

虚拟机放置问题的研究综述

童俊杰 赫 罡 符 刚
(中国联通网络技术研究院 北京 100046)

摘 要 随着云计算数据中心规模和数量的日益增长,以及虚拟化技术的普遍采用,虚拟机放置问题逐步成为产业界和学术界研究的热点。虚拟机放置策略和方法的选择对数据中心的能耗,物理资源的利用率和虚拟机性能具有重大影响。合理的放置方法和策略在保证上层应用和业务不受影响的同时,能有效降低云计算数据中心的能耗,提升物理资源利用率,减少物理资源的浪费。阐述了虚拟机放置问题中的 3 个基本要素:优化目标、约束限制和实现方法,并基于已有的研究工作进行归纳与总结。最后,结合已有成果,展望了未来的研究方向和亟待解决的关键问题。

关键词 虚拟机放置,节能,优化,约束

中图法分类号 TP393 文献标识码 A

Research Survey of Virtual Machine Placement Problem

TONG Jun-jie HE Gang FU Gang

(Network Technology Research Institute of China Unicom, Beijing 100046, China)

Abstract With increasing number and scale of the cloud computing data centers, which generally adopt virtualization technologies, virtual machine placement problem is becoming a hot topic in both industry and academy areas. The choice of policies and methods on virtualization machine placement impacts the energy consumption of data centers, the utilization of resources and the performance of virtual machines. Proper policies and methods protect the up layer applications and service from being affected with lowering energy consumption, increasing resources utilization and decreasing resources wasting. This paper described three essentials of virtual machine placement problem including optimization function, constraint conditions and methodology depending on the existing research works. This paper also presented a summary of the current works and some crucial problems which should be solved pressingly.

Keywords Virtual machine placement, Energy saving, Optimization, Constraints

1 引言

云计算作为一种新的资源供给和计算模式,使用户可以更为简单和便捷地以在线方式按需获取可配置的资源,降低用户上线部署门槛和管理复杂度的同时,带来高可靠、易用和低风险的好处^[1]。该概念自 2006 年提出以来,相关技术便得到了广泛地关注和快速地发展。

当前云计算数据中心大多基于虚拟化技术。虚拟化技术使得多台虚拟机可共享同一物理设备^[2],从而提升资源利用率并节约能耗。云计算数据中心提供 IaaS(Infrastructure as a Service)、PaaS(Platform as a Service)和 SaaS(Software as a Service) 3 种按需服务^[3]。与 3 种服务相关的资源分配和管理问题得到了广泛关注和研究。

虚拟机放置即 VMP(Virtual Machine Placement)问题作为云计算中资源分配和管理的重要组成,在国际会议 CC-Grid、CLOUD、CloudCom 和 INFOCOM 以及 IEEE、ACM 和 Springer 出版的重要期刊上已有不少相关论文发表。该问题指的是根据一定的方法和策略为虚拟机选择合适的 PM (Physical Machine),在满足虚拟机资源需求和约束限制的前

提下,达到提高资源利用率、减少能耗和网络时延等目的。

当前已经有对 VMP 问题的若干综述研究。已有的综述研究主要分为如下 4 个方面:特定目标方面,例如针对 VMP 问题的能耗优化技术^[4,5];场景中云架构方面,例如考虑普通的单数据中心和多数据中心架构^[8]以及特殊架构(例如联合云^[6])中的 VMP 相关技术;放置算法本身,例如大规模云环境下不同放置方法的性能对比^[7];放置方法的设计,例如 VMP 问题特定类别方法(例如启发式算法^[9])

从虚拟机放置方案的部署和应用过程看,其主要部署和应用过程如图 1 所示。

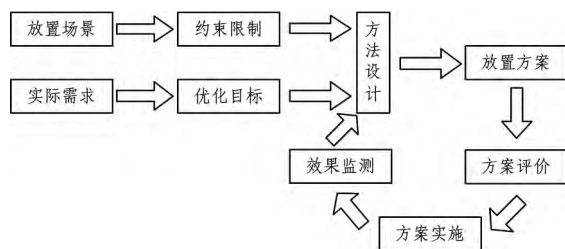


图 1 虚拟机放置方案的部署和应用

根据实际的放置场景和需求,设定相应的放置优化目标

童俊杰(1988—),男,博士,工程师,CCF 会员,主要研究方向为网络功能虚拟化,E-mail: tongjj9@chinaunicom.cn;赫 罡(1974—),男,博士,高级工程师,主要研究方向为 SDN 和核心网虚拟化;符 刚(1977—),男,硕士,高级工程师,主要研究方向为 VoLTE、IMS 和核心网虚拟化。

和约束条件,以作为放置方法设计的输入参数。生成的放置方案需要经过方案评价后方可进入实施,并且在实施后要根据实际检测的效果适当地对放置方案进行调整,以进一步优化放置方案,达到预期效果。

本文主要针对图1过程中的优化目标、约束限制和实现方法3个方面,对VMP问题进行综合阐述,分别对应本文的第2—第4节内容;最后总结全文,指出未来的研究方向并展望后续工作。

2 VMP问题的优化目标

VMP问题的优化目标可以分为单目标和多目标。单目标还可以细分为两类,一类为狭义单目标,该类目标只针对某个具体方面,例如能耗或资源利用率等;另外一类则是广义单目标,这类目标虽然也以单个目标函数的方式呈现,但其中却包含多个目标的组合。多目标则是同时优化多种目标。

多目标是单目标的有机组合,本小节先对单个目标进行解释和总结,然后对多个单目标的组合进行解释和总结。

2.1 单个目标

比较典型的目标有:能耗优化、网络流量优化、资源分配优化和性能优化。

2.1.1 能耗优化

数据中心能耗包括了数据中心各种用能设备能耗的总和。在当前数据中心能耗水平状况下,IT设备的能耗最高,占到总能耗的50%左右,其中服务器能耗占40%左右,存储设备和网络通信设备能耗各占5%左右;空调系统能耗占到40%左右;配电系统能耗占数据中心总能耗的10%左右^[10]。

当前最小化能源的目标,主要也是减少IT设备的能耗,可以通过间接地减少服务器或路由器或同时减少两者的激活数量来达到减少能耗的目的。文献[11,12]分别只考虑了服务器和网络设备的能耗,文献[13,14]同时考虑最小化激活物理机和网络设备来降低能耗水平,文献[15]则考虑跨数据中心场景下数据中心内部服务器和网络设备能耗以及数据中心之间IP层和WDM层网络通信设备能耗。

除了间接方式,也可以通过一定的能耗模型,直接将能耗作为最小化目标,这里的能耗不仅包括服务器和网络通信等设备^[16-18]的能耗,还包括虚拟机迁移等操作上的能耗^[19]。对于能耗模型,物理和网络通信设备主要采用的是基于CPU利用率的线性模型^[16-20]。基于利用率的线性模型,某时刻 t 单个设备的能耗采用式(1)计算。

$$P(t) = C_0 + C_1 \times U(t) \quad (1)$$

式中 C_0 和 C_1 为常量, $U(t)$ 为设备时刻 t 某项资源的利用率,对于服务器来说,一般为CPU的利用率。这样,特定服务器在时间 t_0 到 t_1 内的能耗可通过式(2)计算。

$$E = \int_{t_0}^{t_1} P(t) dt \quad (2)$$

通过减少设备数量来间接减少能源消耗,较为简单直接,但忽略了设备性能和能耗的异质性;通过一定能耗模型直接将能耗作为目标,更为精确,但增加了计算的复杂性。

间接和直接方式主要是针对单个数据中心内的能耗。针对多个分布式数据中心的总能耗,由于不同数据中心外部及内部环境和条件不同(外部环境包括地理位置的温度和当地

电费计价方式^[21,22],内部环境包括数据中心的PUE值等^[23]),影响了能耗和相应的成本。

2.1.2 网络流量优化

单纯考虑虚拟机之间的通信流量,文献[27]通过对VM进行聚类分组达到最小化组件流量、最大化组内流量,从而减少跨机架的通信流量。文献[26]根据策略将有通信需求的VM放置得尽量靠近,以节省网络带宽资源。文献[26,27]中是在假定知晓虚拟机之间确切通信需求基础上进行相关优化,但是在实际中,虚拟机之间的通信流量多变,因此文献[28]则是通过虚拟机迁移,来达到减小网络通信流量的目的。

在业务运行的情况下,虚拟机需要与存储设备或节点进行通信。文献[24]针对的是集中存储时的网络通信开销,文献[25]则是针对采用Hadoop架构运行业务时的最大数据访问时延。

相对于单数据中心的流量考虑,文献[29]则考虑最小化跨数据中心的流量。

总的说来,对于网络流量优化而言,需要根据实际应用和场景在时延和带宽消耗之间进行权衡,较为常见的目标包括:最小化总的访问时延或带宽消耗,最小化最大访问时延,在一定访问时延限制下最小化访问时延或带宽、在一定时间或带宽限制下最小化最大访问时延^[30]。

2.1.3 资源分配优化

虚拟机之间的资源请求和服务器之间的可用物理资源在不同维度资源上均具备差异性,在资源分配时应同时考虑这两种差异性,较常见的目标是最小化资源浪费^[31,34]和最大化资源使用均衡性^[32,33],两者同时考虑时可以达到减少物理机激活数量的目的^[35]。所以,有时也将减少物理机激活数量的目标等同于提升资源使用率。

此外,对于跨数据中心的场景,文献[36]则旨在最小化最大资源利用率。

2.1.4 性能优化

从应用性能的角度,主要是优化应用的计算性能和网络性能,进而减少任务完成的时间。文献[37]主要考虑网络资源对应用的重要性,基于网络流量进行智能的虚拟机放置,通过一定的调度和虚拟机迁移达到提升网络性能的目的。文献[38]中则不仅考虑网络,还考虑虚拟机之间的干扰,从而优化计算性能或网络性能。

从可靠性的角度,主要是保证一个或多个虚拟机、主机、机架甚至数据中心失效时虚拟机的可靠性。文献[39]主要考虑保证虚拟机垂直和水平两种扩展策略下的可靠性。文献[40]则主要是保证 k 个主机失效时虚拟机的可靠性。

从安全性的角度,主要是满足虚拟机对硬件、中间件、客户操作系统、进程以及应用上的安全要求。文献[41]定义了基础设施中的安全指标,并提出相应的虚拟机放置算法以满足特定安全指标要求。

从服务质量的角度,主要是保证尽量满足上层业务和虚拟机对SLA(Service-level Agreement)的要求,尽量减少虚拟机启动数和避免发生SLA冲突等导致云服务器供应商CSP(Cloud Service Provider)的利润受损。文献[42]只考虑资源分配对SaaS层SLA的影响,文献[43]则考虑对IaaS层SLA的影响。文献[44,45]则是分别考虑实时的服务和虚拟机需求对SLA的影响。

2.2 多个单目标的组合

2.2.1 多目标融合为单目标

现有研究普遍采用的方法为线性加权。将采用虚拟化技术的数据中心能耗建模为多个能耗子目标之和,例如文献[13,46]将服务器能耗目标和网络能耗目标线性相加,文献[16]将数据中心总能耗目标和拒绝请求开销目标线性相加。

将多个目标进行融合时,需要对相关问题领域有较深的理解,以保证融合后目标的有效性^[47]。

2.2.2 多目标优化

不同于将多个目标融合为单个目标,多目标优化指的是同时优化多个目标,可以用如下数学模型进行标识:

$$\begin{aligned} & \max/\min f_1(x) \\ & \max/\min f_2(x) \\ & \dots \\ & \max/\min f_k(x) \\ & \text{subject to } x \in S \end{aligned} \quad (3)$$

其中 $k(k \geq 2)$ 为目标的个数, S 为变量的可行域即变量取值的集合, \max/\min 代表目标函数优化的方向。变量的可行域也可以约束条件的形式描述,例如:

$$\begin{aligned} & \text{subject to } g_i(x) \geq 0, i=1, 2, \dots, p \\ & h_i(g)=0, i=1, 2, \dots, q \end{aligned} \quad (4)$$

其中, p 和 q 分别代表约束的个数。

多个目标同时优化时,往往难以存在一个满足所有目标函数的最优解,需要借助其他思路来解决,例如常用的帕累托模型和 ϵ -约束等^[48]。

3 VMP 问题的约束限制

VMP 问题的约束限制主要可分为 4 类:资源类约束、放置类约束、性能约束和可靠性约束。为表述方便,本节先对 VMP 问题进行简单的抽象表示,然后对各类约束进行详细描述。

3.1 VMP 问题的初步描述

给定一系列服务器 $S=\{s_1, s_2, \dots, s_m\}$, 共计 $m(>1)$ 个独立服务器(Physical Machine, PM), 服务器 s_i 的资源容量为 c_i , 一般来说包含资源 c_i^{cpu} 、 c_i^m 、 c_i^b 和 c_i^d 即 CPU、内存、网络带宽和磁盘资源;一系列虚拟机 $V=\{v_1, v_2, \dots, v_n\}$ 共计 $n(>1)$ 个虚拟机,虚拟机 v_i 的资源需求为 d_i , 一般来说包含资源 d_i^{cpu} 、 d_i^m 、 d_i^b 和 d_i^d 即 CPU、内存、网络带宽和磁盘资源的需求。

利用矩阵 X 表示虚拟机的放置情况,该矩阵维度为 $n \times m$, 如果第 i 个虚拟机放置到第 j 个服务器上,则 $x_{ij}=1$ 。

3.2 资源类约束

约束 1:对于任何物理服务器来说,其能满足所有运行在其上的虚拟机的基本资源(包括 CPU、内存、磁盘和带宽)需求时能正常运行。

约束表示: $\forall j=1, 2, \dots, m$, 有 $\sum_{i=1}^n x_{ij} d_i^{cpu} \leq c_j^{cpu} \times r$, $\sum_{i=1}^n x_{ij} d_i^m \leq c_j^m \times r$, $\sum_{i=1}^n x_{ij} d_i^b \leq c_j^b \times r$, $\sum_{i=1}^n x_{ij} d_i^d \leq c_j^d \times r$ 。

约束说明:上述约束中的参数 $r(\leq 1)$ 表示服务器预留一定物理资源,以保证其能正常运行,通常该值设为 0.8^[49]。

约束 2:服务器应满足虚拟机的特殊资源要求。

约束表示:如果 $x_{ij}=1$ 且虚拟机 v_i 需要特殊物理资源,

则 s_i 具备该特殊物理资源且满足虚拟机 v_i 的需求。

约束说明:在特定场景中,特别针对要求较高计算性能或网络转发性能的业务场景,为了保证性能,往往需要 FPGA、SoC 等形态的硬件级加速资源。

约束 3:数据中心应满足虚拟机的总资源需求。

约束表示:假设有一系列可选数据中心 D , 若将虚拟机 V 切分为多个子集后进行分配,应保证单个数据中心资源满足虚拟机需求总和,同时数据中心之间的带宽资源满足虚拟机通信需求。

约束说明:跨数据中心场景下的 VMP 问题^[21,22,29]。

3.3 放置类约束

约束 1:一个虚拟机能且只能在一个物理机上进行部署。

约束表示:对于 $\forall i=1, 2, \dots, n$, 有 $\sum_{j=1}^m x_{ij}=1$ 。

约束说明:一般情况下,认为单个虚拟机只能在一个物理机上进行部署运行,但是也存在虚拟机资源需求过大,对其进行切分后分配资源的情况^[50,51]。

约束 2:任意虚拟机之间可以进行亲和性和反亲和性约束。

约束表示:对于虚拟机 v_i 和虚拟机 v_j , 同服务器的亲和性约束为 (v_i, v_j, SP) ;也可以设定不同服务器但同机架的亲和性约束 $(v_i, v_j, SR) \wedge (v_i, v_j, DP)$;还可设定同数据中心但不同机架和服务器的亲和性约束 $(v_i, v_j, SD) \wedge (v_i, v_j, DR) \wedge (v_i, v_j, DP)$;还可以假定异数据中心 (v_i, v_j, DD) 。

约束说明: SP/DP 表示 Same/Different PM 即同/不同服务器, SR/DR 表示 Same/Different Rack 即同/不同机架, SD/DD 表示 Same/Different DC 即同/不同数据中心。

3.4 性能类约束

约束 1:虚拟机之间的网络延迟在 SLA 所规定的范围内。

约束表示:虚拟机 v_i 和虚拟机 v_j 之间的网络延迟为 (v_i, v_j, μ, δ) , 其中 μ 为平均网络延迟, δ 为网络抖动,即虚拟机之间的任意通信网络延迟的范围为 $(\mu-\delta, \mu+\delta)$ 。

约束说明:网络延迟和网络拓扑紧密相关,当前已有研究主要考虑 4 种典型的数据中心网络拓扑,包括 Fat-Tree、VL2、VL2N-Tree 和 Bcube^[12]。

3.5 可靠性约束

约束 1:应保证虚拟机具备 k 容错特性,即任意 $k-1$ 个物理服务器宕机或失效时,虚拟机的业务不受影响,其服务质量仍在 SLA 规定范围内。

约束表示: $L=\{l_{v_i}\}_{v_i \in V}$ 表示所有虚拟机容错性要求,且 $0 \leq l_{v_i} < m$, 即虚拟机 v_i 通过迁移可以避免 l_{v_i} 个主机失效时带来的业务中断。

约束说明:容错性增加的同时,一般会以牺牲资源为代价,应在容错性和资源利用率两者之间做好均衡。

4 实现方法

已有的主要方法可以分为 4 类:确定性算法、启发式算法、元启发式算法和近似算法。

4.1 确定性算法

在 VMP 问题解决方法中较为常见的确定性算法包括约束规划、线性规划、整数线性规划、混合整数线性规划、伪布尔优化和动态规划等。

确定性较为简单,但存在两个明显缺陷,一个是针对特定场景可能不合适;另外一个也是主要缺陷是可扩展性差,不适合于解决大规模的 VMP 问题^[19]。

4.2 启发式算法

虽然 VMP 问题的实际场景和目标函数不同,但是大多都是 NP-Hard 问题甚至是 NP-完全问题^[11,12,20,25,30]。由于物理服务器和虚拟机数量较多,求解空间大,找到唯一最优解的计算复杂度高,因此通常为了避免较大的开销,采用一些启发式算法来解决 VMP 问题。

较为常见的有首次适应算法(First Fit),即将所有虚拟机放入到单个物理服务器直到其资源耗尽,然后继续选择其它服务器进行放置;首次适应下降算法(First Fit Decreasing),即先将所有虚拟机按照资源需求由大到小进行排序,然后用首次适应算法进行放置;最佳适应算法(Best Fit),即找出放置虚拟机所需物理服务器数量最少的方案,在放置的时候,尽量将虚拟机放置到已经启动的物理服务器中;最佳适应下降算法(Best Fit Decreasing),即先将所有虚拟机按照资源需求由大到小进行排序,然后采用最佳适应算法进行放置^[52]。

除上述经典算法外,较为常见的一类方法分为 3 大步,第一步是根据一定策略对虚拟机进行分组或分类,第二步根据分组或分类的结果进行映射,第三步为按照一定规则对映射进行调优。其中,第一步的分组或分类中,较为常见的方法是基于虚拟机的流量拓扑图,通过最小化组间流量、最大化组内流量减少网络流量,节约带宽^[12,13,27,53,54]。第二步的映射过程中,可能包含虚拟机组到机架的映射和虚拟机到物理服务器的映射^[12,27],也可直接由虚拟机组映射到物理服务器^[13,53,54]。第三步的优化过程中,则主要是以目标函数为目的的进行调优。

4.3 元启发式算法

在解决 VMP 问题的方法中,较为常见的元启发式包括蚁群算法^[55]、遗传算法^[31,35]、粒子群算法^[35]和模拟退火算法^[56]以及相关改进,该类算法可以较好地找到全局最优解,但是其一般时间复杂度较高,导致可扩展性和算法实时性不强。此外,其对多目标进行优化时,通常将多个目标融合为单目标进行求解,无法做到真正多目标优化。

4.4 近似算法

启发式算法和元启发式算法在一定时间内,可以找到合理的解,但较难衡量所得解的优劣性。近似算法较为适合 NP-Hard 问题,能有效地将解约束在最优解的一定范围内。假设 OPT 为最优解, $f(x)$ 为通过 ρ -近似算法所得的解,则有以下公式成立。

$$\begin{cases} OPT \leq f(x) \leq \rho OPT, & \text{if } \rho > 1 \\ \rho OPT \leq f(x) \leq OPT, & \text{if } \rho < 1 \end{cases} \quad (5)$$

该类算法设计要求较高,当前开展的研究相对较少^[30]。

结束语 本文从优化目标、约束限制和实现方法 3 个方面对数据中心虚拟机放置问题进行了系统的阐述。优化目标的建立需要从具体场景出发,根据实际的需求进行选定,当前的目标建立中主要考虑对能耗和网络流量的优化;约束限制中的资源类约束和放置类约束较为通用,而性能类约束和可靠性约束与上层应用及业务的需求相关;当前的实现方法以启发式算法和元启发式算法为主,而较为普遍采取的手段则是在启发式算法和元启发式算法中利用确定性算法来高效快速地解决问题,并保证解的优化性。VMP 问题中的优化目标

是方向,约束限制是条件,而实现方法则是在一定条件下朝特定方向前进的方式,三者紧密联系,构成了 VMP 问题从定义到解决的基本要素。

虽然已有较多针对 VMP 问题的研究工作,但还存在许多问题需要进一步研究。下面根据对前人工作的总结,并结合自己的理解,给出 VMP 问题中需要进一步研究的问题。

(1)在线的大规模 VMP 问题对算法的计算效率和可扩展性要求较高,已有研究较难以满足。数据中心的数量和规模都在急剧增加,使得 VMP 中的可扩展性问题日益突出。此外,由于任务或业务的请求是一个连续过程,这也意味着虚拟机放置是一个持续过程而非一次性过程,在进行虚拟机连续放置过程中,为了避免重新计算的时空复杂度,应设计合理的在线算法,保证任务和业务执行的时间。

(2)VMP 问题解决方案的测试和评估方案多样,难以统一,缺乏对算法较为通用的验证方法。当前对 VMP 解决方法的测试验证手段多样,主要体现在实验中使用的数据集及其中所包含的虚拟机负载特征。数据集中有模拟的^[24,31,32]、基于实际数据中心所得^[20]和基于测试工具^[19]3 种,而虚拟机负载特征主要服从随机分布、均匀分布^[32]、正态分布^[31,32]、泊松分布、高斯分布和幂律分布^[24]等。对于模拟数据集,主要是通过自定义或参考的虚拟机实例类别来产生特定分布的虚拟机请求^[32,53],较常用的虚拟机实例类别可参考 Amazon EC2。

(3)已有研究关注的是提升物理资源利用率,减少激活的物理设备数量,节省能耗,对虚拟机的容错性和可靠性的考虑较为欠缺。虚拟机的容错手段往往是通过设置虚拟机副本,以牺牲资源的方式来提升虚拟机的容错性和可靠性,这样会降低物理资源利用率。需要设计较为合理的 VMP 方法以在资源利用率和虚拟机可靠性之间做到较好均衡。

参考文献

- [1] Qi Z, Lu C, Raouf B. Cloud computing: state-of-the-art and research challenges [J]. Journal of Internet Services and Applications, 2011, 1(1): 7-18
- [2] Barham P, Dragovic B, Fraser K, et al. Xen and the art of virtualization [J]. ACM SIGOPS Operating Systems Review, 2003, 37(5): 164-177
- [3] Mell P, Grance T. The NIST definition of cloud computing [J]. National Institute of Standards and Technology, 2009, 53(6): 50
- [4] Beloglazov A, Abawajy J, Buyya R. Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing [J]. Future Generation Computer Systems, 2012, 28(5): 755-768
- [5] Salimian L, Safi F. Survey of energy efficient data centers in cloud computing [C]//IEEE/ACM 6th International Conference on Utility and Cloud Computing. Dresden, 2013: 369-374
- [6] Gahlawat M, Sharma P. Survey of virtual machine placement in federated clouds [C]//IACC. Gurgaon, 2014: 735-738
- [7] Mills K, Filliben J, Dabrowski C. Comparing vm-placement algorithms for on-demand clouds [C]//CloudCom. Athens, 2011: 91-98
- [8] Adam Z, Mann. Allocation of Virtual Machines in Cloud Data Centers-A Survey of Problem Models and Optimization Algorithms [J]. CSUR, 2015, 48(1): 11
- [9] Zhan Z H, Liu X F, Gong Y J, et al. Cloud Computing Resource

- Scheduling and a Survey of Its Evolutionary Approaches [J]. CSUR, 2015, 47(4): 63
- [10] 谷力精, 周伏秋, 孟辉. 我国数据中心能耗及能效水平研究[J]. 中国能源, 2010, 32(11): 42-45
- [11] Aroca J A, Anta A F, Mosteiro M A, et al. Power-efficient assignment of virtual machines to physical machines [M]// Adaptive Resource Management and Scheduling for Cloud Computing, 2014: 71-78
- [12] Fang W W, Liang X M, Li S X, et al. VMPlanner: Optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers [J]. Computer Networks, 2013, 57(1): 179-196
- [13] 董健康, 王洪波, 李阳阳, 等. IaaS 环境下改进能源效率和网络性能的虚拟机放置方法[J]. 通信学报, 2014, 35(1): 72-81
- [14] Maolin T, Shenchen P. A Hybrid Genetic Algorithm for the Energy-Efficient Virtual Machine Placement Problem in Data Centers[J]. Neural Process Letters, 2015, 41(2): 211-221
- [15] Burak K, Luca F, Antonio C, et al. Inter-and-Intra Data Center VM-Placement for Energy-Efficient Large-Scale Cloud Systems [C]// GC, California, 2012: 708-713
- [16] Aisan D, Mohan G, Kee C C. Power-efficient resource-guaranteed VM placement and routing for time-aware data center applications [J]. Computer Networks, 2015, 88(9): 249-268
- [17] Chen K Y, Xu Y, Xi K, et al. Intelligent virtual machine placement for cost efficiency in geo-distributed cloud systems [C]// ICC, Budapest, 2013: 3498-3503
- [18] Tang M L, Pan S C. A Hybrid Genetic Algorithm for the Energy-Efficient Virtual Machine Placement Problem in Data Centers [J]. Neural Processing Letters, 2015, 41(2): 211-221
- [19] Zhang Z M, Hsu C C, Morris C. CoolCloud: A Practical Dynamic Virtual Machine Placement Framework for Energy Aware Data Centers [C]// CLOUD, Dubai, 2015: 758-765
- [20] Li X, Qian Z Z, Lu S L, et al. Energy efficient virtual machine placement algorithm with balanced and improved resource utilization in a data center [J]. Mathematical and Computer Modelling, 2013, 58(5/6): 1222-1235
- [21] Le K, Zhang J R, Meng J D, et al. Reducing electricity cost through virtual machine placement in high performance computing clouds [C]// SC, Seattle, 2011: 1-12
- [22] Li Y, Chiu D, Liu C B, et al. Towards Dynamic Pricing-Based Collaborative Optimizations for Green Data Centers [C]// ICDEW, Brisbane, 2013: 272-278
- [23] Khosravi A, Garg S K, Buyya R. Energy and Carbon-Efficient Placement of Virtual Machines in Distributed Cloud Data Centers [C]// Euro-Par, Aachen, 2013: 317-328
- [24] Rami C, Liane L E, Joseph N, et al. Almost Optimal Virtual Machine Placement for Traffic Intense Data Centers [C]// INFOCOM, Turin, 2013: 355-359
- [25] Kuo J J, Yang H H, Tsai M J. Optimal Approximation Algorithm of Virtual Machine Placement for Data Latency Minimization in Cloud Systems [C]// INFOCOM, Toronto, 2014: 1303-1311
- [26] Georgion S, Tsakalozos K, Delis A. Exploiting network-topology awareness for vm placement in iaas clouds [C]// CGC, Karlsruhe, 2013: 151-158
- [27] Jayasinghe D, Pu C, Eilam T, et al. Improving Performance and Availability of Services Hosted on IaaS Clouds with Structural Constraint-aware Virtual Machine Placement [C]// SCC, Washington, 2011: 72-79
- [28] Fuang T P, Hamilton G, Oikonomou K, et al. Implementing Scalable, Network-Aware Virtual Machine Migration for Cloud Data Centers [C]// CLOUD, Santa Clara, 2013: 557-564
- [29] Teyeb H, Balma A, Ben N A, et al. Optimal Virtual Machine Placement in Large-Scale Cloud Systems [C]// CLOUD, Anchorage, 2014: 424-431
- [30] Alicherry M, Lakshman T V. Optimizing Data Access Latencies in Cloud Systems by Intelligent Virtual Machine Placement [C]// INFOCOM, Turin, 2013: 647-655
- [31] Zheng Q H, Li R, Li X Q, et al. A Multi-objective Biogeography-Based Optimization for Virtual Machine Placement [C]// CC-Grid, Shenzhen, 2015: 687-696
- [32] Hieu N T, Francesco M D, Antti Y J. A Virtual Machine Placement Algorithm for Balanced Resource Utilization in Cloud Data Centers [C]// CLOUD, Anchorage, 2014: 474-481
- [33] Lin X, Qian Z Z, Lu S L, et al. Energy Efficient Virtual Machine Placement Algorithm with Balanced and Improved Resource Utilization in a Data Center [J]. Mathematical and Computer Modelling, 2013, 58(5/6): 1222-1235
- [34] Gao Y, Guan H, Qi Z, et al. A Multi-objective Ant Colony System Algorithm for Virtual Machine Placement in Cloud Computing [J]. Journal of Computer and System Sciences, 2013, 79(8): 1230-1242
- [35] Adamuthe A C, Pandharpatte R M, Thampi G T. Multiobjective Virtual Machine Placement in Cloud Environment [C]// Cloud& Ubiquitous Computing& Emerging Technologies, Pune, 2013: 8-13
- [36] Guo Y, Stolyar S, Walid A. Shadow-Routing Based Dynamic Algorithms for Virtual Machine Placement in a Network Cloud [J]. IEEE Transactions on Cloud Computing, 2015(99): 1
- [37] Kakadia D, Kopri N, Varma V. Network-aware Virtual Machine Consolidation for Large Data Centers [C]// Third International Workshop on Network-Aware Data Management, Denver, 2013: 6
- [38] Gupta A, Kale L V, Milojicic D, et al. HPC-Aware VM Placement in Infrastructure Clouds [C]// IC2E, Redwood, 2013: 11-20
- [39] Wang W T, Chen H P, Chen X. An Availability-aware Virtual Machine Placement Approach for Dynamic Scaling of Cloud Applications [C]// UIC/ATC, Fukuoka, 2012: 509-516
- [40] Bin E, Biran O, Boni O, et al. Guaranteeing High Availability Goals for Virtual Machine Placement [C]// ICDCS, Minneapolis, 2011: 700-709
- [41] Caron E, Le A D, Lefray A, et al. Definition of Security Metrics for the Cloud Computing and Security-Aware Virtual Machine Placement Algorithms [C]// CyberC, Beijing, 2013: 125-131
- [42] Wu L L, Garg S K, Versteeg S, et al. SLA-Based Resource Provisioning for Hosted Software-as-a-Service Applications in Cloud Computing Environments [J]. IEEE Transactions on Service Computing, 2014, 7(3): 465-485
- [43] Lin J W, Jen F, Chen C H. Interference-aware Virtual Machine Placement in Cloud Computing Systems [C]// ICCIS, Kuala Lumpur, 2012: 598-603
- [44] Ishakian V, Bestavros A. MORPOSYS: Efficient Colocation of QoS-Constrained Workloads in the Cloud [C]// CCGrid, Ottawa, 2012: 90-97
- [45] Dang D T, Hermenier F. Higher SLA Satisfaction in Datacenters with Continuous VM Placement Constrains [C]// Proceedings of

the 9th Workshop on Hot Topics in Dependable System(Hot-Dep'13). 2013;1-6

[46] Yang T, Lee Y C, Zomaya A Y. Energy-Efficient Data Center Networks Planning with Virtual Machine Problem and Traffic Configuration [C]//CloudCom. Singapore, 2014:284-291

[47] Baran B, von Lucken C, Sotelo A. Multi-objective Pump Scheduling Optimisation Using Evolutionary Strategies [J]. Advances in Engineering Software, 2005, 36(1):39-47

[48] von Lucken C, Baran B, Brizuela C. A Survey on Multi-objective Evolutionary Algorithms for Many-objective Problems [J]. Computational Optimization and Applications, 2014, 58(3):707-756

[49] Lee S, Panigraphy R, Prabhakaran V, et al. Validating Heuristics for Virtual Machines Consolidation; MSR-TR-2011-9[R]. Microsoft Research, 2011

[50] Goudarzi H, Pedram M. Energy-Efficient Virtual Machine Replication and Placement in a Cloud Computing System [C] // CLOUD. Honolulu, 2012:750-757

[51] Li X, Wu J, Tang S J, et al. Let's Stay Together: Towards Traf-

fic Aware Virtual Machine Placement in Data Centers [C]//IN-FOCOM. Toronto, 2014:1842-1850

[52] Pires F L, Baran B. A Virtual Machine Placement Taxonomy [C]//CCGrid. Guangdong, 2015:159-168

[53] Sun M, Wu W D, Zhang X C, et al. A Matrix Transformation Algorithm for Virtual Machine Placement in Cloud [C]// 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). Melbourne, 2013:1778-1783

[54] Su K, Xu L, Chen C, et al. Affinity and Conflict-Aware Placement of Virtual Machines in Heterogeneous Data Centers [C]// ISADS. Taichung, 2015:289-294

[55] Gao Y, Guan H, Qi Z, et al. A Multi-Objective Ant Colony System Algorithm for Virtual Machine Placement in Cloud Computing [J]. Journal of Computer and System Sciences, 2013, 79(8):1230-1242

[56] Wu Y Q, Tang M L, Fraser W. A Simulated Annealing Algorithm for Energy Efficient Virtual Machine Placement [C]// SMC. Seoul, 2012:1245-1250

(上接第 202 页)

中,实线线条表示全站仪实测数据,虚线线条表示采用 PTAM 算法运动估计结果,点虚线线条表示采用本文算法进行运动估计后相机运动轨迹。由 5 组重建轨迹对比分析可知,本文算法本文算法远远优于传统的 PTAM 算法。

图 5、图 7、图 9、图 11、图 13 为 5 组数据集的旋转平移误差对比分析,上图为平均平移误差,下图为平均旋转误差,其中 $R_{\text{平}} = \Delta S / S$, $R_{\text{旋}} = \Delta \alpha / S$, 横坐标表示距离。方形实线线条为 PTAM 算法误差,菱形线条为本文算法误差。

表 1 处理时间(ms)

	构建金字塔	稀疏配准	特征配准	优化	共计
台式电脑	11.81	15.96	20.73	38.16	86.66

具体的各测试数据集的误差对比分析详见表 1。

通过表 1 可以看出,本文系统的运行处理时间,基本可以实现实时在线处理,达到一定的时效性。表 2 为两种不同环境下的误差对比情况,数据表明,产生的误差均保证在允许误差的范围之内,实时有效。

表 2 估计误差对比

场景	本文算法/PTAM 算法			
	RMS 3D(%)	最大误差(m)	平均误差(m)	累积误差(m)
数据集 1	1.15/6.85	0.0285/0.0337	0.0179/0.0207	1.533/1.987
数据集 2	1.08/9.46	0.0611/0.0699	0.0323/0.0441	1.986/2.576
数据集 3	2.38/9.61	0.0311/0.0559	0.0161/0.0324	1.649/3.352
数据集 4	0.97/5.24	0.0332/0.0535	0.0144/0.0287	1.124/1.964
数据集 5	3.52/10.28	0.0318/0.0395	0.0228/0.0338	1.524/2.357

结束语 测试过程中出现了一些问题。相机只发生旋转(相机位置固定)时,系统不能通过仅依靠角度信息计算出相机位姿定位。且在移动平台采集数据时,如遇大型车辆(如车辆)经过,会出现相对静止情况出现,对运动估计结果出现极大影响,因此后续工作中会陆续引入其他传感器进行结合,构建更好的定位系统。还需通过引入集束调整策略对生成的初步点云结果进行优化,同时采用并行技术分别对点云位置和相机位置分别进行优化,以提高系统性能。或者引入目前较流行的深度滤波算法,通过深度滤波位姿估计法以减少在特征匹配环节时间的花销,而且可以得到很好的定位精度和重建模型。

参 考 文 献

[1] Lim H, Lim J, Kim H J. Real-time 6-DOF monocular visual SLAM in a large-scale environment[C]// IEEE International Conference on Robotics and Automation (ICRA). 2014:1532-1539

[2] 孟旭炯. 基于单目视觉的同时定位与地图重建算法研究[D]. 杭州:浙江大学, 2010

[3] 郑驰. 基于光流法的单目视觉里程计研究[D]. 杭州:浙江大学, 2013

[4] 卢维. 高精度实时视觉定位的关键技术研究[D]. 杭州:浙江大学, 2015

[5] 段京翼. 室外机器人单目视觉里程计研究[D]. 南京:南京信息工程大学, 2014

[6] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C]// Proc. IEEE Intl. Conf. on Robotics and Automation. 2014:15-22

[7] Engel J, Ops T S, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]// Computer Vision. 2014:834-849

[8] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry [C] // Robotics and Automation (ICRA). 2014:15-22

[9] Mur-Artal R, Montiel J M M, Tard J D. ORB-SLAM: a Versatile and Accurate Monocular SLAM System[C]// CoRR. 2015

[10] Tan W, Liu H, Dong Z, et al. Robust monocular SLAM in dynamic environments [C] // IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2013:209-218