

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2019 年 10 月 3 日 (03.10.2019)



(10) 国际公布号
WO 2019/183861 A1

- (51) 国际专利分类号:
G06F 9/50 (2006.01)
- (21) 国际申请号: PCT/CN2018/080970
- (22) 国际申请日: 2018 年 3 月 28 日 (28.03.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 深圳市大疆创新科技有限公司 (SZ DJI TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省深圳市南山区高新区南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。
- (72) 发明人: 李庆 (LI, Qing); 中国广东省深圳市南山区高新区南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。 夏昌奇 (XIA, Changqi); 中国广东省深圳市南山区高新区南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。 张晓炜

(ZHANG, Xiaowei); 中国广东省深圳市南山区高新区南区粤兴一道9号香港科大深圳产学研大楼6楼, Guangdong 518057 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街9号嘉华大厦B座409室, Beijing 100085 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: METHOD, DEVICE, AND MACHINE READABLE STORAGE MEDIUM FOR TASK PROCESSING

(54) 发明名称: 任务处理方法、设备及机器可读存储介质

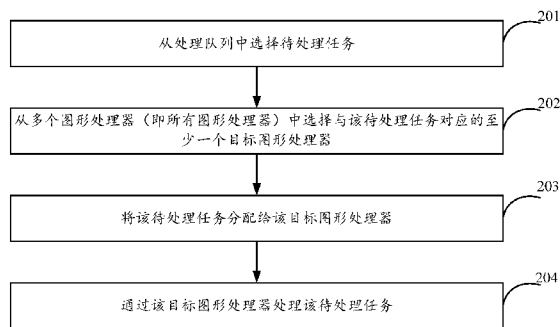


图 2

- 201 Select a task to be processed from a processing queue
202 Select at least one target graphics processor corresponding to the task from a plurality of graphics processors (i.e. all of the graphics processors)
203 Allocate the task to the target graphics processor
204 Process the task by means of the target graphics processor

(57) Abstract: Provided are a method, a device, and a machine readable storage medium for task processing. The method comprises: selecting a task to be processed from a processing queue; selecting at least one target graphics processor corresponding to the task from a plurality of graphics processors; allocating the task to the target graphics processor; and processing the task by means of the target graphics processor. By use of embodiments of the present invention, real-time data processing is ensured, multiple graphics processors can be reasonably managed, multiple tasks to be processed are effectively scheduled, the utilization rates of the graphics processors are maximized, the processing speed is optimized, accuracy and effectiveness of a processing result are effectively ensured, and reliability is improved.

(57) 摘要: 一种任务处理方法、设备及机器可读存储介质, 所述方法包括: 从处理队列中选择待处理任务; 从多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器; 将所述待处理任务分配给所述目标图形处理器; 通过所述目标图形处理器处理所述待处理任务。应用本发明实施例, 保证数据处理的实时性, 可以对多个图形处理器进行合理的管理, 对多个待处理任务进行有效调度, 使图形处理器的利用率得到最大限度的利用, 使处理速度达到最优, 有效保证处理结果的精度和有效性, 增加可靠性。

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW。

- (84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

任务处理方法、设备及机器可读存储介质

技术领域

本发明涉及图像处理技术领域，尤其是涉及一种任务处理方法、设备及机器可读存储介质。

5 背景技术

目前，图像或者雷达等数据的处理过程，可以基于 ARM (Advanced RISC Machines, 高级精简指令集处理器)、DSP (Digital Signal Processing, 数字信号处理) 或者 CPU (Central Processing Unit, 中央处理器) 等平台，其能够处理的数据量受到处理器主频、内存大小、传输带宽等因素的影响，在自动驾驶等人工智能应用场景中，传感器数据和采集频率使得数据量级达到 GB/s 以上，从而导致 ARM、DSP 或者 CPU 等平台已经不能满足实时处理的需求。

发明内容

本发明提供一种任务处理方法、设备及机器可读存储介质。

本发明第一方面，提供一种任务处理方法，应用于包括多个图形处理器的设备，所述方法包括：从处理队列中选择待处理任务；

从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器；将所述待处理任务分配给所述目标图形处理器；

通过所述目标图形处理器处理所述待处理任务。

本发明第二方面，提供一种任务处理设备，包括调度器和多个图形处理器；所述调度器，用于从处理队列中选择待处理任务，并从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器，并将所述待处理任务分配给所述目标图形处理器；

所述目标图形处理器，用于处理所述待处理任务。

本发明第三方面，提供一种计算机可读存储介质，所述计算机可读存储介质上存储有若干计算机指令，计算机指令被执行时实现上述任务处理方法。

基于上述技术方案，本发明实施例中，可以从多个图形处理器中选择与待处理任务对应的目标图形处理器，将待处理任务分配给该目标图形处理器，
5 并通过该目标图形处理器处理待处理任务，从而基于图形处理器进行数据的实时处理，保证数据处理的实时性，可以针对 GB/s 量级的数据进行实时处理。此外，可以对多个图形处理器进行合理的管理，对多个待处理任务进行有效调度，使图形处理器的利用率得到最大限度的利用，使处理速度达到最优，有效保证处理结果的精度和有效性，增加可靠性。

10 附图说明

为了更加清楚地说明本发明实施例或者现有技术中的技术方案，下面将对本发明实施例或者现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明中记载的一些实施例，对于本领域普通技术人员来讲，还可以根据本发明实施例的这些附图获得其它的附图。

15 图 1 是本发明实施例的应用场景示意图；

图 2 是本发明的任务处理方法的一个实施例流程图；

图 3 是本发明的通过目标图形处理器处理待处理任务的示意图；

图 4 是本发明的通过不同的处理线程进行数据共享的示意图；

图 5 是本发明的任务处理设备的一个实施例框图。

20 具体实施方式

下面将结合本发明实施例中的附图，对本发明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发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通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，都属于本发明保护的范围。

另外，在不冲突的情况下，下述的实施例及实施例中的特征可以相互组合。

本发明使用的术语仅仅是出于描述特定实施例的目的，而非用于限制本发明。本发明和权利要求书所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其它含义。应当理解，本文中使用的
5 术语“和/或”是指包含一个或多个相关联的列出项目的任何或所有可能组合。

尽管在本发明可能采用术语第一、第二、第三等来描述各种信息，但
10 些信息不应限于这些术语。这些术语用来将同一类型的信息彼此区分开。例如，在不脱离本发明范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，此外，所使用的词语“如果”可以被解释成为“在……时”，或者“当……时”，或者“响应于确定”。

本发明实施例中提出一种任务处理方法，该方法可以应用于包括多个图形处理器（Graphics Processing Unit，简称 GPU）的处理设备，对此处理设备的类型不做限制，只要具有多个图形处理器即可。其中，该任务处理方法
15 可以是基于多个图形处理器的多任务实时处理方法，是以多路传感器数据作为输入，其可以应用在自动驾驶、辅助驾驶、室内外作业机器人等应用场景中。此外，上述传感器数据可以是图像数据，如相机到采集的图像数据、雷达（如激光雷达、毫米波雷达等）采集到的图像数据等，对此数据类型不做限制。

在人工智能应用场景，如数据需要实时处理的人工智能应用场景，随着
20 传感器数量的增加、采集帧率的增加，数据量成倍增加，ARM、DSP 或者 CPU 等已无法实现数据的实时处理，因此，本实施例提出基于多个图形处理器的多任务实时处理方法，可以保证数据的实时处理、处理结果的精度和有效性。

其中，考虑到需要并行处理多个任务，且不同任务具有不同的处理要求，
25 本实施例中，借助图形处理器强大的计算性能，提出基于多图形处理器的实时任务处理，可以实现多任务的并行调度，并实现任务的公平性、优先级、独立性、中断以及恢复等功能，可以对多个图形处理器进行合理的管理，对

多个待处理任务进行有效调度，使图形处理器的利用率得到最大限度的利用，使处理速度达到最优，有效保证处理结果的精度和有效性，增加可靠性。

参见图 1 所示，为本发明实施例的应用场景示意图，多任务调度模块用于对任务进行缓存，完成任务调度，向上层应用响应任务处理结果等。处理
5 模块由多个图形处理器组成，通过图形处理器完成任务的实时处理。统计监测模块用于对任务的状态信息和图形处理器的资源信息进行监测统计，将任务的状态信息和图形处理器的资源信息反馈给多任务调度模块，由多任务调度模块利用任务的状态信息和图形处理器的资源信息进行任务调度。

实施例一：

10 如图 2 所示，为本发明实施例中的任务处理方法的流程图，该方法包括：
步骤 201，从处理队列中选择待处理任务。

具体的，在接收到任务后，可以将该任务存储到处理队列中，在需要对任务进行处理时，则可以从该处理队列中选择任务，对此选择方式不做限制。为了区分方便，可以将从处理队列中选择的任务称为待处理任务。

15 步骤 202，从多个图形处理器（即所有图形处理器）中选择与该待处理任务对应的至少一个目标图形处理器。其中，为了区分方便，可以将从多个图形处理器中选择的图形处理器，称为与待处理任务对应的目标图形处理器。

具体的，在从处理队列中选择待处理任务后，则可以根据该待处理任务的状态信息和/或每个图形处理器的资源信息，从所有图形处理器中选择与该
20 待处理任务对应的至少一个目标图形处理器，对此选择方式不做限制。

步骤 203，将该待处理任务分配给该目标图形处理器。

具体的，在从处理队列中选择待处理任务，并为该待处理任务选择目标图形处理器之后，就可以将该待处理任务分配给该目标图形处理器，也就是说，可以将该待处理任务分配给该目标图形处理器的空闲处理线程。例如，
25 可以将待处理任务分配给该目标图形处理器的全部或者部分空闲处理线程。

步骤 204, 通过该目标图形处理器处理该待处理任务。

具体的, 在将该待处理任务分配给该目标图形处理器的空闲处理线程后, 就可以通过该空闲处理线程处理该待处理任务, 对此处理过程不做限制。例如, 该待处理任务可以包括数据 (如传感器数据等) 和任务类型, 目标图形
5 处理器的该空闲处理线程可以根据该数据, 执行与该任务类型对应的处理。

基于上述技术方案, 本发明实施例中, 可以从多个图形处理器中选择与待处理任务对应的目标图形处理器, 将待处理任务分配给该目标图形处理器, 并通过该目标图形处理器处理待处理任务, 从而基于图形处理器进行数据的实时处理, 保证数据处理的实时性, 可以针对 GB/s 量级的数据进行实时处理。
10 此外, 可以对多个图形处理器进行合理的管理, 对多个待处理任务进行有效调度, 使图形处理器的利用率得到最大限度的利用, 使处理速度达到最优, 有效保证处理结果的精度和有效性, 增加可靠性。

实施例二:

在步骤 201 之前, 在接收到待处理任务后, 若该待处理任务是时间序列
15 和/或数据具有依赖关系的静态调度任务, 则可以将该待处理任务缓存到静态处理队列; 若该待处理任务是时间序列和/或数据独立的动态调度任务, 则可以将该待处理任务缓存到动态处理队列。在此基础上, 在步骤 201 中, 从处理队列中选择待处理任务, 可以包括: 若该处理队列是静态处理队列, 则可以从该静态处理队列中选择时间序列和/或数据具有依赖关系的待处理任务;
20 若该处理队列是动态处理队列, 则可以从该动态处理队列中选择时间序列和/或数据独立的待处理任务。

在本实施例中, 可以将处理队列划分为静态处理队列 (Static) 和动态处理队列 (Dynamic)。其中, 该静态处理队列可以是用于存储时间序列具有依赖关系的任务, 或者, 存储数据具有依赖关系的任务, 或者, 存储时间序列
25 和数据均具有依赖关系的任务; 为了区分方便, 可以将上述具有依赖关系的

任务称为静态调度任务。此外，该动态处理队列可以是用于存储时间序列独立的任务，或者，存储数据独立的任务，或者，存储时间序列和数据均独立的任务；为了区分方便，可以将上述独立的任务称为动态调度任务。

而且，针对存储在静态处理队列的所有任务，可以在同一个图形处理器
5 中进行处理，或者，可以在相关的图形处理器中进行处理，也就是说，对图形处理器有限制；可以预先配置静态处理队列与图形处理器（如图形处理器 1）的对应关系，因此，该静态处理队列的所有任务均需要由图形处理器 1 进行处理。此外，针对存储在动态处理队列的所有任务，可以在同一个图形处理器中进行处理，也可以在不同图形处理器中进行处理，也就是说，对图形处
10 理器没有限制，可以将动态处理队列的所有任务分配到任意的图形处理器。

其中，时间序列具有依赖关系的任务是指：一个任务的处理依赖于前一个任务。例如，在 H264 编码中，非 I 帧任务依赖于前一次出现的 I 帧任务。基于此，多任务调度模块在接收到任务 1 和任务 2 后，若任务 1 和任务 2 在时间序列具有依赖关系，则可以将任务 1 和任务 2 均存储到静态处理队列。
15 此外，数据具有依赖关系的任务是指：一个数据的处理依赖于前一个数据，对此不做限制。而且，针对存储在静态处理队列的所有任务，可以在同一个图形处理器中进行处理，或者，可以在相关的图形处理器中进行处理。

其中，时间序列独立的任务是指：一个任务的处理不依赖于前一个任务，即任务是独立任务。例如，图像转换和点云算法处理等任务，各任务就是独立的任务，不依赖于前一个任务。基于此，多任务调度模块在接收到任务 3
20 后，若任务 3 在时间序列上独立，则可以将任务 3 存储到动态处理队列。此外，数据独立的任务是指：一个数据的处理与其它数据无关，对此不做限制。而且，针对存储在动态处理队列的所有任务，可以在同一个图形处理器中进行处理，也可以在不同图形处理器中进行处理。

25 针对静态处理队列中的每个待处理任务，多任务调度模块可以从静态处理队列中选择待处理任务，在待处理任务处理完成或者未处理完成时，可以

从静态处理队列选择另一待处理任务。针对动态处理队列中的每个待处理任务，多任务调度模块可以从动态处理队列中选择待处理任务，在待处理任务处理完成或者未处理完成时，即可以从动态处理队列选择另一待处理任务。

实施例三：

5 在步骤 201 之前，在接收到待处理任务后，可以将该待处理任务缓存到任务队列（也可以称为大队列或 Task Queue），该任务队列可以包括多个待处理任务。针对任务队列中的每个待处理任务，若该待处理任务是时间序列和/或数据具有依赖关系的静态调度任务，则可以将该待处理任务缓存到静态处理队列；若该待处理任务是时间序列和/或数据独立的动态调度任务，则可以
10 将待处理任务缓存到动态处理队列。在此基础上，在步骤 201 中，从处理队列中选择待处理任务，可以包括：若该处理队列是静态处理队列，则可以从该静态处理队列中选择时间序列和/或数据具有依赖关系的待处理任务；若该处理队列是动态处理队列，则可以从该动态处理队列中选择时间序列和/或数据独立的待处理任务。实施例三的处理过程与实施例二类似，在此不再赘述。

15 实施例四：

 在步骤 201 中，从处理队列中选择待处理任务，可以包括但不限于：获取该处理队列中的每个待处理任务的优先级；基于优先级优先从该处理队列中选择高优先级的待处理任务。其中，获取该处理队列中的每个待处理任务的优先级，可以包括：针对该处理队列中的每个待处理任务，可以获取该待
20 处理任务的类型；然后，通过该任务类型查询映射表，得到该任务类型对应的优先级；其中，所述映射表用于记录任务类型与优先级的对应关系。

 其中，多任务调度模块可以配置映射表，参见表 1 所示，该映射表用于记录任务类型与优先级的对应关系。在此基础上，假设待处理任务 1 包括任务类型 A，待处理任务 2 包括任务类型 B，待处理任务 3 包括任务类型 C，则
25 多任务调度模块通过任务类型 A 查询映射表，得到待处理任务 1 的优先级 5，

通过任务类型 B 查询映射表，得到待处理任务 2 的优先级 3，通过任务类型 C 查询映射表，得到待处理任务 3 的优先级 1。显然，由于待处理任务 1 的优先级 5 高于待处理任务 2 的优先级 3，待处理任务 2 的优先级 3 高于待处理任务 3 的优先级 1，因此，多任务调度模块先从处理队列中选择待处理任务 1，然后选择待处理任务 2，然后选择待处理任务 3，以此类推。

表 1

任务类型	优先级
任务类型 A	5
任务类型 B	3
任务类型 C	1

实施例五：为了选择目标图形处理器，则可以先获取待处理任务的状态信息和/或图形处理器的资源信息，以下对这些信息的获取过程进行详细说明。

情况一、多任务调度模块获取待处理任务的状态信息。

10 在待处理任务的处理过程中，统计监测模块可以获取已经完成处理的待处理任务的状态信息，并在状态信息表中记录该待处理任务的任务类型与该待处理任务的状态信息的对应关系；将该状态信息表发送给多任务调度模块。在此基础上，多任务调度模块在从处理队列中选择待处理任务后，先获取该待处理任务的任务类型，并通过该任务类型查询状态信息表，得到与该任务类型对应的状态信息，并将得到的状态信息确定为待处理任务的状态信息。

例如，在处理任务 1、待处理任务 2 和待处理任务 3 已经完成处理后，统计监测模块可以得到表 2 所示的状态信息表，并将该状态信息表发送给多任务调度模块。在此基础上，多任务调度模块从处理队列中选择待处理任务 4 后，若待处理任务 4 的任务类型为任务类型 A，则可以得到与任务类型 A 对应的状态信息 A，也就是说，所述待处理任务 4 的状态信息是状态信息 A。

表 2

任务类型	状态信息
任务类型 A	状态信息 A
任务类型 B	状态信息 B
任务类型 C	状态信息 C

情况二、多任务调度模块获取图形处理器的资源信息。

在待处理任务的处理过程中，统计监测模块可以获取（如周期性获取）每个图形处理器的资源信息，并将每个图形处理器的资源信息发送给多任务调度模块，这样，多任务调度模块可以得到每个图形处理器的资源信息。

- 5 在上述实施例中，状态信息可以包括任务处理时间，该任务处理时间是任务完成时间与任务接收时间的时间差。此外，资源信息可以包括：空闲处理线程的数量；和/或，处理线程的状态（如占用状态或者空闲状态）。

- 10 其中，在每个待处理任务的处理过程中，统计监测模块可以对待处理任务的状态信息进行统计，并将任务类型与状态信息的对应关系通知给多任务调度模块。具体的，待处理任务的处理过程可以经过三个阶段，任务接收缓存阶段、任务调度分发阶段和任务处理阶段，因此，统计监测模块可以对上述三个阶段消耗的时间进行统计，而这三个阶段消耗的时间也就是待处理任务的状态信息，即该待处理任务的任务处理时间。此外，任务处理时间可以是任务完成时间与任务接收时间的时间差，因此，统计监测模块还可以对待
- 15 处理任务的任务接收时间和任务完成时间进行统计，并计算任务完成时间与任务接收时间的时间差，该时间差就是该待处理任务的任务处理时间。

- 20 其中，统计监测模块将任务类型与任务处理时间的对应关系发送给多任务调度模块时，由于任务类型对应多个待处理任务，每个待处理任务具有一个任务处理时间，不同待处理任务对应的任务处理时间可以相同或不同。基于此，从任务类型对应的多个待处理任务的任务处理时间中选择最大值，将最大值作为该任务类型对应的任务处理时间。或，从任务类型对应的多个待

处理任务的任务处理时间中选择最小值，将最小值作为该任务类型对应的任务处理时间。或，计算任务类型对应的多个待处理任务的任务处理时间的平均值，将平均值作为该任务类型对应的任务处理时间。或，还可以将上述最大值、最小值、平均值均作为任务类型对应的任务处理时间，对此不做限制。

5 其中，在每个待处理任务的处理过程中，统计监测模块可以对图形处理器的资源信息进行统计，并将图形处理器的资源信息通知给多任务调度模块。具体的，图形处理器可以支持多任务并行处理，每个图形处理器具有多个处理线程，待处理任务是分配到处处理线程进行处理，因此，统计监测模块可以对每个图形处理器的处理线程进行监测，如监测图形处理器的空闲处理线程
10 的数量、监测图形处理器的处理线程的状态（如占用状态或空闲状态）等。

实施例六：

在步骤 202 中，从多个图形处理器中选择与该待处理任务对应的至少一个目标图形处理器，可以包括：根据该待处理任务的状态信息，从多个图形处理器中选择与该待处理任务对应的目标图形处理器。或者，根据每个图形
15 处理器的资源信息，从多个图形处理器中选择与该待处理任务对应的目标图形处理器。或者，根据该待处理任务的状态信息和每个图形处理器的资源信息，从多个图形处理器中选择与该待处理任务对应的目标图形处理器。

情况一、根据待处理任务的状态信息，从多个图形处理器中选择与该待处理任务对应的目标图形处理器，可以包括：根据待处理任务的状态信息确
20 定任务处理时间，针对任务处理时间大于时间阈值的多个待处理任务（表示其任务处理时间比较长），为不同的待处理任务选择不同的目标图形处理器。

其中，可以根据经验设置时间阈值，当任务处理时间大于时间阈值时，表示处理时间比较长，因此，通过将任务处理时间大于时间阈值的多个待处理任务分配到不同的图形处理器，从而保证每个待处理任务被及时调度，对
25 所有图形处理器的资源进行合理分配和利用，防止任务处理超时，保证处理时间比较大的任务不在同一时刻占用同一个图形处理器的多个处理线程。

情况二、根据图形处理器的资源信息，从多个图形处理器中选择与待处理任务对应的目标图形处理器，可以包括：根据图形处理器的资源信息确定空闲处理线程的数量，并从多个图形处理器中选择空闲处理线程最多的图形处理器作为目标图形处理器。由于空闲处理线程数量最多的图形处理器，是资源利用最少的图形处理器，因此，选择空闲处理线程数量最多的图形处理器作为目标图形处理器，可以对所有图形处理器的资源进行合理分配和利用。

情况三、根据图形处理器的资源信息，从多个图形处理器中选择与待处理任务对应的目标图形处理器，可以包括：根据图形处理器的资源信息确定空闲处理线程，并将空闲处理线程对应的图形处理器确定为目标图形处理器。由于是将空闲处理线程对应的图形处理器确定为目标图形处理器，因此将待处理任务分配给目标图形处理器后，目标图形处理器具有处理待处理任务的空闲处理线程，避免目标图形处理器不具有处理待处理任务的空闲处理线程。

其中，当处理线程的状态为空闲状态时，则表示该处理线程是空闲处理线程，当处理线程的状态为占用状态时，则表示该处理线程是占用处理线程。显然，由于是将空闲处理线程对应的图形处理器确定为目标图形处理器，因此，在处理线程空闲时才分配待处理任务，保证待处理任务能够被及时处理。

情况四、根据待处理任务的状态信息和每个图形处理器的资源信息，从多个图形处理器中选择与该待处理任务对应的目标图形处理器，包括：针对任务处理时间大于时间阈值的多个待处理任务，为不同的待处理任务选择不同的目标图形处理器；在选择目标图形处理器时，从多个图形处理器中选择空闲处理线程较多的图形处理器，或选择具有空闲处理线程的图形处理器。

实施例七：

在步骤 202 中，从多个图形处理器中选择与该待处理任务对应的至少一个目标图形处理器，可以包括：针对并行处理的多个待处理任务，为多个待处理任务分别选择对应的目标图形处理器，不同待处理任务对应的目标图形处理器相同或者不同。例如，当不同待处理任务对应不同的目标图形处理器

时，可以将并行处理的多个待处理任务分别到多个图形处理器，从而对所有图形处理器的资源进行合理分配和利用，保证每个待处理任务得到及时处理。

在步骤 202 中，从多个图形处理器中选择与该待处理任务对应的至少一个目标图形处理器，可以包括：若处理队列是静态处理队列，则可以从多个图形处理器中查询与该静态处理队列对应的图形处理器（在上述实施例已经介绍，可以预先配置静态处理队列与图形处理器的对应关系，基于所述对应关系，就可以查询到与该静态处理队列对应的图形处理器），并将查询到的图形处理器确定为该待处理任务对应的目标图形处理器。若处理队列是动态处理队列，则从所有图形处理器中选择目标图形处理器，参见上述实施例。

10 实施例八：

在步骤 204 中，通过该目标图形处理器处理该待处理任务，可以包括：目标图形处理器可以利用图像处理、特征点跟踪、半全局立体块匹配、雷达与相机自标定、点云跟踪、本地地图、深度学习等算法，对待处理任务进行处理，对此过程不做限制。例如，在利用图像处理算法对待处理任务进行处理时，可以对源图像进行格式转换、畸变校正，输出预期格式或者特殊格式的图像数据。又例如，在利用特征点跟踪算法对待处理任务进行处理时，可以利用图像序列中像素在时间域上的变化、相邻帧之间的相关性，获得上一帧与当前帧之间的对应关系，从而计算相邻帧之间物体的运动信息。又例如，在利用深度学习算法对待处理任务进行处理时，可以根据同一目标对象在双目成像中的视差和双目之间的位置关系，计算目标对象相对于相机的深度。

以下结合具体情况，对通过目标图形处理器处理待处理任务进行说明。

情况一、在通过目标图形处理器处理待处理任务（如低优先级的待处理任务）时，若存在比低优先级的待处理任务的优先级更高的待处理任务，则中断低优先级的待处理任务，通过目标图形处理器处理优先级更高的待处理任务；在优先级更高的待处理任务处理完成后，恢复低优先级的待处理任务。

参见图 3 所示，假设目标图形处理器正在处理待处理任务 1，此时接收到一个新的待处理任务 2。目标图形处理器判断是否存在空闲处理线程。如果存

在，则将待处理任务 2 分配给空闲处理线程。如果不存在，则比较待处理任务 1 的优先级与待处理任务 2 的优先级。若待处理任务 1 的优先级高，则继续处理待处理任务 1，待处理任务 2 处于等待状态；在待处理任务 1 处理完成后，处理待处理任务 2。若待处理任务 2 的优先级高，则可以中断待处理任务 1，处理待处理任务 2；在待处理任务 2 处理完成后，恢复待处理任务 1。

情况二、在通过目标图形处理器处理待处理任务时，如果待处理任务出现异常，则中断该待处理任务，提高该待处理任务的优先级，将该待处理任务缓存到处理队列中。由于该待处理任务的优先级被提高，使得该待处理任务能够被优先从处理队列中选取出来，避免该待处理任务出现等待超时。

情况三、通过目标图形处理器并行处理分配给该目标图形处理器的多个待处理任务；其中，并行处理可以包括同步串行处理、内核异步处理。

其中，在目标图形处理器处理多个待处理任务时，可以采用同步串行处理或者内核（kernel）异步处理等方式，对多个待处理任务进行并行处理，对此处理方式不做限制，从而可以有效的节省处理时间，并提高处理效率。

情况四、在目标图形处理器对待处理任务进行处理时，对中央处理器的地址进行锁存，并通过 DMA 控制器传输目标图形处理器与中央处理器之间的交互数据。例如，当待处理任务的数据量较大时，通过调用 `cudaHostRegister` 接口对中央处理器的地址进行页锁存，并通过 DMA 控制器实现中央处理器与图形处理器之间的数据交互，从而显著提高带宽，减少非必要的拷贝。

情况五、在目标图形处理器对待处理任务进行处理时，通过不同的处理线程进行数据共享。具体的，如果多个待处理任务由多个处理线程完成，可以进行处理线程之间的数据共享，避免经过多次内存与显存的拷贝，可以在处理线程内实现显存与内存的共享，省去拷贝过程，可以节省大量处理时间。

本实施例中，目标图形处理器可以利用图像处理、特征点跟踪、半全局立体块匹配、雷达与相机自标定、点云跟踪、本地地图、深度学习等算法，对待处理任务进行处理，因此，处理线程可以包括：图像处理线程、特征点跟踪处理线程、半全局立体块匹配处理线程（Semi-Global Block Matching，简

称 SGBM)、雷达与相机自标定处理线程(后续称为自校准处理线程)、点云跟踪处理线程、地图处理线程、深度学习处理线程(Deep Learning)。

其中,图像处理线程用于接收原始数据,并对原始数据进行处理,得到去畸变和极线矫正后的灰度图像和 RGB 图像。特征点跟踪处理线程用于对图像进行特征点检测和跟踪。半全局立体块匹配处理线程用于对双目图像进行匹配,得到视差图(disparity map),并计算出三维点云。自校准处理线程用于标定相机与相机,雷达与相机之间的外参。点云跟踪处理线程用于将三维点云分割为不同的物体,并进行目标跟踪和区域检测,将结果发送给地图处理线程。深度学习处理线程用于对 RGB 图像进行检测并跟踪目标,将结果发送给地图处理线程。地图处理线程用于接收点云跟踪处理线程的处理结果和深度学习处理线程的处理结果,并利用接收到的信息生成局部地图。

参见图 4 所示,为通过不同的处理线程进行数据共享的示意图。

其中,图像处理线程对待处理任务进行处理时,还可以将输出数据提供给自校准处理线程,由所述自校准处理线程根据输出数据对待处理任务进行处理;或者,可以将输出数据提供给特征点跟踪处理线程,由所述特征点跟踪处理线程根据输出数据对待处理任务进行处理;或者,可以将输出数据提供给半全局立体块匹配处理线程,由所述半全局立体块匹配处理线程根据输出数据对待处理任务进行处理;或者,可以将输出数据提供给深度学习处理线程,由所述深度学习处理线程根据输出数据对待处理任务进行处理。

其中,在半全局立体块匹配处理线程对待处理任务进行处理时,还可以将输出数据提供给点云跟踪处理线程,由所述点云跟踪处理线程根据输出数据对待处理任务进行处理。进一步的,在点云跟踪处理线程对待处理任务进行处理时,还可以将输出数据提供给地图处理线程;此外,在深度学习处理线程对待处理任务进行处理时,也可以将输出数据提供给地图处理线程;基于此,所述地图处理线程可以根据所述点云跟踪处理线程提供的输入数据和所述深度学习处理线程提供的输入数据,对待处理任务进行处理。

实施例八:

基于与上述方法同样的发明构思，本发明实施例还提供一种任务处理设备，如图 5 所示，所述任务处理设备包括调度器和多个图形处理器；所述调度器，用于从处理队列中选择待处理任务，并从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器，并将所述待处理任务分配给所述目标图形处理器；所述目标图形处理器，用于处理所述待处理任务。

所述调度器从处理队列中选择待处理任务时具体用于：若处理队列是静态处理队列，则从所述静态处理队列中选择时间序列和/或数据具有依赖关系的待处理任务。

所述调度器从处理队列中选择待处理任务时具体用于：若处理队列是动态处理队列，则从动态处理队列中选择时间序列和/或数据独立的待处理任务。

所述调度器还用于接收到待处理任务；若所述待处理任务是时间序列和/或数据具有依赖关系的静态调度任务，则将所述待处理任务缓存到静态处理队列；若所述待处理任务是时间序列和/或数据独立的动态调度任务，则将所述待处理任务缓存到动态处理队列。

所述调度器还用于接收待处理任务，将待处理任务缓存到任务队列；针对任务队列中的待处理任务，若待处理任务是时间序列和/或数据具有依赖关系的静态调度任务，将待处理任务缓存到静态处理队列；若待处理任务是时间序列和/或数据独立的动态调度任务，将待处理任务缓存到动态处理队列。

所述调度器从处理队列中选择待处理任务时用于：获取处理队列中待处理任务的优先级；基于优先级优先从处理队列中选择高优先级的待处理任务。

所述调度器获取处理队列中待处理任务的优先级时具体用于：获取所述处理队列中的待处理任务的任务类型；通过所述任务类型查询映射表，得到所述任务类型对应的优先级；映射表用于记录任务类型与优先级的对应关系。

在一个例子中，所述调度器从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器。

所述调度器还用于获取所述待处理任务的任务类型；通过所述任务类型查询状态信息表，得到与所述任务类型对应的状态信息；其中，所述状态信息表用于记录任务类型与状态信息的对应关系；将得到的状态信息确定为所述待处理任务的状态信息。

5 在一个例子中，参见图 4 所示，所述任务处理设备还包括：监测器，用于获取已经完成处理的待处理任务的状态信息，并在状态信息表中记录所述待处理任务的任务类型与所述待处理任务的状态信息的对应关系，并将所述状态信息表发送给所述调度器。

10 在上述实施例中，所述状态信息包括任务处理时间；任务处理时间是任务完成时间与任务接收时间的时间差。

所述监测器，还用于获取图形处理器的资源信息，将所述图形处理器的资源信息发送给所述调度器；所述调度器还用于获取图形处理器的资源信息；

其中，所述资源信息包括：空闲处理线程的数量；和/或，处理线程的状态；所述状态为占用状态或者空闲状态。

15 所述调度器根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：根据所述状态信息确定任务处理时间，针对任务处理时间大于时间阈值的多个待处理任务，为不同的待处理任务选择不同的目标图形处理器。

20 所述调度器根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：根据所述资源信息确定空闲处理线程的数量，并从多个图形处理器中选择空闲处理线程最多的图形处理器作为目标图形处理器。

25 所述调度器根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：根据所述资源信息确定空闲处理线程，并将所述空闲处理线程对应的图形处理器确定为所述待处理任务对应的目标图形处理器。

在一个例子中，所述调度器从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：针对并行处理的多个待处理任务，为所述多个待处理任务分别选择对应的目标图形处理器，不同待处理任务对应的目标图形处理器相同或者不同。

5 所述调度器从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：若所述处理队列是静态处理队列，则从所述多个图形处理器中查询与所述静态处理队列对应的图形处理器，并将查询到的图形处理器确定为所述待处理任务对应的目标图形处理器。

10 所述目标图形处理器处理所述待处理任务时具体用于：若存在比所述待处理任务的优先级更高的待处理任务，则中断所述待处理任务，并处理所述优先级更高的待处理任务；在所述优先级更高的待处理任务处理完成后，恢复所述待处理任务。所述目标图形处理器处理所述待处理任务时具体用于：如果所述待处理任务出现异常，则中断所述待处理任务，并提高所述待处理任务的优先级，并将所述待处理任务缓存到所述处理队列。所述目标图形处
15 理器处理所述待处理任务时具体用于：并行处理分配给所述目标图形处理器的多个待处理任务；其中，所述并行处理包括同步串行处理、内核异步处理。

所述目标图形处理器处理所述待处理任务时具体用于：在对所述待处理任务进行处理时，对中央处理器的地址进行锁存，通过 DMA 控制器传输所述目标图形处理器与中央处理器之间的交互数据。

20 所述目标图形处理器处理所述待处理任务时具体用于：在对所述待处理任务进行处理时，通过不同的处理线程进行数据共享。

所述目标图形处理器通过不同的处理线程进行数据共享时具体用于：

在图像处理线程对待处理任务进行处理时，将输出数据提供给自校准处理线程，由自校准处理线程根据所述输出数据对待处理任务进行处理；或者，
25 将输出数据提供给特征点跟踪处理线程，由所述特征点跟踪处理线程根据所述输出数据对待处理任务进行处理；或者，

将输出数据提供给半全局立体块匹配处理线程，由所述半全局立体块匹

配处理线程根据所述输出数据对待处理任务进行处理；或者，

将输出数据提供给深度学习处理线程，由所述深度学习处理线程根据所述输出数据对待处理任务进行处理。

所述目标图形处理器通过不同的处理线程进行数据共享时具体用于：在
5 半全局立体块匹配处理线程对待处理任务进行处理时，将输出数据提供给点
云跟踪处理线程，由点云跟踪处理线程根据所述输出数据对待处理任务进行
处理。所述目标图形处理器通过不同的处理线程进行数据共享时具体用于：
在点云跟踪处理线程对待处理任务进行处理时，将输出数据提供给地图处理
线程；在深度学习处理线程对待处理任务进行处理时，将输出数据提供给地
10 图处理线程；所述地图处理线程根据点云跟踪处理线程提供的输入数据和深
度学习处理线程提供的输入数据，对待处理任务进行处理。

实施例九：

基于与上述方法同样的发明构思，本发明实施例中还提供一种计算机可
读存储介质，所述计算机可读存储介质上存储有若干计算机指令，所述计算
15 机指令被执行时，实现权利要求上述的任务处理方法。

上述实施例阐明的系统、装置、模块或单元，可以由计算机芯片或实体
实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机，
计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、
智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游
20 戏控制台、平板计算机、可穿戴设备或者这些设备中的任意几种设备的组合。

为了描述的方便，描述以上装置时以功能分为各种单元分别描述。当然，
在实施本发明时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

本领域内的技术人员应明白，本发明实施例可提供为方法、系统、或计
算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或
25 结合软件和硬件方面的实施例的形式。而且，本发明实施例可采用在一个或
多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于

磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可以由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其它可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其它可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

而且,这些计算机程序指令也可以存储在能引导计算机或其它可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或者多个流程和/或方框图一个方框或者多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其它可编程数据处理设备,使得在计算机或者其它可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其它可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

本领域技术人员应明白,本发明的实施例可提供为方法、系统或计算机程序产品。因此,本发明可以采用完全硬件实施例、完全软件实施例、或者结合软件和硬件方面的实施例的形式。而且,本发明可以采用在一个或者多个其中包含有计算机可用程序代码的计算机可用存储介质(可以包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

以上所述仅为本发明实施例而已,并不用于限制本发明。对于本领域技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原理之内所作的任何修改、等同替换、改进,均应包含在本发明的权利要求范围之内。

权利要求书

1、一种任务处理方法，其特征在于，应用于包括多个图形处理器的设备，所述方法包括：

从处理队列中选择待处理任务；

5 从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器；将所述待处理任务分配给所述目标图形处理器；

通过所述目标图形处理器处理所述待处理任务。

2. 根据权利要求 1 所述的方法，其特征在于，

所述从处理队列中选择待处理任务，包括：

10 若所述处理队列是静态处理队列，则从所述静态处理队列中选择时间序列和/或数据具有依赖关系的待处理任务。

3. 根据权利要求 1 所述的方法，其特征在于，

所述从处理队列中选择待处理任务，包括：若所述处理队列是动态处理队列，则从所述动态处理队列中选择时间序列和/或数据独立的待处理任务。

15 4. 根据权利要求 2 或 3 所述的方法，其特征在于，

所述从处理队列中选择待处理任务之前，还包括：

接收到待处理任务；

若所述待处理任务是时间序列和/或数据具有依赖关系的静态调度任务，将所述待处理任务缓存到静态处理队列；

20 若所述待处理任务是时间序列和/或数据独立的动态调度任务，将所述待处理任务缓存到动态处理队列。

5. 根据权利要求 2 或 3 所述的方法，其特征在于，

所述从处理队列中选择待处理任务之前，还包括：

接收待处理任务，将所述待处理任务缓存到任务队列；

25 针对任务队列中的待处理任务，若所述待处理任务是时间序列和/或数据具有依赖关系的静态调度任务，则将所述待处理任务缓存到静态处理队列；

若所述待处理任务是时间序列和/或数据独立的动态调度任务，则将所述

待处理任务缓存到动态处理队列。

6. 根据权利要求 1 所述的方法, 其特征在于,
所述从处理队列中选择待处理任务, 包括:

获取所述处理队列中的待处理任务的优先级;

5 基于所述优先级优先从所述处理队列中选择高优先级的待处理任务。

7. 根据权利要求 6 所述的方法, 其特征在于,
所述获取所述处理队列中的待处理任务的优先级, 包括:

获取所述处理队列中的待处理任务的任务类型;

通过所述任务类型查询映射表, 得到所述任务类型对应的优先级;

10 其中, 所述映射表用于记录任务类型与优先级的对应关系。

8. 根据权利要求 1 所述的方法, 其特征在于, 所述从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器, 包括:

根据所述待处理任务的状态信息和/或图形处理器的资源信息, 从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器。

15 9. 根据权利要求 8 所述的方法, 其特征在于, 所述根据所述待处理任务的状态信息和/或图形处理器的资源信息, 从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器之前, 还包括:

获取所述待处理任务的任务类型;

20 通过所述任务类型查询状态信息表, 得到与所述任务类型对应的状态信息; 其中, 所述状态信息表用于记录任务类型与状态信息的对应关系;

将得到的状态信息确定为所述待处理任务的状态信息。

10. 根据权利要求 9 所述的方法, 其特征在于, 所述通过所述任务类型查询状态信息表, 得到与所述任务类型对应的状态信息之前, 还包括:

25 获取已经完成处理的待处理任务的状态信息, 并在状态信息表中记录所述待处理任务的任务类型与所述待处理任务的状态信息的对应关系。

11. 根据权利要求 8 所述的方法, 其特征在于, 所述状态信息包括任务处理时间; 任务处理时间是任务完成时间与任务接收时间的时间差。

12. 根据权利要求 8 所述的方法，其特征在于，所述根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与
所述待处理任务对应的至少一个目标图形处理器之前，还包括：

获取图形处理器的资源信息；其中，所述资源信息包括：空闲处理线程
5 的数量；和/或，处理线程的状态；所述状态为占用状态或者空闲状态。

13. 根据权利要求 8 所述的方法，其特征在于，所述根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与
所述待处理任务对应的至少一个目标图形处理器，包括：

根据所述状态信息确定任务处理时间，针对任务处理时间大于时间阈值
10 的多个待处理任务，为不同的待处理任务选择不同的目标图形处理器。

14. 根据权利要求 8 所述的方法，其特征在于，所述根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与
所述待处理任务对应的至少一个目标图形处理器，包括：

根据所述资源信息确定空闲处理线程的数量，并从所述多个图形处理器
15 中选择空闲处理线程最多的图形处理器作为所述目标图形处理器。

15. 根据权利要求 8 所述的方法，其特征在于，所述根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与
所述待处理任务对应的至少一个目标图形处理器，包括：

根据所述资源信息确定空闲处理线程，并将所述空闲处理线程对应的图
20 形处理器确定为所述待处理任务对应的目标图形处理器。

16. 根据权利要求 1 所述的方法，其特征在于，所述从所述多个图形处理器中选择与
所述待处理任务对应的至少一个目标图形处理器，包括：

针对并行处理的多个待处理任务，为所述多个待处理任务分别选择对应的
目标图形处理器，不同待处理任务对应的目标图形处理器相同或者不同。

25 17. 根据权利要求 1 所述的方法，其特征在于，所述从所述多个图形处理器中选择与
所述待处理任务对应的至少一个目标图形处理器，包括：

若所述处理队列是静态处理队列，则从所述多个图形处理器中查询与所

述静态处理队列对应的图形处理器，并将查询到的图形处理器确定为所述待处理任务对应的目标图形处理器。

18. 根据权利要求 1 所述的方法，其特征在于，在通过所述目标图形处理器处理所述待处理任务时，所述方法还包括：

5 若存在比所述待处理任务的优先级更高的待处理任务，则中断所述待处理任务，通过所述目标图形处理器处理所述优先级更高的待处理任务；

在所述优先级更高的待处理任务处理完成后，恢复所述待处理任务。

19. 根据权利要求 1 所述的方法，其特征在于，在通过所述目标图形处理器处理所述待处理任务时，所述方法还包括：

10 如果所述待处理任务出现异常，则中断所述待处理任务，并提高所述待处理任务的优先级，并将所述待处理任务缓存到所述处理队列中。

20. 根据权利要求 1 所述的方法，其特征在于，所述通过所述目标图形处理器处理所述待处理任务，包括：

15 通过所述目标图形处理器并行处理分配给所述目标图形处理器的多个待处理任务；其中，所述并行处理包括同步串行处理、内核异步处理。

21. 根据权利要求 1 所述的方法，其特征在于，所述通过所述目标图形处理器处理所述待处理任务，包括：

在对所述待处理任务进行处理时，对中央处理器的地址进行锁存，并通过 DMA 控制器传输所述目标图形处理器与所述中央处理器之间的交互数据。

20 22. 根据权利要求 1 所述的方法，其特征在于，所述通过所述目标图形处理器处理所述待处理任务，包括：

在对所述待处理任务进行处理时，通过不同的处理线程进行数据共享。

23. 根据权利要求 22 所述的方法，其特征在于，所述通过不同的处理线程进行数据共享，包括：

25 在图像处理线程对待处理任务进行处理时，将输出数据提供给自校准处理线程，由自校准处理线程根据所述输出数据对待处理任务进行处理；或者，将输出数据提供给特征点跟踪处理线程，由所述特征点跟踪处理线程根

据所述输出数据对待处理任务进行处理；或者，

将输出数据提供给半全局立体块匹配处理线程，由所述半全局立体块匹配处理线程根据所述输出数据对待处理任务进行处理；或者，

5 将输出数据提供给深度学习处理线程，由所述深度学习处理线程根据所述输出数据对待处理任务进行处理。

24. 根据权利要求 22 所述的方法，其特征在于，所述通过不同的处理线程进行数据共享，包括：在半全局立体块匹配处理线程对待处理任务进行处理时，将输出数据提供给点云跟踪处理线程，由所述点云跟踪处理线程根据所述输出数据对待处理任务进行处理。

10 25. 根据权利要求 22 所述的方法，其特征在于，所述通过不同的处理线程进行数据共享，包括：

在点云跟踪处理线程对待处理任务进行处理时，将输出数据提供给地图处理线程；在深度学习处理线程对待处理任务进行处理时，将输出数据提供给地图处理线程；所述地图处理线程根据所述点云跟踪处理线程提供的输入
15 数据和所述深度学习处理线程提供的输入数据，对待处理任务进行处理。

26. 一种任务处理设备，其特征在于，包括调度器和多个图形处理器；

所述调度器，用于从处理队列中选择待处理任务，并从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器，并将所述待处理任务分配给所述目标图形处理器；

20 所述目标图形处理器，用于处理所述待处理任务。

27. 根据权利要求 26 所述的设备，其特征在于，所述调度器从处理队列中选择待处理任务时具体用于：若处理队列是静态处理队列，则从所述静态处理队列中选择时间序列和/或数据具有依赖关系的待处理任务。

25 28. 根据权利要求 26 所述的设备，其特征在于，所述调度器从处理队列中选择待处理任务时具体用于：若处理队列是动态处理队列，则从所述动态处理队列中选择时间序列和/或数据独立的待处理任务。

29. 根据权利要求 27 或 28 所述的设备, 其特征在于,

所述调度器还用于接收到待处理任务; 若所述待处理任务是时间序列和/或数据具有依赖关系的静态调度任务, 则将所述待处理任务缓存到静态处理队列; 若所述待处理任务是时间序列和/或数据独立的动态调度任务, 则将所述待处理任务缓存到动态处理队列。

30. 根据权利要求 27 或 28 所述的设备, 其特征在于, 所述调度器还用于接收待处理任务, 将所述待处理任务缓存到任务队列; 针对任务队列中的待处理任务, 若所述待处理任务是时间序列和/或数据具有依赖关系的静态调度任务, 将所述待处理任务缓存到静态处理队列; 若所述待处理任务是时间序列和/或数据独立的动态调度任务, 将所述待处理任务缓存到动态处理队列。

31. 根据权利要求 26 所述的设备, 其特征在于, 所述调度器从处理队列中选择待处理任务时具体用于: 获取所述处理队列中的待处理任务的优先级; 基于所述优先级优先从所述处理队列中选择高优先级的待处理任务。

32. 根据权利要求 31 所述的设备, 其特征在于, 所述调度器获取所述处理队列中的待处理任务的优先级时具体用于: 获取所述处理队列中的待处理任务的任务类型; 通过所述任务类型查询映射表, 得到所述任务类型对应的优先级; 其中, 所述映射表用于记录任务类型与优先级的对应关系。

33. 根据权利要求 26 所述的设备, 其特征在于, 所述调度器从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于: 根据所述待处理任务的状态信息和/或图形处理器的资源信息, 从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器。

34. 根据权利要求 33 所述的设备, 其特征在于, 所述调度器还用于获取所述待处理任务的任务类型; 通过所述任务类型查询状态信息表, 得到与所述任务类型对应的状态信息; 其中, 所述状态信息表用于记录任务类型与状态信息的对应关系; 将得到的状态信息确定为所述待处理任务的状态信息。

35. 根据权利要求 34 所述的设备, 其特征在于, 还包括:

监测器, 用于获取已经完成处理的待处理任务的状态信息, 并在状态信

息表中记录所述待处理任务的任务类型与所述待处理任务的状态信息的对应关系，并将所述状态信息表发送给所述调度器。

36. 根据权利要求 33 所述的设备，其特征在于，所述状态信息包括任务处理时间；任务处理时间是任务完成时间与任务接收时间的时间差。

5 37. 根据权利要求 33 所述的设备，其特征在于，还包括监测器，用于获取图形处理器的资源信息，将所述图形处理器的资源信息发送给所述调度器；

所述调度器还用于获取图形处理器的资源信息；

其中，所述资源信息包括：空闲处理线程的数量；和/或，处理线程的状态；所述状态为占用状态或者空闲状态。

10 38. 根据权利要求 33 所述的设备，其特征在于，所述调度器根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：

根据所述状态信息确定任务处理时间，针对任务处理时间大于时间阈值的多个待处理任务，为不同的待处理任务选择不同的目标图形处理器。

15 39. 根据权利要求 33 所述的设备，其特征在于，所述调度器根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：

根据所述资源信息确定空闲处理线程的数量，并从所述多个图形处理器中选择空闲处理线程最多的图形处理器作为所述目标图形处理器。

20 40. 根据权利要求 33 所述的设备，其特征在于，所述调度器根据所述待处理任务的状态信息和/或图形处理器的资源信息，从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：

根据所述资源信息确定空闲处理线程，并将所述空闲处理线程对应的图形处理器确定为所述待处理任务对应的目标图形处理器。

25 41. 根据权利要求 26 所述的设备，其特征在于，所述调度器从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：针对并行处理的多个待处理任务，为所述多个待处理任务分别选择对

应的目标图形处理器，不同待处理任务对应的目标图形处理器相同或者不同。

42. 根据权利要求 26 所述的设备，其特征在于，

所述调度器从所述多个图形处理器中选择与所述待处理任务对应的至少一个目标图形处理器时具体用于：若所述处理队列是静态处理队列，则从所述多个图形处理器中查询与所述静态处理队列对应的图形处理器，并将查询到的图形处理器确定为所述待处理任务对应的目标图形处理器。

43. 根据权利要求 26 所述的设备，其特征在于，所述目标图形处理器处理所述待处理任务时具体用于：若存在比所述待处理任务的优先级更高的待处理任务，则中断所述待处理任务，并处理所述优先级更高的待处理任务；在所述优先级更高的待处理任务处理完成后，恢复所述待处理任务。

44. 根据权利要求 26 所述的设备，其特征在于，

所述目标图形处理器处理所述待处理任务时具体用于：如果所述待处理任务出现异常，则中断所述待处理任务，并提高所述待处理任务的优先级，并将所述待处理任务缓存到所述处理队列中。

45. 根据权利要求 26 所述的设备，其特征在于，所述目标图形处理器处理所述待处理任务时具体用于：并行处理分配给所述目标图形处理器的多个待处理任务；其中，所述并行处理包括同步串行处理、内核异步处理。

46. 根据权利要求 26 所述的设备，其特征在于，

所述目标图形处理器处理所述待处理任务时具体用于：在对所述待处理任务进行处理时，对中央处理器的地址进行锁存，通过 DMA 控制器传输所述目标图形处理器与中央处理器之间的交互数据。

47. 根据权利要求 26 所述的设备，其特征在于，

所述目标图形处理器处理所述待处理任务时具体用于：在对所述待处理任务进行处理时，通过不同的处理线程进行数据共享。

48. 根据权利要求 47 所述的设备，其特征在于，

所述目标图形处理器通过不同的处理线程进行数据共享时具体用于：

在图像处理线程对待处理任务进行处理时，将输出数据提供给自校准处

理线程，由自校准处理线程根据所述输出数据对待处理任务进行处理；或者，

将输出数据提供给特征点跟踪处理线程，由所述特征点跟踪处理线程根据所述输出数据对待处理任务进行处理；或者，

5 将输出数据提供给半全局立体块匹配处理线程，由所述半全局立体块匹配处理线程根据所述输出数据对待处理任务进行处理；或者，

将输出数据提供给深度学习处理线程，由所述深度学习处理线程根据所述输出数据对待处理任务进行处理。

49. 根据权利要求 47 所述的设备，其特征在于，所述目标图形处理器通过不同的处理线程进行数据共享时具体用于：在半全局立体块匹配处理线程
10 对待处理任务进行处理时，将输出数据提供给点云跟踪处理线程，由点云跟踪处理线程根据所述输出数据对待处理任务进行处理。

50. 根据权利要求 47 所述的设备，其特征在于，

所述目标图形处理器通过不同的处理线程进行数据共享时具体用于：

在点云跟踪处理线程对待处理任务进行处理时，将输出数据提供给地图
15 处理线程；在深度学习处理线程对待处理任务进行处理时，将输出数据提供给地图处理线程；所述地图处理线程根据点云跟踪处理线程提供的输入数据和深度学习处理线程提供的输入数据，对待处理任务进行处理。

51、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储有若干计算机指令，所述计算机指令被执行时，实现权利要求 1-25 任
20 一项所述的任务处理方法。

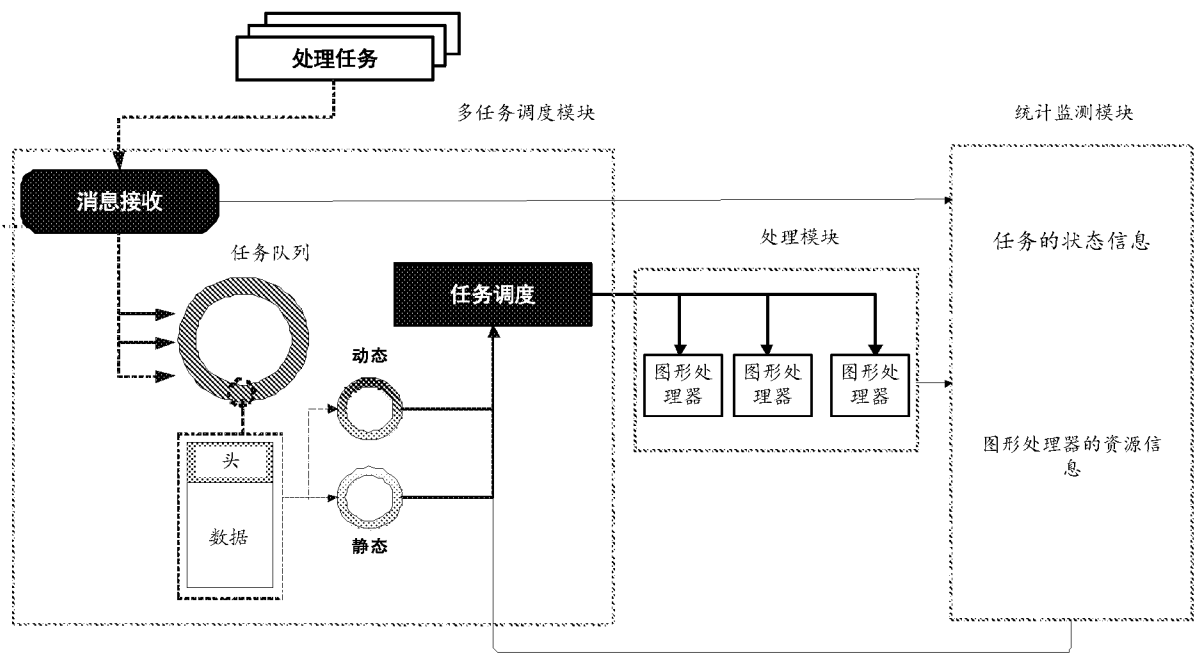


图 1

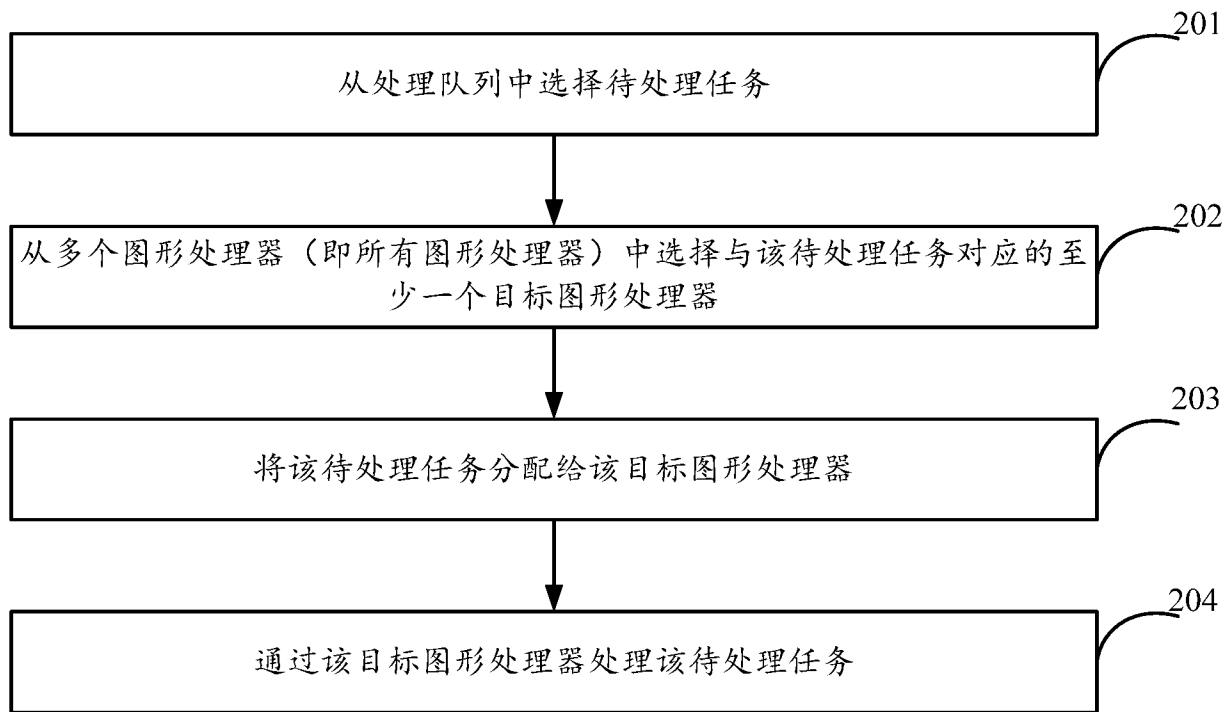


图 2

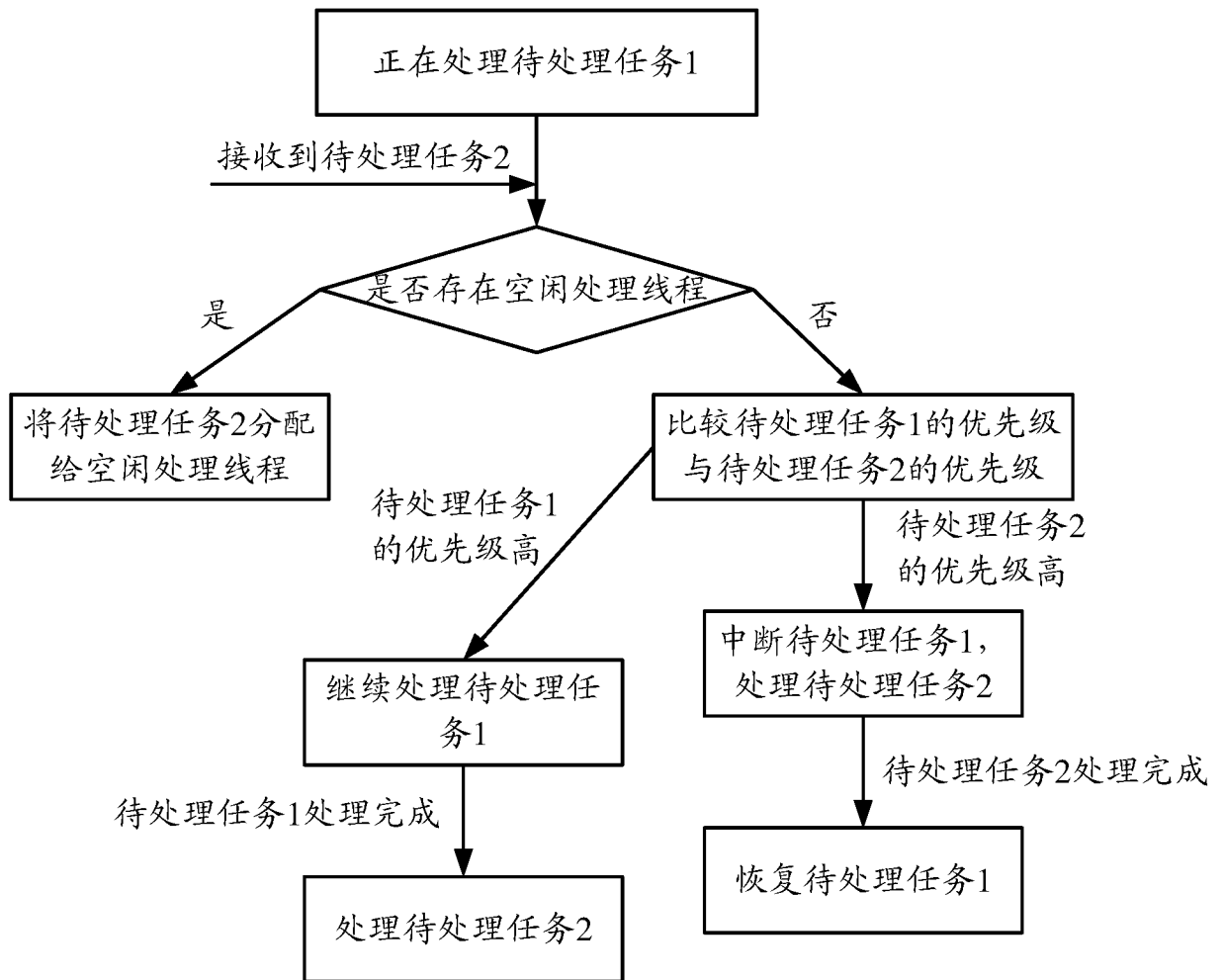


图 3

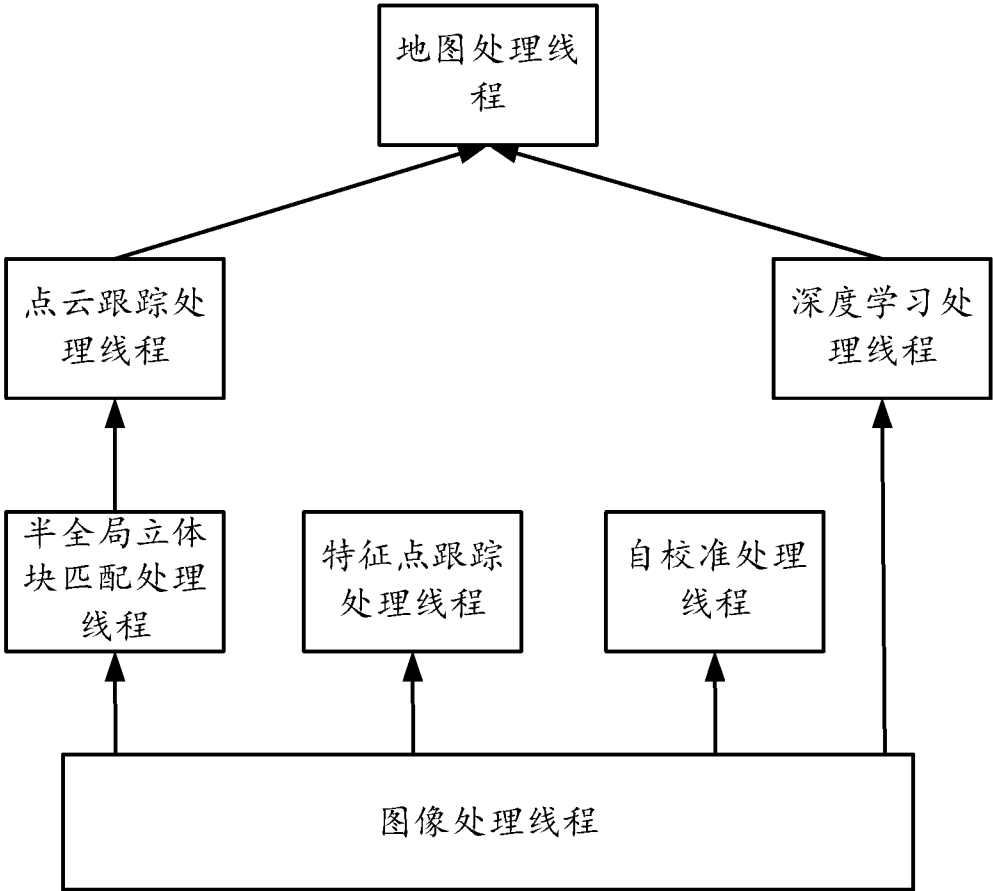


图 4

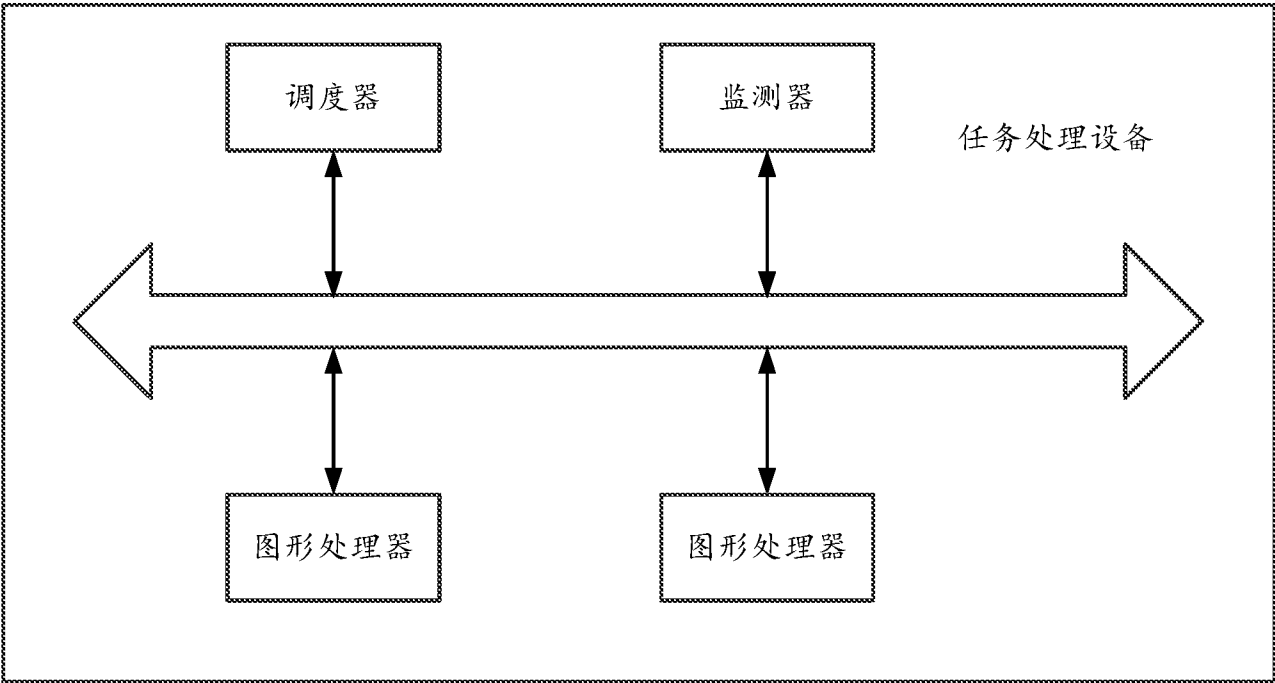


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/080970

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/50(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; VEN; USTXT; EPTXT; WOTXT: 图形处理器, 待处理, 任务, 选择, 优先级, 静态, 动态, 队列, GPU, pending, task, select, priority, static, dynamic, queue

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107122243 A (HANGZHOU YUANSUAN CLOUD CALCULATING CO., LTD. ET AL.) 01 September 2017 (2017-09-01) description, paragraphs [0045]-[0074] and [0153], and figures 1-3	1-51
A	CN 107273331 A (SHANDONG CHAOYUE NC ELECTRONIC CO., LTD.) 20 October 2017 (2017-10-20) entire document	1-51
A	US 2011161675 A1 (NVIDIA CORPORATION) 30 June 2011 (2011-06-30) entire document	1-51

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

29 November 2018

Date of mailing of the international search report

19 December 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/080970

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	107122243	A	01 September 2017	CN 107122243 B	24 July 2018
CN	107273331	A	20 October 2017	None	
US	2011161675	A1	30 June 2011	None	

国际检索报告

国际申请号

PCT/CN2018/080970

A. 主题的分类

G06F 9/50(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNABS;CNTXT;CNKI;VEN;USTXT;EPTXT;WOTXT;图形处理器, 待处理, 任务, 选择, 优先级, 静态, 动态, 队列, GPU, pending, task, select, priority, static, dynamic, queue

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 107122243 A (杭州远算云计算有限公司 等) 2017年 9月 1日 (2017 - 09 - 01) 说明书第45-74段、第153段, 图1-3	1-51
A	CN 107273331 A (山东超越数控电子有限公司) 2017年 10月 20日 (2017 - 10 - 20) 全文	1-51
A	US 2011161675 A1 (NVIDIA CORPORATION) 2011年 6月 30日 (2011 - 06 - 30) 全文	1-51

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2018年 11月 29日

国际检索报告邮寄日期

2018年 12月 19日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

刘细金

电话号码 86-(20)-28950395

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/080970

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	107122243	A	2017年 9月 1日	CN 107122243 B	2018年 7月 24日
CN	107273331	A	2017年 10月 20日	无	
US	2011161675	A1	2011年 6月 30日	无	