(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114972412 A
(43)申请公布日 2022.08.30

(21)申请号 202210515831.7

(22)申请日 2022.05.12

(71)申请人 深圳锐视智芯科技有限公司
地址 518000 广东省深圳市南山区桃源街道福光社区留仙大道3370号南山智园崇文园区3号楼2001

(72)发明人 沈益冉 郭旭成 杜博闻

(74)专利代理机构 深圳市恒申知识产权事务所(普通合伙) 44312
专利代理师 廖厚琪

(51)Int.Cl.
G06T 7/207(2017.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06V 10/80(2022.01)
G06V 10/82(2022.01)

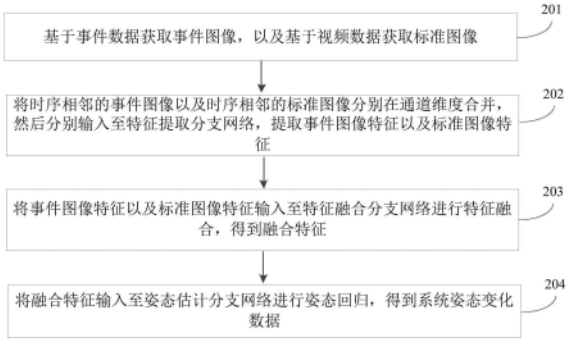
权利要求书3页 说明书10页 附图4页

(54)发明名称

一种姿态估计方法、装置、系统及可读存储介质

(57)摘要

本申请提供了一种姿态估计方法、装置、系统及可读存储介质，该方法包括：基于事件数据获取事件图像及基于视频数据获取标准图像；将相邻事件图像及相邻标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征及标准图像特征；将事件图像特征及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。通过本申请方案的实施，结合事件相机和标准相机两者收集的数据共同估计系统在运动过程中的姿态变化，当系统处于高速和高动态范围的场景下，可提供更多运动信息用以估计系统自身姿态变化，提高了姿态估计的精度和稳定性。



CN 114972412 A

1.一种姿态估计方法,其特征在于,应用于姿态估计系统,所述姿态估计系统配置有事件相机以及标准相机,所述事件相机用于采集事件数据,所述标准相机用于采集视频数据,所述姿态估计方法包括:

基于所述事件数据获取事件图像,以及基于所述视频数据获取标准图像;

将时序相邻的所述事件图像以及时序相邻的所述标准图像分别在通道维度合并,然后分别输入至特征提取分支网络,提取事件图像特征以及标准图像特征;

将所述事件图像特征以及所述标准图像特征输入至特征融合分支网络进行特征融合,得到融合特征;

将所述融合特征输入至姿态估计分支网络进行姿态回归,得到系统姿态变化数据。

2.根据权利要求1所述的姿态估计方法,其特征在于,所述基于所述事件数据获取事件图像的步骤,包括:

将所述事件数据根据事件划分为多个体素;其中,事件数据表示为$e=\{x_i,y_i,t_i,p_i\}$,$i\in 0,1,\cdots,n-1$,x_i,y_i表示像素坐标位置,t_i表示时间戳,p_i表示事件极性;

基于预设转换公式将各所述体素转换为事件图像;所述转换公式表示为:

$$t_i^* = (B-1)(t_i - t_0)/(t_{n-1} - t_0)$$

$$V(x,y,t) = \sum_i p_i \max(0,1-|x-x_i|)\max(0,1-|y-y_i|)\max(0,1-|t-t_i^*|)$$

其中,V(x,y,t)表示所述体素,B表示所述体素的数量。

3.根据权利要求1所述的姿态估计方法,其特征在于,所述将时序相邻的所述事件图像以及时序相邻的所述标准图像分别在通道维度合并的步骤之前,还包括:

针对所有所述事件图像以及所述标准图像分别计算均值和方差,对所有所述事件图像以及所述标准图像进行归一化处理以及标准化处理;

将各所述事件图像以及各所述标准图像分别复制三份后置于色彩三通道中,以预设维度的张量对所述事件图像以及所述标准图像进行表达。

4.根据权利要求1所述的姿态估计方法,其特征在于,所述特征融合分支网络包括自注意力网络,所述自注意力网络包括第一自注意力网络以及第二自注意力网络;

所述将所述事件图像特征以及所述标准图像特征输入至特征融合分支网络进行特征融合,得到融合特征的步骤,包括:

将所述事件图像特征输入至所述特征融合分支网络的所述第一自注意力网络,获取所述事件图像特征相应的第一掩码,以及将所述标准图像特征输入至所述第二自注意力网络,获取所述标准图像特征相应的第二掩码;

通过所述第一自注意力网络对所述事件图像特征与所述第一掩码进行乘法运算,得到第一粗过滤特征,以及通过所述第二自注意力网络对所述标准图像特征与所述第二掩码进行乘法运算,得到第二粗过滤特征;

基于所述第一粗过滤特征以及所述第二粗过滤特征进行特征融合,得到融合特征。

5.根据权利要求4所述的姿态估计方法,其特征在于,所述获取所述事件图像特征相应的第一掩码的步骤,包括:

通过预设第一掩码计算公式获取所述事件图像特征相应的第一掩码;所述第一掩码计

算公式表示为：$\tilde{e}_i = Sigmoid((W_i e_i)^T W_j e_j) g(e_i)$，其中，$e_i$和$e_j$分别表示所述事件图像特征的一个通道的特征向量，$W_i$和$W_j$分别表示将$e_i$和$e_j$投影到嵌入空间的可学习权重矩阵，g()表示将特征向量投影到新嵌入空间的函数；

所述获取所述标准图像特征相应的第二掩码的步骤，包括：

通过预设第二掩码计算公式获取所述标准图像特征相应的第二掩码；所述第二掩码计算公式表示为：$\tilde{f}_i = Sigmoid((W_i f_i)^T W_j f_j) g(f_i)$，其中，$f_i$和$f_j$分别表示所述标准图像特征的一个通道的特征向量。

6.根据权利要求4所述的姿态估计方法，其特征在于，所述特征融合分支网络还包括交叉注意力网络；

所述基于所述第一粗过滤特征以及所述第二粗过滤特征进行特征融合，得到融合特征的步骤，包括：

将所述第一粗过滤特征输入至所述交叉注意力网络，获取所述标准图像特征相应的第三掩码，以及将所述第二粗过滤特征输入至所述交叉注意力网络，获取所述事件图像特征相应的第四掩码；

通过所述交叉注意力网络对所述第一粗过滤特征与所述第四掩码进行乘法运算，得到第一细过滤特征，以及对所述第二粗过滤特征与所述第三掩码进行乘法运算，得到第二细过滤特征；

对所述第一细过滤特征以及所述第二细过滤特征进行特征融合，得到融合特征。

7.根据权利要求6所述的姿态估计方法，其特征在于，所述获取所述标准图像特征相应的第三掩码的步骤，包括：

通过预设第三掩码计算公式获取所述标准图像特征相应的第三掩码；所述第三掩码计算公式表示为：$m_{e\to f}=\mathrm{Sigmoid}((W_{e\to f}e_i)^T W_{e\to f}e_j)$；

所述获取所述事件图像特征相应的第四掩码的步骤，包括：

通过预设第四掩码计算公式获取所述事件图像特征相应的第四掩码；所述第四掩码计算公式表示为：$m_{f\to e}=\mathrm{Sigmoid}((W_{f\to e}e_i)^T W_{f\to e}e_j)$。

8.根据权利要求1至7中任意一项所述的姿态估计方法，其特征在于，所述姿态估计分支网络包括长短期记忆单元以及全连接层；

所述将所述融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据的步骤，包括：

将所述融合特征输入至所述姿态估计分支网络的所述长短期记忆单元进行时序上的建模，得到估计数据；

将所述估计数据输入至所述全连接层进行处理，输出系统姿态变化数据。

9.一种姿态估计装置，其特征在于，应用于姿态估计系统，所述姿态估计系统配置有事件相机以及标准相机，所述事件相机用于采集事件数据，所述标准相机用于采集视频数据，所述姿态估计装置包括：

获取模块，用于基于所述事件数据获取事件图像，以及基于所述视频数据获取标准图像；

提取模块，用于将时序相邻的所述事件图像以及时序相邻的所述标准图像分别在通道

维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；

融合模块，用于将所述事件图像特征以及所述标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；

估计模块，用于将所述融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。

10.一种姿态估计系统，其特征在于，包括事件相机、标准相机、存储器及处理器，其中：

所述事件相机用于采集事件数据，所述标准相机用于采集视频数据；

所述处理器用于执行存储在所述存储器上的计算机程序；

所述处理器执行所述计算机程序时，实现权利要求1至8中任意一项所述方法中的步骤。

11.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时，实现权利要求1至8中的任意一项所述方法中的步骤。

一种姿态估计方法、装置、系统及可读存储介质

技术领域

[0001] 本申请涉及图像处理技术领域，尤其涉及一种姿态估计方法、装置、系统及可读存储介质。

背景技术

[0002] 对于智能车辆的自主导航来说，车辆在运动过程中的自定位能力非常重要，通过融合各种传感器的数据来估计车辆的位置被定义为里程计。早期的车辆定位系统通常采用轮速编码器来计算车辆里程，然而这种方法存在严重的累积误差，而且轮胎打滑等因素会造成更加不可预测的后果。随着计算机视觉技术的发展，视觉传感器越来越多的被用来进行车辆的定位与运动估计中。视觉传感器不仅能提供丰富的感知信息，还具有低成本、小体积等特点。研究人员通常把通过视觉来获取相机姿态的问题称为视觉里程计(VO,Visual Odometry)，在最近二十多年时间里，VO已经广泛应用于各类机器人的导航定位中。

[0003] 然而，目前主流的VO方法主要是基于图像中物体的几何特性来估计相机的姿态，所以要求图像含有大量稳定的纹理特征，一旦场景中出现遮挡物或在雾天雨天取景，在没有其他传感器(IMU、激光、雷达等)的情况下，几何法的求解精度都会受到很严重的干扰。

发明内容

[0004] 本申请实施例提供了一种姿态估计方法、装置、系统及可读存储介质，至少能够解决相关技术基于图像中物体的几何特性来估计相机的姿态，所导致的姿态估计精度较低的问题。

[0005] 本申请实施例第一方面提供了一种姿态估计方法，应用于姿态估计系统，所述姿态估计系统配置有事件相机以及标准相机，所述事件相机用于采集事件数据，所述标准相机用于采集视频数据，所述姿态估计方法包括：

[0006] 基于所述事件数据获取事件图像，以及基于所述视频数据获取标准图像；

[0007] 将时序相邻的所述事件图像以及时序相邻的所述标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；

[0008] 将所述事件图像特征以及所述标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；

[0009] 将所述融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。

[0010] 本申请实施例第二方面提供了一种姿态估计装置，应用于姿态估计系统，所述姿态估计系统配置有事件相机以及标准相机，所述事件相机用于采集事件数据，所述标准相机用于采集视频数据，所述姿态估计装置包括：

[0011] 获取模块，用于基于所述事件数据获取事件图像，以及基于所述视频数据获取标准图像；

[0012] 提取模块，用于将时序相邻的所述事件图像以及时序相邻的所述标准图像分别在

通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；

[0013] 融合模块，用于将所述事件图像特征以及所述标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；

[0014] 估计模块，用于将所述融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。

[0015] 本申请实施例第三方面提供了一种姿态估计系统，包括：事件相机、标准相机、存储器及处理器，其中，事件相机用于采集事件数据，标准相机用于采集视频数据；处理器用于执行存储在存储器上的计算机程序，处理器执行计算机程序时，实现上述本申请实施例第一方面提供的姿态估计方法中的各步骤。

[0016] 本申请实施例第四方面提供了一种计算机可读存储介质，其上存储有计算机程序，计算机程序被处理器执行时，实现上述本申请实施例第一方面提供的姿态估计方法中的各步骤。

[0017] 由上可见，根据本申请方案所提供的姿态估计方法、装置、系统及可读存储介质，基于事件数据获取事件图像，以及基于视频数据获取标准图像；将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。通过本申请方案的实施，结合事件相机和标准相机两者收集的数据共同估计系统在运动过程中的姿态变化，当系统处于高速和高动态范围的场景下，可以提供更多运动信息用以估计系统自身的姿态变化，有效提高了姿态估计的精度和稳定性。

附图说明

[0018] 图1为本申请第一实施例提供的一种姿态估计神经网络的结构示意图；

[0019] 图2为本申请第一实施例提供的一种姿态估计方法的基础流程示意图；

[0020] 图3为本申请第一实施例提供的一种事件图像示意图。

[0021] 图4为本申请第二实施例提供的一种姿态估计方法的细化流程示意图；

[0022] 图5为本申请第三实施例提供的姿态估计装置的程序模块示意图；

[0023] 图6为本申请第四实施例提供的姿态估计系统的结构示意图。

具体实施方式

[0024] 为使得本申请的发明目的、特征、优点能够更加的明显和易懂，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而非全部实施例。基于本申请中的实施例，本领域技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0025] 在本申请实施例的描述中，需要理解的是，术语“长度”、“宽度”、“上”、“下”、“前”、“后”、“左”、“右”、“竖直”、“水平”、“顶”、“底”“内”、“外”等指示的方位或位置关系为基于附图所示的方位或位置关系，仅是为了便于描述本申请实施例和简化描述，而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作，因此不能理解为对本发明的限制。

[0026] 此外，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。在本申请实施例的描述中，“多个”的含义是两个或两个以上，除非另有明确具体的限定。

[0027] 在本申请实施例中，除非另有明确的规定和限定，术语“安装”、“相连”、“连接”、“固定”等术语应做广义理解，例如，可以是固定连接，也可以是可拆卸连接，或成一体；可以是机械连接，也可以是电连接；可以是直接相连，也可以通过中间媒介间接相连，可以是两个元件内部的连通或两个元件的相互作用关系。对于本领域的普通技术人员而言，可以根据具体情况理解上述术语在本申请实施例中的具体含义。

[0028] 以上所述仅为本申请的较佳实施例而已，并不用以限制本发明，凡在本申请的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本申请的保护范围之内。

[0029] 在最近二十多年时间里，VO已经广泛应用于各类机器人的导航定位中，如美国NASA开发的火星探测器“勇气号”和“机遇号”，还有无人空中飞行器、水下机器人、陆地机器人等。这些主流的VO方法主要是基于图片中物体的几何特性来估计相机的位置，所以要求图片含有大量稳定的纹理特征，一旦场景中出现遮挡物或在雾天雨天取景，在没有其他传感器(IMU、激光、雷达等)的情况下，几何法的求解精度都会受到很严重的干扰。而很多实际应用中，诸多其他传感器也可能派不上用场，所以只通过视觉来定位的方法还有很大的研究空间。

[0030] 为了解决相关技术中基于图像中物体的几何特性来估计相机的姿态，所导致的姿态估计精度较低的问题，本申请第一实施例提供了一种姿态估计方法，应用于姿态估计系统，该系统部署在移动物体上，姿态估计系统配置有事件相机以及标准相机，事件相机用于采集事件数据，标准相机用于采集视频数据，事件相机是一种新型的仿生传感器，可在相机上的所有像素上异步捕捉光强度的变化，与标准相机相比，它在捕捉快速移动和高动态范围的场景时具有更好的性能，标准相机优选的可以采用深度相机实现。应当理解的是，本实施例的事件相机以及标准相机可以是相互独立的两个相机，也可以是一个集成相机上独立设置的两个不同传感器单元，在实际应用中，两者协同工作同时触发数据采集，事件相机以事件流的形式记录事件数据，事件数据可以txt格式存储，标准相机以连续的图像帧的形式记录视频数据，视频数据可以一系列png格式的图片进行存储。

[0031] 应当说明的是，为了实现本实施例的姿态估计方法，本实施例采用姿态估计神经网络执行姿态估计，如图1所示为本实施例提供的一种姿态估计神经网络的结构示意图，整体网络包括：特征提取分支网络A、特征融合分支网络B以及姿态估计分支网络C。

[0032] 如图2为本实施例提供的姿态估计方法的基础流程示意图，该姿态估计方法包括以下的步骤：

[0033] 步骤201、基于事件数据获取事件图像，以及基于视频数据获取标准图像。

[0034] 具体的，在本实施例中，视频数据有多张时序相连的标准图像组成，视频中每一张标准图像记录的时间定义为$t\in[t_0,t_1,\cdots,t_{n-1}]$，收集时间在$t_m$和$t_{m+1}$之间的事件数据，这些收集到的事件数据和第m张标准图像匹配在一起。应当说明的是，事件数据的表示形式为$e=\{x_i,y_i,t_i,p_i\}$，$i\in0,1,\cdots,n-1$，在实际应用中可将事件数据根据时间分为多个体素，然后将一个体素的事件转换为一张二维事件图像，每一张事件图像与一张标准图像在时间上

对齐。

[0035] 在本实施例一种实施方式中，上述基于事件数据获取事件图像的步骤，包括：将事件数据根据事件划分为多个体素；其中，事件数据表示为$e=\{x_i,y_i,t_i,p_i\}$，$i\in 0,1,\cdots,n-1$，x_i,y_i表示像素坐标位置，t_i表示时间戳，p_i表示事件极性；基于预设转换公式将各体素转换为事件图像。该转换公式表示为：

[0036] $$t_i^* = (B-1)(t_i - t_0)/(t_{n-1} - t_0)$$

[0037] $$V(x,y,t) = \sum_i p_i \max(0,1-|x-x_i|)\max(0,1-|y-y_i|)\max(0,1-|t-t_i^*|)$$

[0038] 其中，V(x,y,t)表示体素，B表示体素的数量。

[0039] 应当理解的是，本实施例为了避免局部事件的过分堆积导致画面模糊，可以将特定时间采集的事件数据转换为一张高为260、宽为346的二维图像，具体请参阅如图3所示的事件图像示意图。

[0040] 步骤202、将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征。

[0041] 请再次参阅图1，本实施例将源于标准相机(Standard Camera)的相邻标准图像以及源于事件相机(Event Camera)的相邻事件图像分别沿通道维度进行合并，然后作为特征提取分支网络的输入，由特征提取分支网络进行特征提取。应当理解的是，本实施例所选取的时序相邻的事件图像的数量以及时序相邻的标准图像的数量，优选地均可以为两张。

[0042] 在本实施例一种实施方式中，上述将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并的步骤之前，还包括：针对所有事件图像以及标准图像分别计算均值和方差，对所有事件图像以及标准图像进行归一化处理以及标准化处理；将各事件图像以及各标准图像分别复制三份后置于色彩三通道中，以预设维度的张量对事件图像以及标准图像进行表达。

[0043] 具体的，在本实施例中，一方面，计算所有事件图像像素的均值和方差，对所有事件图像进行归一化和标准化的处理，将每一张代表事件的二维图像复制三份，放在三个不同的通道中，变成一个大小为[3,260,346]的张量，相邻的事件图像在通道这一维合并，变成大小为[6,260,346]的张量，得到的张量再输入至特征提取分支网络中，最后得到一个大小为[1024,5,6]的张量，也即最终所提取的事件图像特征；另一方面，计算整段视频数据中所有标准图像像素的均值和方差，对所有图像进行标准化和归一化的操作，视频中的每一张图像高为260宽为346，有RGB三个通道，是一个大小为[3,260,346]的张量，将相邻的图像在通道这一维合并，变成大小为[6,260,346]的张量，得到的张量输入至特征提取分支网络的网络结构中，最后得到一个大小为[1024,5,6]的张量，也即标准图像特征。

[0044] 步骤203、将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征。

[0045] 具体的，在本实施例一种实施方式中，特征融合分支网络包括自注意力网络，请再次参阅图1，自注意力网络包括第一自注意力网络(也即图1中Self-attention(event))以及第二自注意力网络(也即图1中Self-attention(image))。相应的，上述将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征的步骤，包括：将

事件图像特征输入至特征融合分支网络的第一自注意力网络，获取事件图像特征相应的第一掩码，以及将标准图像特征输入至第二自注意力网络，获取标准图像特征相应的第二掩码；通过第一自注意力网络对事件图像特征与第一掩码进行乘法运算，得到第一粗过滤特征，以及通过第二自注意力网络对标准图像特征与第二掩码进行乘法运算，得到第二粗过滤特征；基于第一粗过滤特征以及第二粗过滤特征进行特征融合，得到融合特征。

[0046] 应当理解的是，自注意力网络通过比较单个通道与所有通道的相似性来捕获长期依赖关系和全局通道相关性，其关注来自同一传感器的主要特征，允许更好的选择标准相机和事件相机收集的特征。

[0047] 进一步地，在本实施例一种实施方式中，上述获取事件图像特征相应的第一掩码的步骤，包括：通过预设第一掩码计算公式获取事件图像特征相应的第一掩码；第一掩码计算公式表示为：$\tilde{e}_i = Sigmoid((W_i e_i)^T W_j e_j) g(e_i)$，其中，$e_i$和$e_j$分别表示事件图像特征的一个通道的特征向量，$W_i$和$W_j$分别表示将$e_i$和$e_j$投影到嵌入空间的可学习权重矩阵，g()表示将特征向量投影到新嵌入空间的函数。

[0048] 以及，上述获取标准图像特征相应的第二掩码的步骤，包括：通过预设第二掩码计算公式获取标准图像特征相应的第二掩码；第二掩码计算公式表示为：$\tilde{f}_i = Sigmoid((W_i f_i)^T W_j f_j) g(f_i)$，其中，$f_i$和$f_j$分别表示标准图像特征的一个通道的特征向量。

[0049] 具体的，本实施例针对事件图像特征以及标准图像特征分别进行上述掩码计算公式中的操作得到相应的掩码，事件图像特征和标准图像特征的掩码$\tilde{e}_i$和$\tilde{f}_i$大小均为[1024,1,1]，$\tilde{e}_i$和$\tilde{f}_i$分别与e和f相乘后输出的张量e_{att}和f_{att}大小均为[1024,5,6]，e_{att}和f_{att}也即第一粗过滤特征和第二粗过滤特征，两者为图像特征中经过自过滤后所得到的有用特征。

[0050] 更进一步地，在本实施例一种实施方式中，特征融合分支网络还包括交叉注意力网络(也即图1中Cross-attention)，交叉注意力网络旨在融合不同传感器互补的数据，更关注不同传感器在不同场景下的特征。

[0051] 相应的，上述基于第一粗过滤特征以及第二粗过滤特征进行特征融合，得到融合特征的步骤，包括：将第一粗过滤特征输入至交叉注意力网络，获取标准图像特征相应的第三掩码，以及将第二粗过滤特征输入至交叉注意力网络，获取事件图像特征相应的第四掩码；通过交叉注意力网络对第一粗过滤特征与第四掩码进行乘法运算，得到第一细过滤特征，以及对第二粗过滤特征与第三掩码进行乘法运算，得到第二细过滤特征；对第一细过滤特征以及第二细过滤特征进行特征融合，得到融合特征。

[0052] 本实施例优选地，上述获取标准图像特征相应的第三掩码的步骤，包括：通过预设第三掩码计算公式获取标准图像特征相应的第三掩码；第三掩码计算公式表示为：$m_{e\to f}$=Sigmoid($(W_{e\to f}e_i)^T W_{e\to f}e_j$)。

[0053] 以及，上述获取事件图像特征相应的第四掩码的步骤，包括：通过预设第四掩码计算公式获取事件图像特征相应的第四掩码；第四掩码计算公式表示为：$m_{f\to e}$=Sigmoid($(W_{f\to e}e_i)^T W_{f\to e}e_j$)。

[0054] 应当说明的是，掩码$m_{f\to e}$和$m_{e\to f}$的大小均为[1024,1,1]，e_{att}和f_{att}分别乘以相应掩

码得到输出e_{out}和f_{out}，e_{out}和f_{out}的大小均为[1024,5,6]，e_{out}和f_{out}也即第一细过滤特征和第二细过滤特征，两者为原始提取的图像特征中更为有用的特征，把e_{out}和f_{out}从三个维度变成一个维度得到张量[30720]，e_{out}和f_{out}在通道处连接，最后得到输出结果，特征融合公式表示为：$\widetilde{V_{ef}} = [m_{f\to e} \odot f_{att};\ m_{e\to f} \odot e_{out}]$。

[0055] 步骤204、将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。

[0056] 具体的，在本实施例中，姿态估计分支网络包括长短期记忆单元(也即图1中LSTM)以及全连接层，LSTM部分用于估计相机经历的帧到帧运动，在实际应用中，优选的采用两个LSTM堆叠，以获得更好的性能。本实施例将融合特征输入至姿态估计分支网络的长短期记忆单元进行时序上的建模，得到估计数据；将估计数据输入至全连接层进行处理，输出系统姿态变化数据。

[0057] 具体的，本实施例将上述得到的融合特征$\widetilde{V_{ef}}$输入至LSTM进行时序上的建模，从而输出out大小为[1000]，最后由全连接层输出代表位移的x、y、z和代表旋转的欧拉角，从而得到相邻时刻间系统的姿态变化。

[0058] 基于上述本申请实施例的技术方案，基于事件数据获取事件图像，以及基于视频数据获取标准图像；将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。通过本申请方案的实施，结合事件相机和标准相机两者收集的数据共同估计系统在运动过程中的姿态变化，当系统处于高速和高动态范围的场景下，可以提供更多运动信息用以估计系统自身的姿态变化，有效提高了姿态估计的精度和稳定性。

[0059] 图4中的方法为本申请第二实施例提供的一种细化的姿态估计方法，该姿态估计方法包括：

[0060] 步骤401、基于事件数据获取事件图像，以及基于视频数据获取深度图像。

[0061] 具体的，在本实施例中，事件相机与深度相机协同工作同时触发数据采集，事件相机以事件流的形式记录事件数据，深度相机以连续的图像帧的形式记录视频数据。

[0062] 步骤402、将时序相邻的两张事件图像以及时序相邻的两张深度图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及深度图像特征。

[0063] 步骤403、将事件图像特征输入至特征融合分支网络的第一自注意力网络，获取事件图像特征相应的第一掩码，以及将深度图像特征输入至第二自注意力网络，获取深度图像特征相应的第二掩码。

[0064] 步骤404、通过第一自注意力网络对事件图像特征与第一掩码进行乘法运算，得到第一粗过滤特征，以及通过第二自注意力网络对深度图像特征与第二掩码进行乘法运算，得到第二粗过滤特征。

[0065] 具体的，自注意力网络通过比较单个通道与所有通道的相似性来捕获长期依赖关系和全局通道相关性，其关注来自同一传感器的主要特征，允许更好的选择标准相机和事件相机收集的特征。

[0066] 步骤405、将第一粗过滤特征输入至交叉注意力网络，获取深度图像特征相应的第三掩码，以及将第二粗过滤特征输入至交叉注意力网络，获取事件图像特征相应的第四掩码。

[0067] 步骤406、通过交叉注意力网络对第一粗过滤特征与第四掩码进行乘法运算，得到第一细过滤特征，以及对第二粗过滤特征与第三掩码进行乘法运算，得到第二细过滤特征。

[0068] 具体的，交叉注意力网络旨在融合不同传感器互补的数据，更关注不同传感器在不同场景下的特征，可从自注意力网络所提取的有用特征中进一步提取更为有用的特征。

[0069] 步骤407、对第一细过滤特征以及第二细过滤特征进行特征融合，得到融合特征。

[0070] 步骤408、将融合特征输入至姿态估计分支网络的长短期记忆单元进行时序上的建模，得到估计数据；将估计数据输入至全连接层进行处理，输出系统姿态变化数据。

[0071] 具体的，本实施例将上述得到的融合特征输入至LSTM进行时序上的建模，从而输出out大小为[1000]，最后由全连接层输出代表位移的x、y、z和代表旋转的欧拉角，从而得到相邻时刻间系统的姿态变化。

[0072] 应当理解的是，本实施例中各步骤的序号的大小并不意味着步骤执行顺序的先后，各步骤的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成唯一限定。

[0073] 基于本申请实施例上述方案，能够在多种场景下准确的预测系统自身的姿态变化，将系统部署在移动的设备上，根据事件相机和深度相机拍摄到的内容就可以得到系统的姿态变化。通过大量实验，结果表明，与仅在各种场景中使用深度相机的方法相比，本申请实施例通过自动采用来自不同来源的优势信息，可以大大提高轨迹估计精度。

[0074] 图5为本申请第三实施例提供的一种姿态估计装置。该姿态估计装置应用于姿态估计系统，姿态估计系统配置有事件相机以及标准相机，事件相机用于采集事件数据，标准相机用于采集视频数据。如图5所示，该姿态估计装置主要包括：

[0075] 获取模块501，用于基于事件数据获取事件图像，以及基于视频数据获取标准图像；

[0076] 提取模块502，用于将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；

[0077] 融合模块503，用于将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；

[0078] 估计模块504，用于将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。

[0079] 在本实施例的一些实施方式中，获取模块在执行上述基于事件数据获取事件图像的功能时，具体用于：将事件数据根据事件划分为多个体素；其中，事件数据表示为$e=\{x_i, y_i, t_i, p_i\}$，$i\in 0,1,\cdots,n-1$，x_i, y_i表示像素坐标位置，t_i表示时间戳，p_i表示事件极性；基于预设转换公式将各体素转换为事件图像；转换公式表示为：

[0080] $$t_i^* = (B-1)(t_i - t_0)/(t_{n-1} - t_0)$$

[0081] $$V(x,y,t) = \sum_i p_i \max(0,1-|x-x_i|)\max(0,1-|y-y_i|)\max(0,1-|t-t_i^*|)$$

[0082] 其中,V(x,y,t)表示体素,B表示体素的数量。

[0083] 在本实施例的一些实施方式中,该姿态估计装置还包括:处理模块,用于针对所有事件图像以及标准图像分别计算均值和方差,对所有事件图像以及标准图像进行归一化处理以及标准化处理;将各事件图像以及各标准图像分别复制三份后置于色彩三通道中,以预设维度的张量对事件图像以及标准图像进行表达。

[0084] 在本实施例的一些实施方式中,特征融合分支网络包括自注意力网络,自注意力网络包括第一自注意力网络以及第二自注意力网络。相应的,融合模块具体用于:将事件图像特征输入至特征融合分支网络的第一自注意力网络,获取事件图像特征相应的第一掩码,以及将标准图像特征输入至第二自注意力网络,获取标准图像特征相应的第二掩码;通过第一自注意力网络对事件图像特征与第一掩码进行乘法运算,得到第一粗过滤特征,以及通过第二自注意力网络对标准图像特征与第二掩码进行乘法运算,得到第二粗过滤特征;基于第一粗过滤特征以及第二粗过滤特征进行特征融合,得到融合特征。

[0085] 进一步地,在本实施例的一些实施方式中,融合模块在执行上述获取事件图像特征相应的第一掩码的功能时具体用于:通过预设第一掩码计算公式获取事件图像特征相应的第一掩码;第一掩码计算公式表示为:$\tilde{e}_i = Sigmoid((W_i e_i)^T W_j e_j) g(e_i)$,其中,$e_i$和$e_j$分别表示事件图像特征的一个通道的特征向量,$W_i$和$W_j$分别表示将$e_i$和$e_j$投影到嵌入空间的可学习权重矩阵,g()表示将特征向量投影到新嵌入空间的函数。另外,融合模块在执行上述获取标准图像特征相应的第二掩码的功能时具体用于:通过预设第二掩码计算公式获取标准图像特征相应的第二掩码;第二掩码计算公式表示为:$\tilde{f}_i = Sigmoid((W_i f_i)^T W_j f_j) g(f_i)$,其中,$f_i$和$f_j$分别表示标准图像特征的一个通道的特征向量。

[0086] 进一步地,在本实施例的另一些实施方式中,特征融合分支网络还包括交叉注意力网络。相应的,融合模块在执行上述基于第一粗过滤特征以及第二粗过滤特征进行特征融合,得到融合特征的功能时,具体用于:将第一粗过滤特征输入至交叉注意力网络,获取标准图像特征相应的第三掩码,以及将第二粗过滤特征输入至交叉注意力网络,获取事件图像特征相应的第四掩码;通过交叉注意力网络对第一粗过滤特征与第四掩码进行乘法运算,得到第一细过滤特征,以及对第二粗过滤特征与第三掩码进行乘法运算,得到第二细过滤特征;对第一细过滤特征以及第二细过滤特征进行特征融合,得到融合特征。

[0087] 更进一步地,在本实施例的一些实施方式中,融合模块在执行上述获取标准图像特征相应的第三掩码的功能时,具体用于:通过预设第三掩码计算公式获取标准图像特征相应的第三掩码;第三掩码计算公式表示为:$m_{e\to f}$=Sigmoid$((W_{e\to f}e_i)^T W_{e\to f}e_j)$。另外,融合模块在执行上述获取事件图像特征相应的第四掩码的功能时,具体用于:通过预设第四掩码计算公式获取事件图像特征相应的第四掩码;第四掩码计算公式表示为:$m_{f\to e}$=Sigmoid$((W_{f\to e}e_i)^T W_{f\to e}e_j)$。

[0088] 在本实施例的一些实施方式中,姿态估计分支网络包括长短期记忆单元以及全连接层。相应的,估计模块具体用于:将融合特征输入至姿态估计分支网络的长短期记忆单元进行时序上的建模,得到估计数据;将估计数据输入至全连接层进行处理,输出系统姿态变化数据。

[0089] 应当说明的是，第一、二实施例中的姿态估计方法均可基于本实施例提供的姿态估计装置实现，所属领域的普通技术人员可以清楚的了解到，为描述的方便和简洁，本实施例中所描述的姿态估计装置的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0090] 根据本实施例所提供的姿态估计装置，基于事件数据获取事件图像，以及基于视频数据获取标准图像；将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征；将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征；将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据。通过本申请方案的实施，结合事件相机和标准相机两者收集的数据共同估计系统在运动过程中的姿态变化，当系统处于高速和高动态范围的场景下，可以提供更多运动信息用以估计系统自身的姿态变化，有效提高了姿态估计的精度和稳定性。

[0091] 图6为本申请第四实施例提供的一种姿态估计系统。该姿态估计系统可用于实现前述实施例中的姿态估计方法，主要包括：

[0092] 存储器601、处理器602、事件相机603、标准相机604及存储在存储器601上并可在处理器602上运行的计算机程序605，存储器601和处理器602通过通信连接。处理器602执行该计算机程序605时，实现前述实施例一或二中的方法。其中，处理器的数量可以是一个或多个。

[0093] 存储器601可以是高速随机存取记忆体(RAM，Random Access Memory)存储器，也可为非不稳定的存储器(non-volatile memory)，例如磁盘存储器。存储器601用于存储可执行程序代码，处理器602与存储器601耦合。

[0094] 进一步的，本申请实施例还提供了一种计算机可读存储介质，该计算机可读存储介质可以是设置于上述姿态估计系统中，该计算机可读存储介质可以是前述图6所示实施例中的存储器。

[0095] 该计算机可读存储介质上存储有计算机程序，该程序被处理器执行时实现前述实施例中的姿态估计方法。进一步的，该计算机可存储介质还可以是U盘、移动硬盘、只读存储器(ROM，Read-Only Memory)、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0096] 在本申请所提供的几个实施例中，应该理解到，所揭露的装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个模块或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或模块的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0097] 作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理模块，即可以位于一个地方，或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0098] 另外，在本申请各个实施例中的各功能模块可以集成在一个处理模块中，也可以是各个模块单独物理存在，也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。

[0099] 集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个可读存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本申请各个实施例方法的全部或部分步骤。而前述的可读存储介质包括:U盘、移动硬盘、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0100] 需要说明的是，对于前述的各方法实施例，为了简便描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其它顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定都是本申请所必须的。

[0101] 在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其它实施例的相关描述。

[0102] 以上为对本申请所提供的姿态估计方法、装置、系统及可读存储介质的描述，对于本领域的技术人员，依据本申请实施例的思想，在具体实施方式及应用范围上均会有改变之处，综上，本说明书内容不应理解为对本申请的限制。

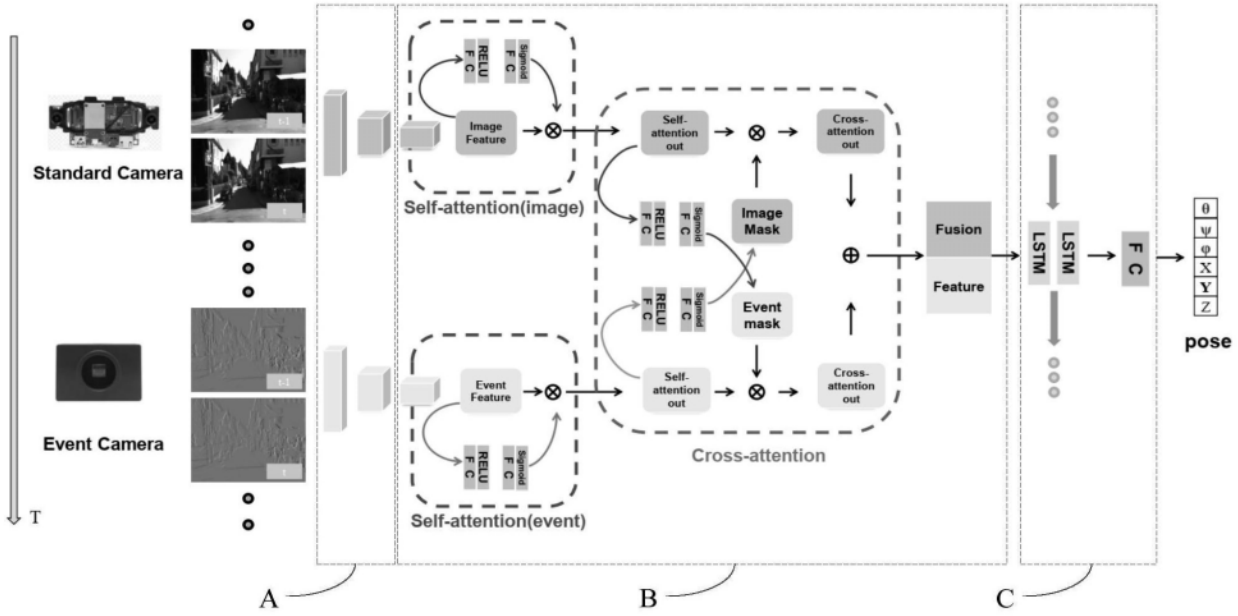

Standard Camera
Event Camera
T
Image Feature
Self-attention(image)
Event Feature
Self-attention(event)
Self-attention out
Cross-attention out
Image Mask
Event mask
Cross-attention
Fusion Feature
LSTM
LSTM
F C
pose
A
B
C

图1

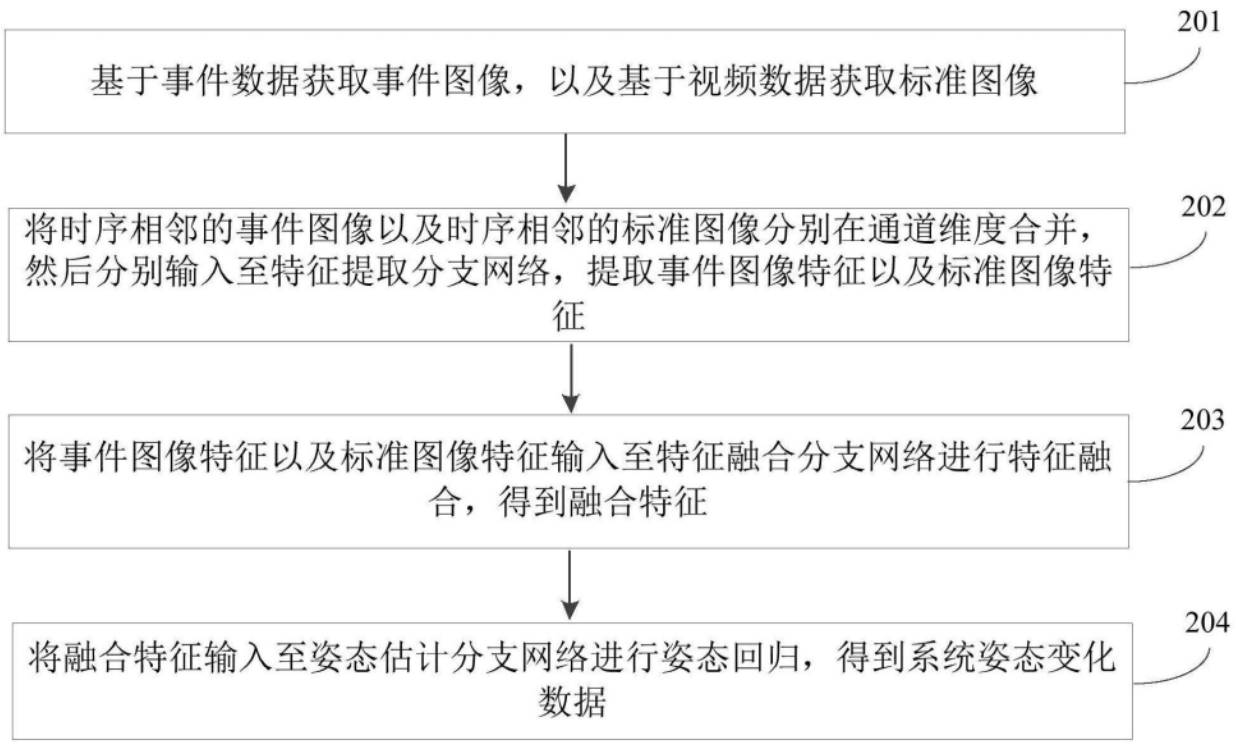

201
基于事件数据获取事件图像，以及基于视频数据获取标准图像
202
将时序相邻的事件图像以及时序相邻的标准图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及标准图像特征
203
将事件图像特征以及标准图像特征输入至特征融合分支网络进行特征融合，得到融合特征
204
将融合特征输入至姿态估计分支网络进行姿态回归，得到系统姿态变化数据

图2

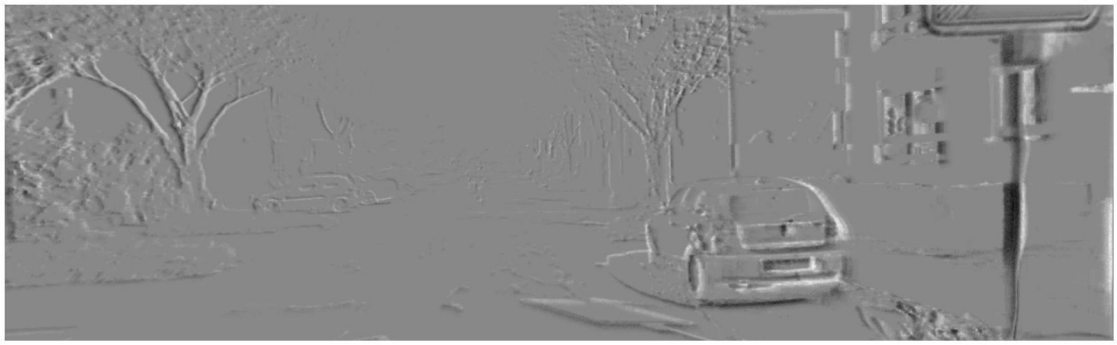


图3

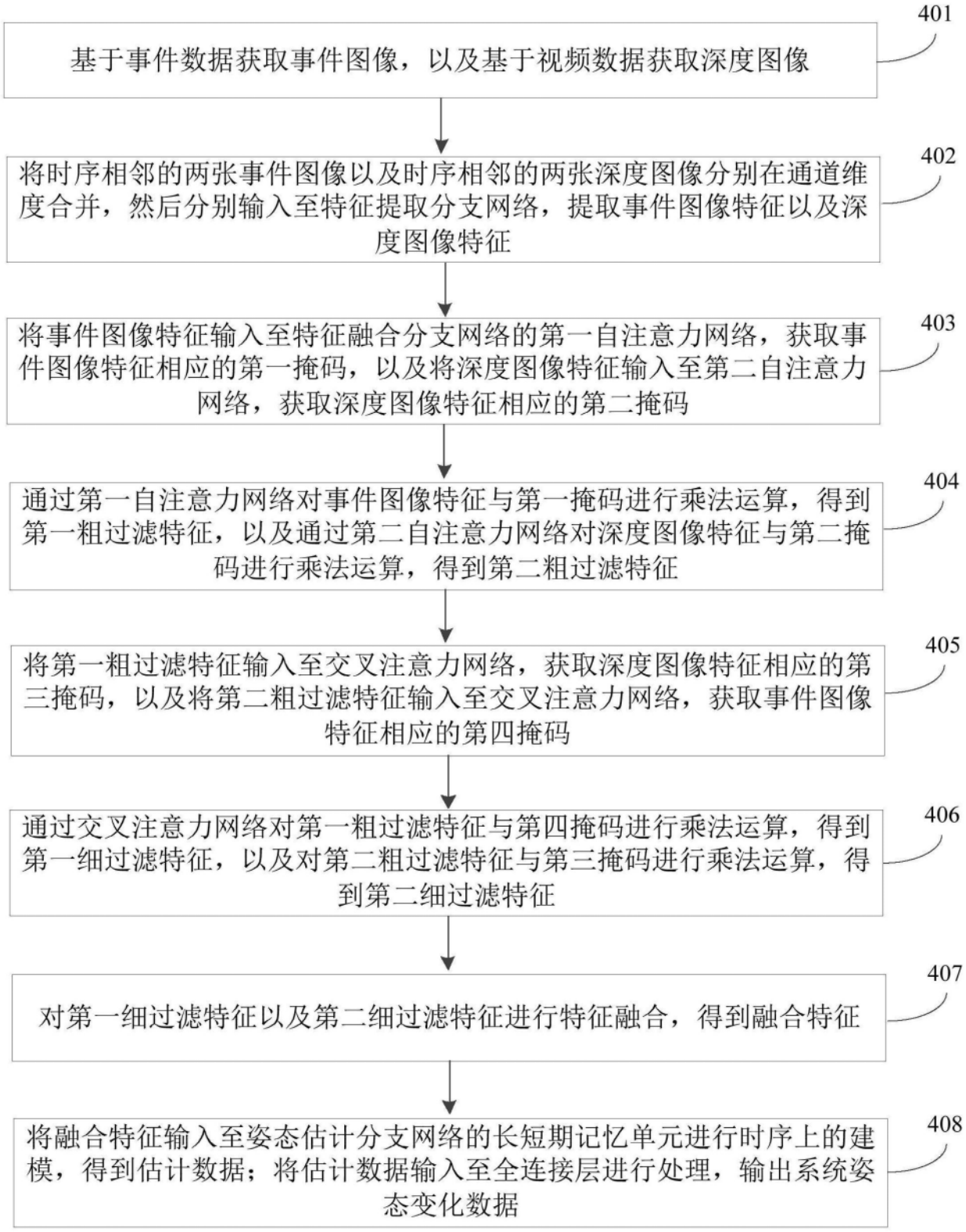

401
基于事件数据获取事件图像，以及基于视频数据获取深度图像
402
将时序相邻的两张事件图像以及时序相邻的两张深度图像分别在通道维度合并，然后分别输入至特征提取分支网络，提取事件图像特征以及深度图像特征
403
将事件图像特征输入至特征融合分支网络的第一自注意力网络，获取事件图像特征相应的第一掩码，以及将深度图像特征输入至第二自注意力网络，获取深度图像特征相应的第二掩码
404
通过第一自注意力网络对事件图像特征与第一掩码进行乘法运算，得到第一粗过滤特征，以及通过第二自注意力网络对深度图像特征与第二掩码进行乘法运算，得到第二粗过滤特征
405
将第一粗过滤特征输入至交叉注意力网络，获取深度图像特征相应的第三掩码，以及将第二粗过滤特征输入至交叉注意力网络，获取事件图像特征相应的第四掩码
406
通过交叉注意力网络对第一粗过滤特征与第四掩码进行乘法运算，得到第一细过滤特征，以及对第二粗过滤特征与第三掩码进行乘法运算，得到第二细过滤特征
407
对第一细过滤特征以及第二细过滤特征进行特征融合，得到融合特征
408
将融合特征输入至姿态估计分支网络的长短期记忆单元进行时序上的建模，得到估计数据；将估计数据输入至全连接层进行处理，输出系统姿态变化数据

图4


姿态估计装置
获取模块
501
提取模块
502
融合模块
503
估计模块
504

图5

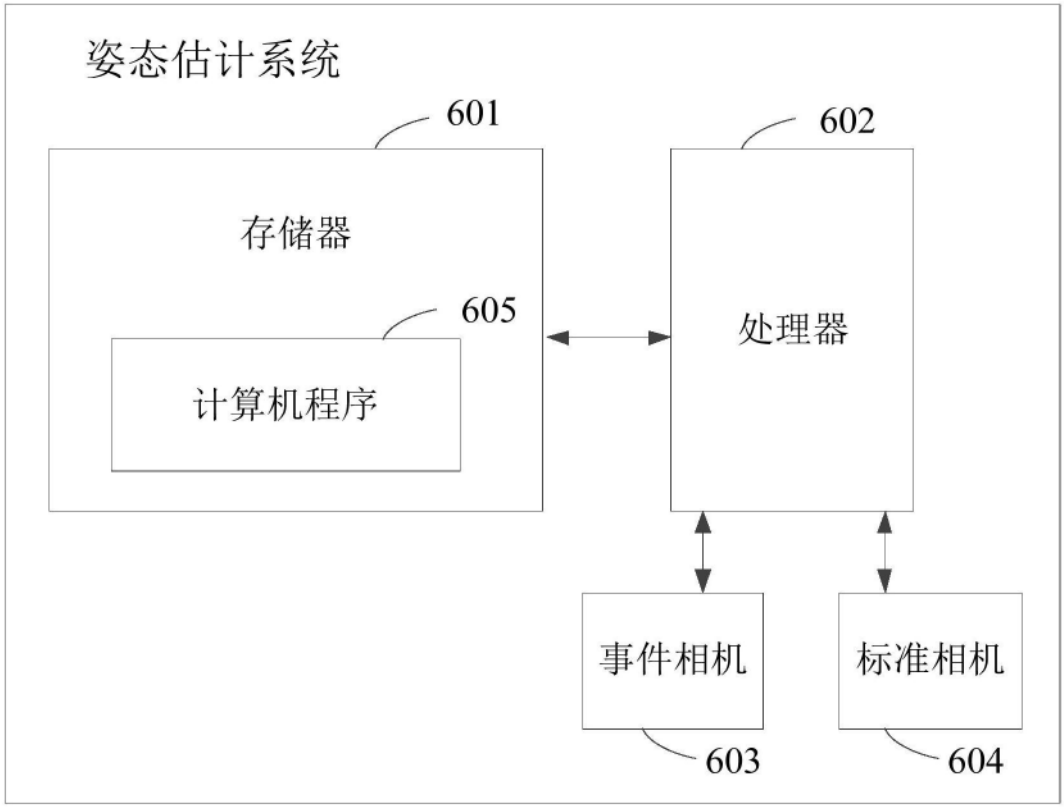

姿态估计系统
601
存储器
605
计算机程序
602
处理器
事件相机
603
标准相机
604

图6