将那些由实时图形引擎生成的图像帧和那些来自照相机的图像帧匹配起来。

背景技术

在视觉或图像处理中有两个不同的任务。一方面有图像分析和特征识别这样困难的任务，另一方面有计算给定输入图像的照相机的3D世界的位置的不那么困难的任务。

在生物视觉中，这两个任务交织在一起使得将它们区分开是困难的。我们通过从我们周围的特征识别和做三角测量来感知我们在世界坐标中的位置。看起来如果我们没有首先识别出我们从中进行三角测量的特征的话，我们不能进行三角测量，并且除非我们可以在我们居住的3D世界的某个位置放置特征，我们不能真正识别。

如果不是所有的视觉系统，那么现有技术中绝大多数的视觉系统尝试在相同的系统中实现这两个任务。例如，参考专利号US5801970包括这两个任务；参考专利号US6704621看起来只包括三角测量，但是它实际上要求道路的识别。

发明内容

如果三角测量任务可以实际上与分析和特征识别任务分开并独立于分析和特征识别任务来进行，那么我们将需要不执行后者任务的系统中的一半的计算资源。通过利用通信处理中的当前的进展，本发明允许照相机位置的三角测量而不进行通常的景物分析和特征识别。本发明利用在视觉领域的世界的先验的、正确的模型。使用最新的图形处理单元在图形表面上绘制3D模型。然后在图形表面上绘制的多个候选中搜索来自照相机的每个帧用于最佳匹配。通过

从一个帧到另一个帧计算照相机位置和视角的改变使得绘制的要比较的图像的计数是小的，然后使用这样计算的结果来限制下一个可能的位置和视角以绘制先验的世界模型。

本发明和现有技术相比较的主要优点是将真实世界映射到世界模型。本发明的一个最适合的应用是机器人编程。以先验的地图引导的机器人以及知道自己在地图中的位置的机器人要远远优于没有被这样引导的机器人。对于导航、引导、寻找路径、避免障碍、瞄准关注点以及其他机器人任务来说是卓越的。

附图说明

图 1 是示出在真实世界中照相机的活动是如何在 3D 模型的世界中被追踪的本发明的实施例的图表。

图 2 是分成区域的透视图表面或者照相机帧的示例。

图 3 是下面描述的算法的高层流程图。

具体实施方式

图 1 中示出了本发明的优选实施例的图表。使用当前可用的先进的图形处理器 101 在绘制的图像 102、103 和 104 上绘制（render）了世界 100 的先验的模型。该模型是真实世界 110 的正确的但是不必要是完整的模型。本发明的目的在于分别追踪在时间 t 和 t + 1 产生帧 107 和 108 的照相机 309 的位置和视角。帧 107 和 108 用作主要实时输入，输入至设备。使用现有技术水平的方法从帧 107 和 108 计算光流矢量。从这些光流矢量根据现有技术以对噪声和异常值具有鲁棒性（robust）的方式可以得出正确的取向和照相机视角。然后在有当前的取向定义的线上的点周围，假设在距当前的位置由当前的速度决定的距离的地方为下一个可能的位置（105）。通过图形处理器或者处理器 101 在 N 个候选图像 102、103 和 104 中绘制可能的候选位置 N。然后将每个绘制的图像和当前的照相机帧进行比较，并且选出最匹配的图像（106）。从选出的图像，也可以从候选位置中选出照相机的最精确的位置，瞬态速度、视角以及角速度。

使用图 3 所示的流程图中的下面的步骤在本发明中实现动态的，帧与帧的三角测量（追踪）。在下面的步骤描述中，对于每个来自照相机的视频帧，具有假设的一批可能的帧，其由图形处理器绘制来进行比较。在本发明中，这样的比较最耗费计算能力。视频帧在垂直和水平分辨率上都等于绘制的图像。如

图 2 所示，每个帧和每个绘制的图像被分成多个矩形区域，其可以相互之间重叠多个象素。

1. 以来自照相机的一帧开始，当获得该帧时，已知在瞬时时间 ‘t’ 照相机的绝对世界位置 P(t)、视角 V(t)、零速度 u(t) = 0 以及零角速度 w(t) = 0。计算这一帧的所有区域（Cs）的离散快速傅立叶变换（FFT），并且提取变换的相位分量，在区域 ‘a’ 时间 ‘t’ 的 PFC（a，t）。

2. 取得下一帧。计算所有 PFC（a，t + 1），在区域 a 时间 t + 1 中的 FFT 的相位分量。

3. 计算 PFC（a，t）和 PFC（a，t + 1）之间的相位差，并且为了获得相位相关面对于相位差矩阵执行反 FFT 变换。如果照相机从 ‘t’ 到 ‘t + 1’ 既没有摇动也没有移动，那么每个区域的相位相关面将指示在区域 ‘a’ 的中心的最大值。如果照相机移动或者摇动了，那么最大值将在除了每个区域的中心之外的某处产生。对于每个区域 OP（a，t + 1）计算光流矢量，其被定义为在相位相关面从中心到最大值点的偏差。（如果在布景的区域中存在移动的物体，每个移动的物体将引起相位相关面上的额外的峰值，但是只要来自随后被比较的帧的两个区域由例如建筑物或者墙壁或者地面的静态物体统治，那么那些其他的峰值将低于对应于照相机位置和/或视角改变的峰值。）

4. 从所有这样的 OP（a，t + 1），并且使用绝对世界位置 P(t)、视角 V(t)、当前速度 u(t)以及当前角速度 w(t)来计算在时间 t + 1 的所有可能的绝对照相机位置（矢量 Pi（t + 1））和视角（单位矢量 Vi（t + 1））的范围。Pi 可以被选择位于活动线（瞬态取向）中，如在由 The MIT Press 1986 年出版的作者为 B.K.P. Horn 的参考书 “Robot Vison” 的 17 章中详述的那样，Pi 可以从 OP（a，t + 1）中容易地确定。

5. 假设可能的照相机位置 Pi（t + 1）的小的数目（例如 N）以及视角 Vi（t + 1）使用先验模型来绘制。这将导致 N 个图像透视图 Mi（a，t + 1）。计算每个 Mi（a，t + 1）的 FFT，并且提取转换的相位分量，PFMi（a，t + 1）。

6. 在 t + 1 时刻对于照相机帧的最佳匹配是考虑了所有的区域 Mi 的每个区域的 PFMi（a，t + 1）和 PFC（a，t + 1）的相位偏差导致了反 FFT 变换，其是具有最大最接近的中心的 2D 图形。从这里也可以选出最佳的可能的位置

P（t+1）和视角 V（t+1）。然后将瞬态速度确定为 u（t+1）=P（t+1）-P（t），同时瞬态角速度为 w（t+1）=V（t+1）-V（t）。

7. 通过将 P（t+1）复制到 P（t）、V（t+1）复制到 V（t）、u（t+1）复制到 u（t）、w（t+1）复制到 w（t）以及 PFC（a，t+1）复制到 PFC（a，t）来丢弃原来的时间 t 的计算和帧，并且使得 t+1 成为当前时间。返回步骤 2。

只要照相机的视角域由静态实体（相当于世界坐标静态，图像的更少的区域由活动实体占据）统治，那么就可能动态三角测量或者追踪。只要照相机帧以及由此区域由静态实体统治，在相位相关面上的峰值对应于照相机运动。这在现有技术中是公知的，如在 1987 年由 British Broadcasting Corporation (BBC) 公布的作者为 G.A.Thomas 的名称为“Television Motion Measurement for DATV and other Applications”的参考文章中详述的。

可选实施例

在本发明可选实施例中，步骤 5 和 6 的计算花费在 K 个帧中分期偿还，并且结果的校正延续至将来的帧。例如，如果对每 5 个照相机帧（K=5）选择参考值，那么第一帧是参考帧，并且在从第一帧抽样到第五帧抽样（t+1 到 t+5）的时间间隔中进行步骤 5 和 6。同时，对于所有抽样，对于所有抽样帧的 P 和 V 使用未校准的值来执行所有其他步骤（步骤 1 到 4 以及 7）。当在第五帧最后选定了第一帧的最佳匹配，应用误差校正。相同的误差校正可以被应用至所有五个 P 和 V 的值，以及由于在 t+5 时所有原来的 P 和 V 的值已经被丢弃，仅仅需要校正 P（t+5）和 V（t+5）。

在本发明的另一个实施例中，步骤 5 和 6 的计算花费通过使用多个低成本游戏图形处理器来解决，每个处理器用于每个假设的照相机位置。

在本发明的另一个实施例中，取代在步骤 5 和 6 中计算照相机帧和绘制的图像之间的相位相关面，可以计算亮度值的差值的平方和（在现有技术中称为“直接法”）。最佳匹配是绘制的图像和最小平方和。

上面描述的是本发明的优选实施例。然而，也可能以除了上面描述的优选实施例中的形式的其他特定形式体现本发明。例如，替代正方形和矩形区域‘a’，可以使用圆形区域。

本发明的示例应用是追踪照相机的位置和视角。然而，本领域的普通技术

人员可以理解和认识到根据本发明的设备和操作方法可以应用至其中物体位置的确定、导航和引导是必须的任意场景。优选实施例仅是示例性的，并且不应该被认为在任意方式是限制性的。本发明的范围由附带的权利要求给出，而不是由上面的说明书给出，并且在此包括在权利要求的精神范围内的所有变型和等效。

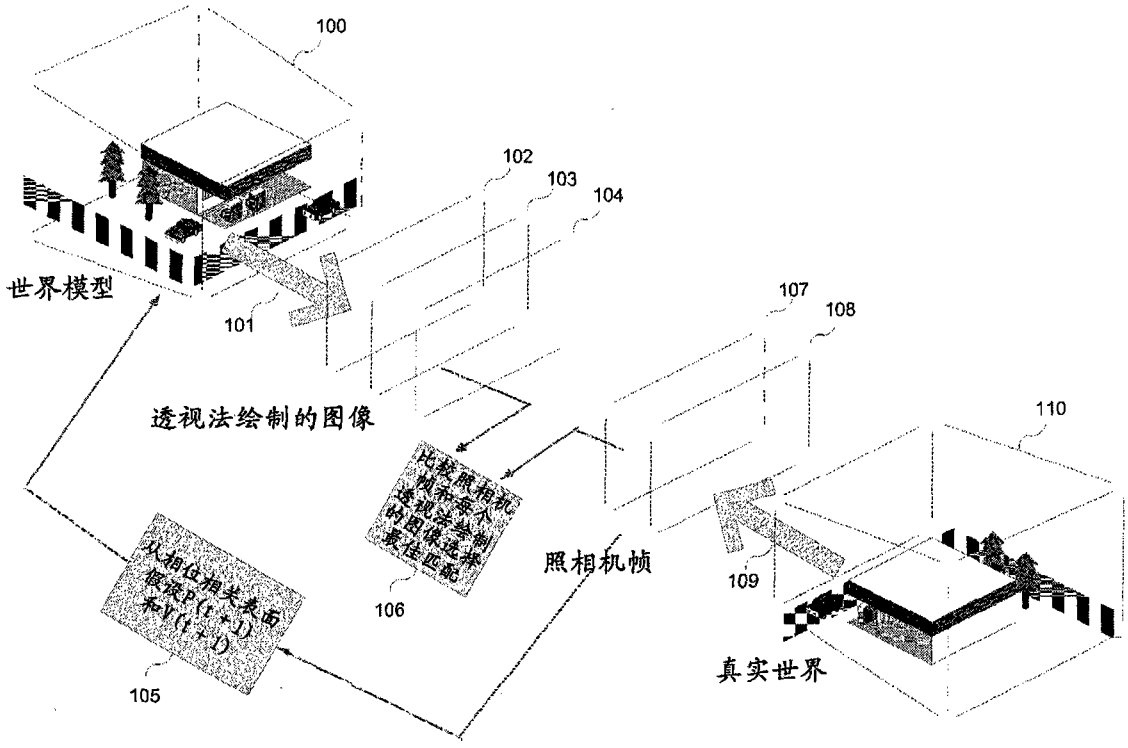


图 1

区域	区域	区域	区域
区域	区域	区域	区域
区域	区域	区域	区域

图 2

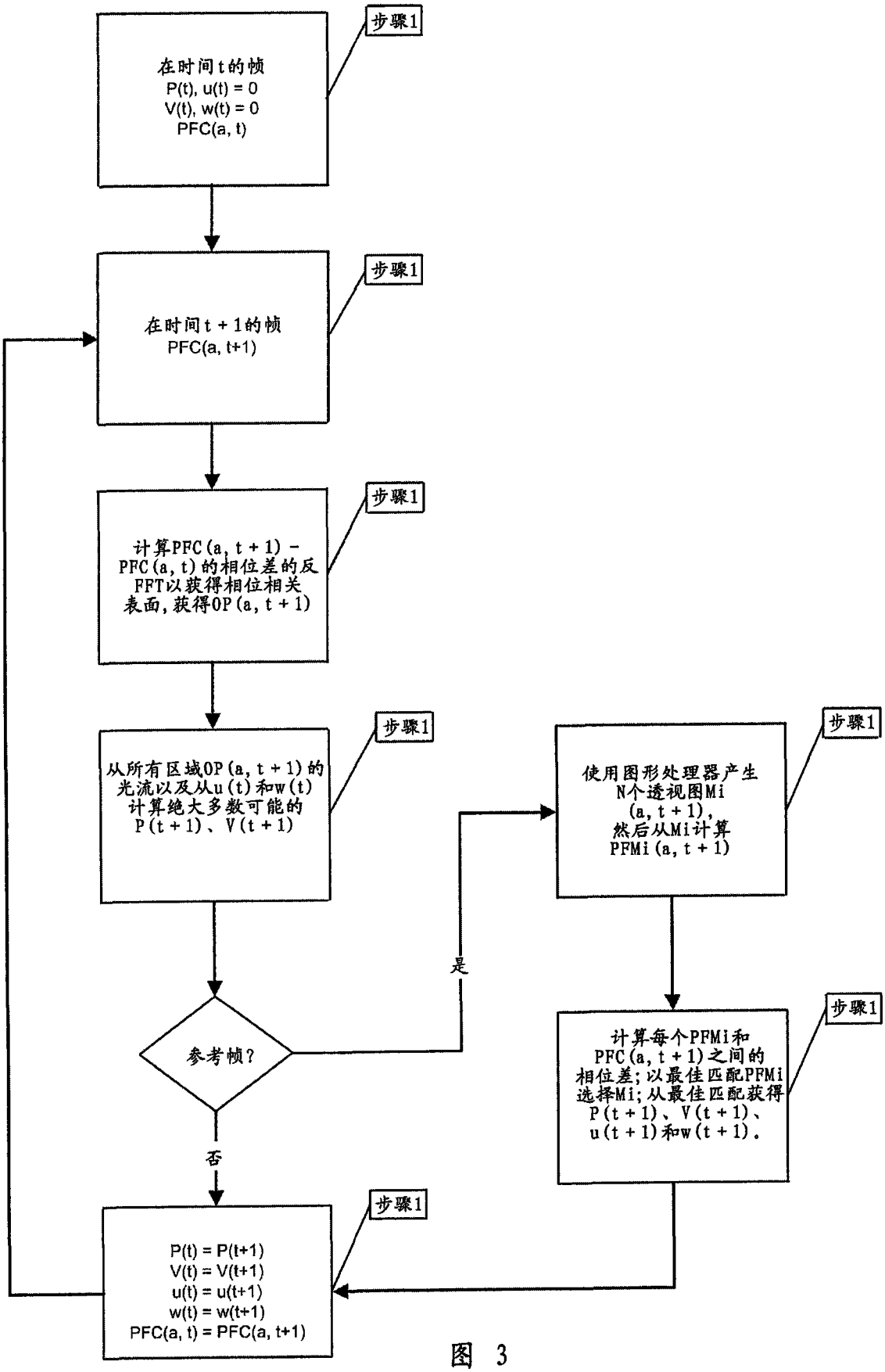

在时间t的帧
P(t), u(t) = 0
V(t), w(t) = 0
PFC(a, t)
步骤1
在时间t + 1的帧
PFC(a, t+1)
步骤1
计算PFC(a, t + 1) - PFC(a, t)的相位差的反FFT以获得相位相关表面，获得OP(a, t + 1)
步骤1
从所有区域OP(a, t + 1)的光流以及从u(t)和w(t)计算绝大多数可能的P(t + 1)、V(t + 1)
步骤1
参考帧？
是
否
使用图形处理器产生N个透视图Mi (a, t + 1)，然后从Mi计算PFMi(a, t + 1)
步骤1
计算每个PFMi和PFC(a, t + 1)之间的相位差；以最佳匹配PFMi选择Mi；从最佳匹配获得P(t + 1)、V(t + 1)、u(t + 1)和w(t + 1)。
步骤1
P(t) = P(t+1)
V(t) = V(t+1)
u(t) = u(t+1)
w(t) = w(t+1)
PFC(a, t) = PFC(a, t+1)
步骤1

图 3