平行视觉: 基于 ACP 的智能视觉计算方法

王坤峰 [1] 苟 超 [1,2] 王飞跃 [1,3]

摘 要 在视觉计算研究中, 对复杂环境的适应能力通常决定了算法能否实际应用, 已经成为该领域的研究焦点之一. 由人工社会 (Artificial societies)、计算实验 (Computational experiments)、平行执行 (Parallel execution) 构成的 ACP 理论在复杂系统建模与调控中发挥着重要作用. 本文将 ACP 理论引入智能视觉计算领域, 提出平行视觉的基本框架与关键技术. 平行视觉利用人工场景来模拟和表示复杂挑战的实际场景, 通过计算实验进行各种视觉模型的训练与评估, 最后借助平行执行来在线优化视觉系统, 实现对复杂环境的智能感知与理解. 这一虚实互动的视觉计算方法结合了计算机图形学、虚拟现实、机器学习、知识自动化等技术, 是视觉系统走向应用的有效途径和自然选择.

关键词 平行视觉, 复杂环境, ACP 理论, 数据驱动, 虚实互动

引用格式 王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法. 自动化学报, 2016, **42**(10): 1490−1500

DOI 10.16383/j.aas.2016.c160604

Parallel Vision: An ACP-based Approach to Intelligent Vision Computing

WANG Kun-Feng[1] GOU Chao[1,2] WANG Fei-Yue[1,3]

Abstract In vision computing, the adaptability of an algorithm to complex environments often determines whether it is able to work in the real world. This issue has become a focus of recent vision computing research. Currently, the ACP theory that comprises artificial societies, computational experiments, and parallel execution is playing an essential role in modeling and control of complex systems. This paper introduces the ACP theory into the vision computing field, and proposes parallel vision and its basic framework and key techniques. For parallel vision, photo-realistic artificial scenes are used to model and represent complex real scenes, computational experiments are utilized to train and evaluate a variety of visual models, and parallel execution is conducted to optimize the vision system and achieve perception and understanding of complex environments. This virtual/real interactive vision computing approach integrates many technologies including computer graphics, virtual reality, machine learning, and knowledge automation, and is developing towards practically effective vision systems.

Key words Parallel vision, complex environments, ACP theory, data-driven, virtual/real interaction

Citation Wang Kun-Feng, Gou Chao, Wang Fei-Yue. Parallel vision: an ACP-based approach to intelligent vision computing. *Acta Automatica Sinica*, 2016, **42**(10): 1490−1500

何谓平行视觉? 为什么要研究发展平行视觉?

平行视觉是复杂系统建模与调控的 ACP (Artificial societies, computational experiments, and parallel execution) 理论[1−3] 在视觉计算领域的推广应用. 其核心是利用人工场景来模拟和表示复杂挑战的实际场景, 通过计算实验进行各种视觉模型的训练与评估, 最后借助虚实互动的平行执行来在线优化视觉模型, 实现对复杂环境的智能感知与理解. 这一虚实互动的视觉计算方法结合了计算机图形学、虚拟现实、机器学习、知识自动化等技术, 是视觉系统走向应用的有效途径和自然选择.

在智能视觉计算研究中, 一个受到广泛关注的问题是算法在复杂环境下的有效性[4−8], 它直接决定了算法能否实际应用. 以交通环境为例, 雨雪雾等恶劣天气、强阴影、夜间低照度等因素经常导致图像细节模糊, 目标具有各种类型、外观和运动特征, 并且目标之间可能存在遮挡, 又进一步增加了视觉算法的设计难度. 许多视觉算法没有经过充分测试, 尽管在简单的受约束环境下有效, 但是在实际应用时面对复杂的开放环境, 算法很容易失败[4−8].

在深度学习热潮之前, 传统视觉计算方法的基本思路是手动设计图像特征 (例如 Harr 小波、SIFT

收稿日期 2016-08-24 录用日期 2016-09-26
Manuscript received August 24, 2016; accepted September 26, 2016
国家自然科学基金 (61533019, 61304200), 国家留学基金资助
Supported by National Natural Science Foundation of China (61533019, 61304200) and China Scholarship Council
本文责任编委 刘德荣
Recommended by Associate Editor LIU De-Rong
1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 2. 青岛智能产业技术研究院 青岛 266000 3. 国防科学技术大学军事计算实验与平行系统技术研究中心 长沙 410073
1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 2. Qingdao Academy of Intelligent Industries, Qingdao 266000 3. Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073

(Scale invariant feature transform)、HOG (Histogram of oriented gradient)、LBP (Local binary pattern) 等), 然后利用标记数据集训练模式分类器 (例如 SVM (Support vector machine)、Adaboost、随机森林等), 取得了较好的实验效果 (例如 DPM (Deformable parts model) 目标检测器[9]). 然而由于模型限制, 这类方法通常依赖于小规模的标记数据集 (例如 INRIA Person[10]、Caltech Pedestrian[11]、KITTI[12] 等数据集), 样本数大致在几千到几十万之间, 难以覆盖复杂环境对应的特征空间. 近年流行的深度学习方法[13−16] 具有强大的特征表达能力, 能够利用标记数据集通过端到端训练 (End-to-end training) 得到分层特征描述, 在图像分类、目标检测等竞赛中显著优于传统方法, 并且性能仍在持续提升. 深度学习依赖于大规模标记数据集 (例如 ImageNet[17]、PASCAL VOC[18]、MS COCO[19] 等), 样本数通常在百万级以上, 能够覆盖更大的特征空间.

由于实际环境的复杂性, 为了建立有效的视觉模型, 不但要求标记数据集规模足够大, 还要求具有足够的多样性 (Diversity). ImageNet 等数据集尽管规模庞大, 却并不满足多样性要求, 不能覆盖复杂挑战的实际环境. 这一状况来自两方面原因. 1) 在复杂环境下采集大规模多样性数据集需要耗费大量人力, 目前 ImageNet[17] 主要从 Internet 上搜集图像, 但是网络空间与物理空间并不等价[20]. 2) 对大规模多样性数据集进行标注需要耗费大量人力并且容易出错, 尤其在恶劣天气、夜间低照度等环境下, 由于图像细节模糊, 由人眼观察标注图像中的目标位置、姿态、运动轨迹都很困难. 标记数据集的不足, 降低了视觉模型的泛化能力, 无法保证实际应用时的有效性.

为了解决大规模多样性数据集的采集和标注困难, 一种可选方案是建立人工场景, 模拟和替代复杂挑战的实际场景, 生成人工场景数据集. 近年来随着游戏引擎[21−22]、虚拟现实[23−25] 等技术的发展, 使构建色彩逼真的人工场景成为可能. 利用人工场景, 可以模拟实际场景中的各种要素, 包括光照时段 (白天、夜间、黎明、黄昏)、天气 (晴、多云、雨、雪、雾等)、目标类型 (行人、车辆、道路、建筑物、植物等) 和子类等. 并且可以灵活地设计各种场景类型、目标外观、目标行为、摄像机配置等. 由此可以生成大规模多样性的视频图像数据集, 并且可以自动得到精确的标注信息, 包括目标位置、运动轨迹、语义分割、深度、光流等.

平行视觉建立在实际场景与人工场景之上, 是一种虚实互动的智能视觉计算方法. 它借鉴了复杂系统建模与调控的 ACP 理论[1−3], 即人工社会 (Artificial societies)、计算实验 (Computational experiments) 和平行执行 (Parallel execution). 通过构建色彩逼真的人工场景, 模拟实际场景中可能出现的环境条件, 并且自动得到精确的标注信息. 结合大规模的人工场景数据集和适当规模的实际场景数据集, 能够训练出更有效的机器学习和视觉计算模型. 利用人工场景, 能够进行各种计算实验, 全面评价视觉算法在复杂环境下的有效性, 或者优化设置模型的自由参数. 如果将视觉模型在实际场景与人工场景中平行执行, 使模型训练和评估在线化、长期化, 则能够持续优化视觉系统, 提高其在复杂环境下的运行效果.

本文其他部分内容安排如下: 第 1 节对相关工作进行综述; 第 2 节提出平行视觉的基本框架; 第 3 节介绍平行视觉的核心算法和关键技术; 第 4 节对本文进行总结, 并对平行视觉的发展趋势进行展望.

1 相关工作综述

正如 Bainbridge 在 *Science* 上发表的论文[21] 所述, 虚拟世界以视频游戏和计算机游戏的形式, 在视觉上模拟复杂的物理空间, 为科学研究提供一个新的环境. 构建虚拟世界或人工场景的相关技术正在快速发展, 在科学研究、人类生活等方面发挥着重要作用.

科幻电影 “阿凡达 (Avatar)” 以令人震撼的视觉效果, 构建了潘多拉星球这一虚拟世界, 呈现了参天巨树、群山、怪兽、Na′vi 族人等虚拟对象, 给观众留下了深刻印象. Miao 等[22] 提出一种基于游戏引擎的平台, 进行人工交通系统的建模和计算. 作者将人工人口设计为游戏中的角色, 利用 Delta3D 游戏引擎构建 3D 仿真环境, 利用 Delta3D 的动态角色层机制管理所有移动的角色 (包括车辆、行人等), 设计了一种面向 Agent 的模块化分布式仿真平台. Sewall 等[23] 提出虚拟化交通 (Virtualized traffic) 概念, 基于离散时空数据来重建和可视化连续交通流, 使用户能够在虚拟世界中观看虚拟化交通事件. 给定路段上每个车辆的两个位置点和对应的行驶时间, 该方法能够重建交通流, 实现虚拟城市的沉浸式可视化. 该方法可应用于高密度交通, 包括任意的车道数, 同时考虑了车辆的几何、运动和动态约束. Prendinger 等[24] 利用 OpenStreetMap、CityEngine、Unity3D 等软件构建虚拟生活实验室 (Virtual Living Lab), 用于交通仿真和用户驾驶行为研究. 作者基于免费地图数据生成车辆出行路网, 并通过车辆 Agent 与路段 Agent 的交互实现环境感知. Karamouzas 等[25] 提出一种新的行人小群体运动模型, 描述群体成员如何与其他成员、其他群体和个体交互, 并且通过构建人工场景来

验证所提模型的有效性. 这些工作虽然不是直接针对视觉计算研究, 但是对人工场景构建很有启发意义.

构建的人工场景可用于摄像机网络控制方法研究. Qureshi 等[26] 利用 OpenGL 构建虚拟火车站和虚拟行人, 并在场景中设置虚拟摄像机, 组成摄像机网络, 如图 1 所示. 该工作建立的人工场景规模较小 (最多仿真 16 台虚拟摄像机、100 个行人), 并且逼真度较低, 没有仿真阴影、复杂光照、反射高光等成像细节. 作者从人工场景视频中提取目标检测和跟踪信息, 在此基础上研究 PTZ 摄像机控制算法, 包括摄像机指派、交接等. Starzyk 等[27] 基于 Panda3D 游戏引擎, 设计了一套分布式虚拟视觉仿真器, 建立了支持摄像机网络研究的软件实验室. 他们仿真办公室场景, 生成人工场景视频, 进行行人检测、跟踪等视觉处理. 根据视觉分析结果进行摄像机操作, 例如摄像机控制、协调、交接等. 该系统在多台计算机上联网实现, 具有较强的可扩展性, 能够仿真大尺度摄像机网络. 作者设计了由 100 多台虚拟摄像机组成的视觉网络.

一些工作基于人工场景数据集进行视觉模型训练. Sun 等[28] 利用 Google 3D Warehouse 获得目标的 3D 模型, 并通过 3D 模型旋转生成 2D 图像数据, 得到虚拟图像集. 在此基础上利用判别去相关 (Discriminative decorrelation) 方法训练 2D 目标检测器, 在缺少实际场景标记图像的情况下进行领域适应. 实验发现, 与基于实际图像集训练出的目标检测器相比, 他们基于虚拟图像集的方法能够获得类似的精度. Hattori 等[29] 在缺少实际场景训练图像的情况下, 完全依靠虚拟数据, 训练面向特定场景 (Scene-specific) 的行人检测器. 已知场景几何信息和摄像机标定参数, 他们利用 Autodesk 3DS Max 软件建立人工场景, 生成虚拟行人数据, 作为训练集. 他们的行人检测器在精度上超过了以 DPM 为代表的通用检测器 (Generic detector), 并且超过了基于少量实际行人数据训练出来的面向特定场景的检测器.

此外, 还有更多的工作结合人工场景数据集和实际场景数据集进行视觉模型训练. 例如, Vázquez 等[30] 利用 Half-Life 2 游戏引擎生成逼真的虚拟世界图像, 训练行人检测器. 他们发现, 基于虚拟世界的训练能够在真实世界中产生很高的测试精度, 但是存在数据集偏移 (Dataset shift) 问题. 于是他们设计了一种领域适应框架 V-AYLA, 先基于虚拟世界数据集训练行人检测器, 然后利用真实世界图像进行主动学习, 发掘困难的正例和反例, 迭代调节检测器参数. 与基于大量真实世界标记样本训练的检测器相比, 虽然 V-AYLA 只利用了少量的真实世界标记样本, 却能够获得相同的性能. 该研究组进一步提出利用虚拟世界数据集训练基于 DPM 的行人检测器[31].

Gaidon 等[32] 利用 Unity 游戏引擎克隆 KITTI

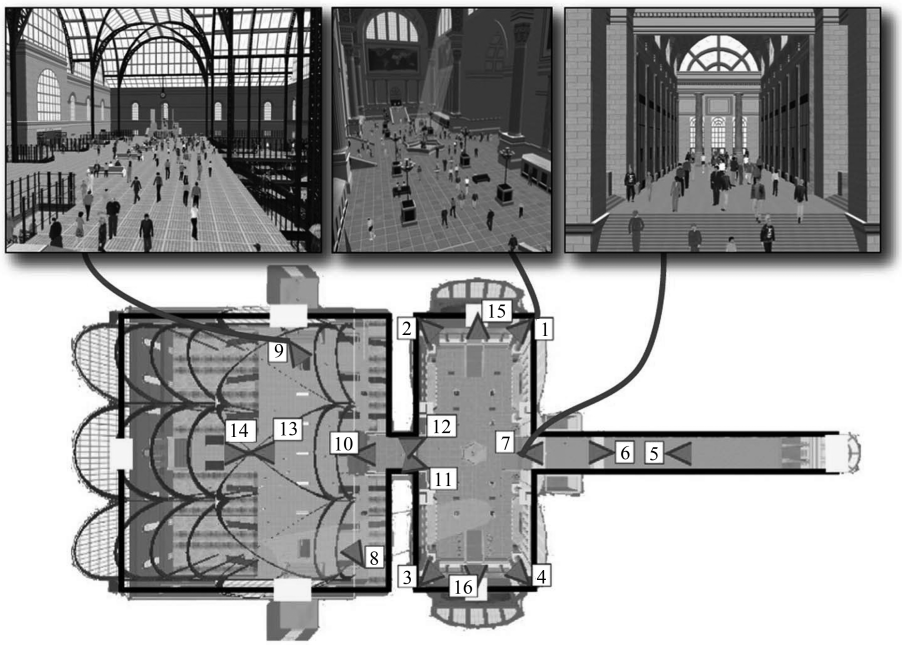

图 1 虚拟火车站的平面图[26] (包括站台和火车轨道 (左)、主候车室 (中) 和购物商场 (右). 该摄像机网络包括 16 台虚拟摄像机)

Fig. 1 Plan view of the virtual train station[26] (Revealing the concourses and train tracks (left), the main waiting room (middle), and the shopping arcade (right). An example camera network comprising 16 virtual cameras is illustrated.)

数据集[12], 生成 "虚拟 KITTI" 数据集, 并自动生成目标检测、跟踪、语义分割、深度和光流的标注信息, 如图 2 所示. 另外, 对每段克隆的虚拟视频, 模拟环境条件 (包括摄像机朝向、光照和天气条件等) 变化, 得到更加多样化的虚拟数据. 他们实验发现: 基于真实数据训练的深度学习算法, 当应用于真实世界和虚拟世界时表现相似; 首先利用虚拟 KITTI 数据做模型预训练, 然后利用真实 KITTI 数据做模型参数微调, 能够提高性能. 他们还将基于真实数据训练的目标跟踪器应用于环境条件变化的虚拟视频, 发现光照和天气条件显著降低跟踪性能, 恶劣天气 (例如雾天) 导致性能的最大下降. 对此进一步感兴趣的读者, 可以参考项目网址 http://www.xrce.xerox.com/Research-Development/Computer-Vision/Proxy-Virtual-Worlds.

Handa 等[33] 利用 CAD 模型仓库, 建立人工室内场景数据集 SceneNet, 包括床、书、天花板、桌子、椅子、地板、沙发等虚拟对象, 自动生成像素级语义标注. 他们研究基于深度 (Depth) 的图像语义标注, 先利用人工场景数据集训练卷积神经网络, 然后利用实际场景数据集对网络参数进行微调. 实验发现, 虽然只是将深度作为输入, 由于人工场景数据集的辅助, 训练出的 CNN (Convolutional neural network) 模型达到了接近甚至优于 State-of-the-art 的性能. 与此同时, Ros 等[34] 利用 Unity 游戏引擎, 建立虚拟城市图像集 SYNTHIA, 包括街区、高速路、郊区、商店、公园、植物、各种路面、车道标线、交通标志、灯柱、行人、车辆等元素, 并且自动生成像素级语义标注, 如图 3 所示. SYNTHIA 中的图像具有较高的逼真度, 可以模拟季节变化 (例如冬季地面有雪、春季植物开花等)、动态光照、投射阴影、恶劣天气等自然现象. 由于手动标注图像语义需要耗费大量人力并且容易出错, 该工作能够显著增大训练数据集的规模和多样性. 他们利用虚拟城市图像集和真实城市图像集共同训练深度卷积神经网络, 实验结果表明这项工作显著提高了图像语义分割的精度. 对此进一步感兴趣的读者, 可以参考项目网址 http://synthia-dataset.net/.

Movshovitz-Attias 等[35] 利用 3DS MAX 软件和 91 种精细的 3D CAD 车辆模型, 生成虚拟车辆图像集 RenderCar, 并且自动得到精确的视角 (Viewpoint) 标注. 在图像渲染时考虑了光源的位置、强度和颜色、摄像机的光圈大小、快门速度和镜头渐晕效应、复杂背景、图像噪声、随机遮挡等因素, 使生成的虚拟图像非常逼真, 同时增加了图像的多样性, 如图 4 所示. 作者利用 RenderCar、PASCAL3D+、CMU-Car 三个图像集, 训练深度卷积神经网络, 进行目标视角估计. 实验发现, 基于虚拟图像集训练出的模型与基于真实图像集训练出的模型性能相近, 都存在数据集偏移问题; 如果结合虚拟和真实图像集, 训练出的模型具有更高的精度.

图 2　虚拟 KITTI 数据集[32] (上: KITTI 多目标跟踪数据集中的一帧图像; 中: 虚拟 KITTI 数据集中对应的图像帧, 叠加了被跟踪目标的标注边框; 下: 自动标注的光流 (左)、语义分割 (中) 和深度 (右))

Fig. 2　The virtual KITTI dataset[32]. (Top: a frame of a video from the KITTI multi-object tracking benchmark. Middle: the corresponding synthetic frame from the virtual KITTI dataset with automatic tracking ground truth bounding boxes. Bottom: automatically generated ground truth for optical flow (left), semantic segmentation (middle), and depth (right).)

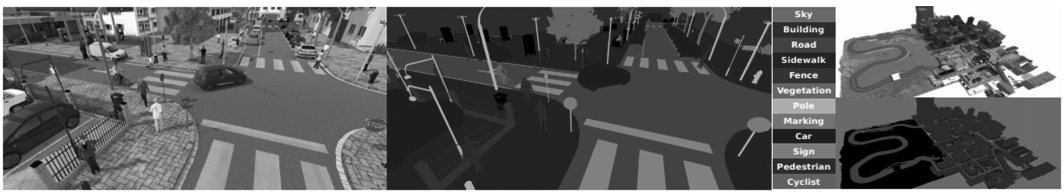

图 3　SYNTHIA 数据集[34] (左: 人工场景中的一帧图像; 中: 对应的语义标记; 右: 虚拟城市的全貌)

Fig. 3　The SYNTHIA dataset[34] (A sample frame (left) with its semantic labels (middle) and a general view of the virtual city (right).)

图 4 RenderCar 中的样本图像[35]

Fig. 4 Sample images from RenderCar[35]

还有许多工作利用人工场景数据集进行算法测评, 例如利用 SABS[36] 或 BMC[37] 数据集验证背景消减算法、利用 CROSS 数据集[38] 验证行为分析算法、利用 MPI-Sintel 数据集[39] 评价光流算法、利用虚拟城市和自由女神雕像数据集[40] 评价图像特征、利用 OVVV 数据集[41] 评价跟踪和监控算法等. Zitnick 等[42] 利用剪贴画组合技术创建了 1 002 个语义场景, 每个场景包含 10 个语义相似的抽象图像, 来研究视觉数据的高层语义理解. 该方法能够创建大量语义相似的场景, 并且避免了目标检测错误, 便于直接进行高层语义研究. 作者通过数据集分析, 研究了视觉特征的语义重要性、目标的显著性与可记忆性, 以及这些概念之间的关系. Veeravasarapu 等[43−44] 利用 Blender 渲染软件构建人工交通场景, 来验证视觉系统在复杂环境 (光照变化、恶劣天气、高频噪声等) 下的性能. 作者从亮度不变性、梯度不变性、二色大气散射等角度证明人工场景视频能够用于视觉模型训练和评估, 并且以背景消减、行人检测为例验证了几种视觉算法.

综上所述, 近年来平行视觉的相关研究呈现出两个趋势. 1) 开源和商业 3D 仿真工具越来越丰富, 功能也越来越强大, 使构建的人工场景越来越逼真. 通过对比图 1 (2008 年成果) 和图 2 ∼ 图 4 (2016 年成果), 可以清晰地感受到这一趋势. 2) 对人工场景的构建和利用已经触及视觉计算研究的方方面面, 从低层的光流估计、目标检测、语义分割等, 到中层的目标跟踪, 再到高层的行为分析、语义理解等, 虚拟现实和人工场景技术都开始发挥作用. 2016 年 10 月召开的欧洲计算机视觉会议 (ECCV) 将举行第 1 届 Virtual/Augmented Reality for Visual Artificial Intelligence 研讨会, 表明该方向已经引起国际同行的重视. 但是目前来看, 基于人工场景的视觉计算研究工作较为分散, 缺少统一的理论支持. 因此, 本文提出平行视觉的基本框架和关键技术, 希望能够为视觉计算研究人员带来一些启发, 促进该领域更好更快地发展.

2 平行视觉的基本框架

王飞跃于 2004 年提出了复杂系统建模与调控的 ACP 理论[1−3], 即:

$$\text{ACP} = \text{Artificial societies} + \text{Computational experiments} + \text{Parallel execution}$$

ACP 理论通过这一组合, 将人工的虚拟空间 Cyberspace 变成解决复杂问题的新的另一半空间, 同自然的物理空间一起构成求解复杂问题之完整的 “复杂空间”. 新兴的物联网、云计算、大数据等技术, 是支撑 ACP 理论的核心技术. 从本质上讲, ACP 的核心就是把复杂系统 “虚” 的和 “软” 的部分建立起来, 通过可定量、可实施的计算实验, 使之 “硬化”, 真正地用于解决实际的复杂问题. 在 ACP 理论的基础上, 形成了实际系统与人工系统并行互动的平行系统. 目前, ACP 理论和平行系统思想已经在城市交通控制、乙烯生产管理、社会计算等领域获得示范应用[2−3], 其中平行交通被国家发改委列入 “互联网 +” 便捷交通重点示范项目[45]. 基于 ACP 的平行方法在计算机视觉方面, 也进行了一些初步的探讨[46].

本文提出的平行视觉是 ACP 理论在视觉计算领域的推广应用, 目标是解决复杂环境 “视觉计算方案” 的科学难题. 图 5 显示了平行视觉的基本框架和体系结构. 总体上, 平行视觉之 ACP 由 “三步曲” 组成.

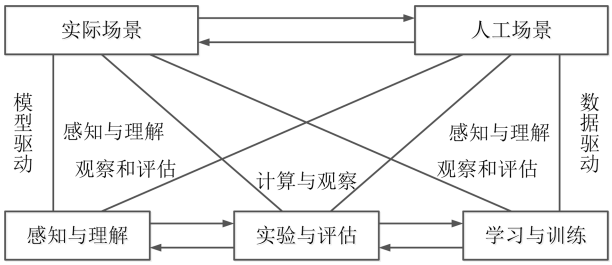

图 5 平行视觉的基本框架与体系结构

Fig. 5 Basic framework and architecture for parallel vision

第一步 (A 步). 构建色彩逼真的人工场景, 模拟实际场景中可能出现的环境条件, 自动得到精确的标注信息, 生成大规模多样性数据集. 一定意义下, 可以把人工场景看作 “视频游戏”, 就是用类似于计算机游戏的技术来建模. 这里主要运用了计算机图形学、虚拟现实、微观仿真等技术. 大体上, 可以把计算机图形学和计算机视觉看作一对正反问题.

计算机图形学是给定 3D 世界模型及其参数, 按照实际摄像机图像生成的原理和过程, 合成出人工场景图像. 而计算机视觉是给定图像序列, 反求 3D 世界模型、参数和语义信息. 平行视觉正是利用了计算机图形学和计算机视觉之间的这种正反关系.

在许多情况下, 由于数据采集和标注困难, 从实际场景中无法获得令人满意的数据集, 影响视觉算法的设计与评估. 利用人工场景数据集, 可以解决这些问题. 首先, 借助计算机平台, 人工场景可以提供 "无限" 规模的数据, 通过在图像生成过程中设定各种物理模型和参数, 可以得到 "无限" 多样的数据, 并且自动生成标注信息, 从而满足对标注数据集的 "大规模" 和 "多样性" 要求. 其次, 实际场景通常不可重复, 而人工场景具有 "可重复性", 通过固定一些物理模型和参数, 改变另外一些, 可以 "定制" 图像生成要素, 以便从各种角度评价视觉算法. 然后, 某些实际场景由于特殊性, 无法从中获得实际数据集, 人工场景可以避免这一问题. 例如为战场环境设计视觉监控系统, 可能无法事先得到敌方活动的视频图像, 可以在计算机上建立人工场景数据集, 对视觉算法进行设计和评估. 又例如为火星无人车设计视觉导航系统, 我们现在无法获得火星地面的大规模实际图像集, 可以通过构建人工场景来辅助设计视觉算法. 总之, 构建人工场景意义重大, 能够为视觉算法设计与评估提供一种可靠的数据来源, 是对实际场景数据的有效补充.

第二步 (C 步). 结合人工场景数据集和实际场景数据集, 进行各种计算实验, 设计和优化视觉算法, 评价视觉算法在复杂环境下的性能. 这里主要运用了机器学习、领域适应、统计分析等技术. 已有的多数视觉系统, 由于应用环境太复杂, 没有经过全面实验, 只是在有限环境下做算法设计和评估, 然后不管三七二十一实施了再说, 对实施效果却是 "心中无数". 若要视觉系统真正有效, 必须在人工场景中进行全面充分的实验. 就是把计算机变成视觉计算 "实验室", 利用人工场景做 "计算实验", 全面设计和评估视觉算法. 与基于实际场景的实验相比, 在人工场景中实验过程可控、可观、可重复, 并且可以真正地产生 "大数据", 用于后续的知识提取和算法优化.

计算实验有两种操作模式, 即学习与训练、实验与评估. "学习与训练" 是针对视觉算法设计而言, 机器学习是智能视觉计算的核心, 无论传统的统计学习方法 (SVM、Adaboost、随机森林等), 还是目前流行的深度学习, 主要依靠 "Learning from data", 训练数据集起着至关重要的作用. 结合大规模人工场景数据集和适当规模的实际场景数据集, 有监督训练机器学习模型, 能够提高视觉算法的性能. 尤其对于深度学习技术, 训练数据增多, 性能会更好[47−49]. 由于机器学习过程中普遍存在数据集偏移问题, 即源领域数据和目标领域数据具有不同的统计分布, 因此必须进行领域适应. 可以首先利用人工场景数据集预训练模型, 然后利用目标领域的实际场景数据集微调模型参数; 也可以为人工场景数据和实际场景数据设定比例, 同时利用它们训练模型. "实验与评估" 是针对视觉算法评价而言, 也就是利用人工场景数据集 (以及实际场景数据集) 评价算法的性能. 由于可以完全控制人工场景的环境条件 (例如光照、天气、目标外观和运动等), 对视觉算法的测试会更充分, 结合统计分析技术, 能够在系统实施之前定量评价视觉算法在各种环境条件下的表现, 做到 "心中有数". 总之, 将计算实验从实际场景扩展到人工场景, 不但拓宽了实验的广度, 更增加了实验的深度, 有助于提高视觉算法性能.

第三步 (P 步). 将视觉模型在实际场景与人工场景中平行执行, 使模型训练和评估在线化、长期化, 通过实际与人工之间的虚实互动和人机混合, 持续优化视觉系统. 这里主要运用了在线学习、知识自动化等技术. 从相关工作综述可知, 许多学者都有类似于 ACP 的想法, 主要集中在前两步, 但是要解决复杂环境的视觉计算问题, "三步曲" 缺一不可. 由于应用环境的复杂性、挑战性和变化性, 不存在一劳永逸的解决方案. 只能接受这些困难, 在运行过程中不断调节和改善, 即将虚实互动和人机混合常态化, 以平行执行的方式持续优化视觉系统, 在复杂环境下进行有效的感知与理解.

平行执行的最大特色是 "把人工场景构建在环内" (The artificial scenes are constructed in the loop), 依靠数据来驱动. 除物理空间的实时视频数据外, 还包括实时光照和天气条件, 以及来自 Web 和 Cyberspace 丰富的虚拟对象模型等数据. 在海量数据的基础上, 自动生成各种有实际意义的人工场景. 在物联网和云计算技术的支持下, 与实际场景对应的人工场景可以有多个, 不是为了 "复制" 或 "重建" 实际场景, 而是为了 "预测"、"培育" 实际场景的可能存在, 为视觉计算增加主动性. 通过实际与人工的虚实互动, 在线训练和评估视觉模型, 不断改善视觉系统, 一方面提高在当前场景中的运行效果, 另一方面为应对未来场景做好准备. 总之, 平行执行是一种基于大数据, 以在线仿真和优化为主要手段的感知与理解复杂环境的方法, 它可以实现视觉计算的知识自动化, 迈向智能视觉计算.

至此, 我们可以进一步明确**平行视觉的基本原则:** 在物理和网络空间大数据的驱动下, 结合计算机图形学、虚拟现实、机器学习、知识自动化等技术, 利用人工场景、计算实验、平行执行等理论和方法,

建立复杂环境下视觉感知与理解的理论和方法体系.

3 平行视觉的核心算法和关键技术

本节分别针对平行视觉之 ACP 三步曲, 提出若干核心算法和关键技术, 希望为本领域研究人员带来一些启发.

3.1 人工场景的核心算法和关键技术

我们以室外场景为例, 对人工场景构建进行说明. 首先应当指出, 构建人工场景不需要从头做起, 而是借助已有的开源或商业游戏引擎和仿真工具, 例如 Unity、Half-Life 2、Delta3D、OpenGL、Panda3D、Google 3D Warehouse、3DS MAX、OVVV、VDrift 等. 每种工具都有其特点, 可以根据具体应用需要进行选择.

人工场景由许多要素构成, 包括静态物体、动态物体、季节、天气、光源等. 用 Agent 表示场景要素, 按照物理规律进行多 Agent 仿真. 人工室外场景的构成要素如表 1 所示. 可以利用 Web 空间 (例如 Google 3D Warehouse) 海量且丰富的静态和动态物体的 3D 模型. 动态物体应具有路径生成和障碍物规避功能. 季节和天气直接影响人工场景的渲染效果, 要求与物理世界的自然规律一致, 例如春季植物开花、冬季地面有雪、晴天投射阴影、雾天物体模糊等. 白天光源主要是太阳, 夜间光源主要是路灯和车灯. 从白天向夜间过渡时, 会自动开启路灯和车灯; 从夜间向白天过渡时, 会自动关闭路灯和车灯. 总之, 要求人工场景的构成要素尽可能逼真并且多样化. 图 6 显示了同一种车型 (货车) 的 3D 模型样例.

在人工场景中设置虚拟摄像机, 生成人工场景图像序列. 虚拟摄像机可以是枪式、云台式或全景式. 摄像机可以是固定的, 例如模拟视频监控; 也可以是移动的, 例如模拟自动驾驶或航拍监控. 相应地, 摄像机位置可以在路口、路段或车载 (机载). 图像生成过程是复杂的: 光从光源发出, 经过大气散射, 到达物体表面; 然后, 被物体漫反射或镜面反射, 再次经过大气散射, 到达摄像机镜头; 最后, 经过光电转换, 生成数字图像. 每个环节都对最终图像有所影响, 例如光源影响光强和色温、天气条件影响大气散射、物体表面影响光的反射、摄像机影响镜头扭曲和图像噪声等. 要想生成色彩逼真的人工场景图像, 必须模拟所有这些过程.

表 1 人工室外场景的构成要素

Table 1 Components for artificial outdoor scenes

场景要素	内容
静态物体	建筑物、天空、道路、人行道、围墙、植物、立柱、交通标志、路面标线等
动态物体	汽车 (轿车、货车、公交车)、自行车、摩托车、行人等
季节	春、夏、秋、冬
天气	晴、阴、雨、雪、雾、霾等
光源	太阳、路灯、车灯等

基于实际场景图像, 难以获得复杂环境下的目标姿态、运动轨迹、语义分割、深度、光流等标注信息. 而人工场景图像是从 3D 模型出发, 自底向上生成的, 因此无论光照和天气条件多么恶劣, 图像细节多么模糊, 都很容易自动得到详细且精确的标注信息. 根据应用需要, 标注应该各有不同. 但总体上, 可以标注的信息包括目标边框、目标区域、目标类型、目标姿态、运动轨迹、图像语义分割、深度、光流等. 基于上述方法和技术, 能够生成色彩逼真的大规模多样性人工场景数据集.

3.2 计算实验的核心算法和关键技术

利用人工场景数据集, 进行各种计算实验, 把计算机变成视觉计算 "实验室". 我们首先为计算实验的两种操作模式 (学习与训练、实验与评估) 分别提出一个例子, 然后简要说明更多的实验思路.

作为第一个例子, 复杂交通环境下的目标检测是一项困难的视觉任务. 在实际应用时, 光照和天气条件、目标和背景外观都很复杂. 在白天和夜间, 光源不同, 光照条件差别很大. 在恶劣天气、夜间低照度、白天强阴影区域等条件下, 目标与背景模糊不清. 相对于摄像机, 目标姿态多样, 并且可能被部分

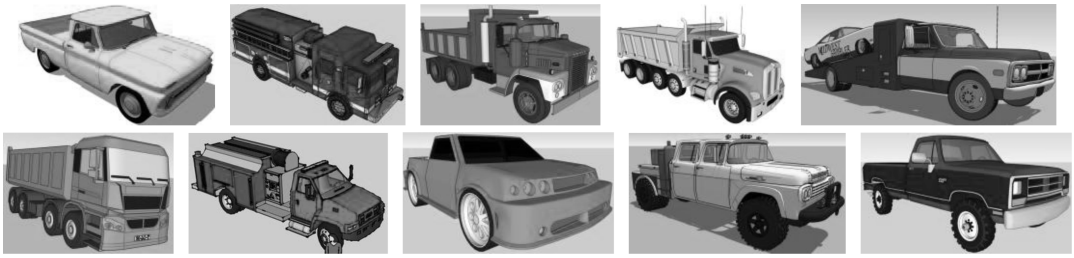

图 6 货车的 3D 模型样例

Fig. 6 Sample 3D models of trucks

遮挡, 为检测增加了新的难度. 在这些因素的综合影响下, 很难设计一个鲁棒的目标检测器. Faster R-CNN[15−16] 是目前精度最高且实时性较好的目标检测器之一, 它由区域提议网和深度残差网组成, 二者共用卷积特征, 如图 7 所示. 在文献 [15 − 16] 中, Faster R-CNN 利用 ImageNet、PASCAL VOC 和 MS COCO 数据集进行学习与训练. 但是这些数据集是从 Internet 上搜集得到, 图像清晰度较高, 缺少恶劣天气和夜间低照度条件的图像, 因此训练的模型在实际应用时很可能失败. 而人工场景能够模拟复杂挑战的交通环境, 提供色彩逼真的大规模多样性数据集, 作为实际场景数据集的补充. 结合人工场景数据集和实际场景数据集, 共同训练 Faster R-CNN 模型, 在每一批训练数据中为人工场景数据和实际场景数据设定比例 (例如 1 : 1), 在训练时能够降低数据集偏移和实现领域适应, 生成更加鲁棒的目标检测器.

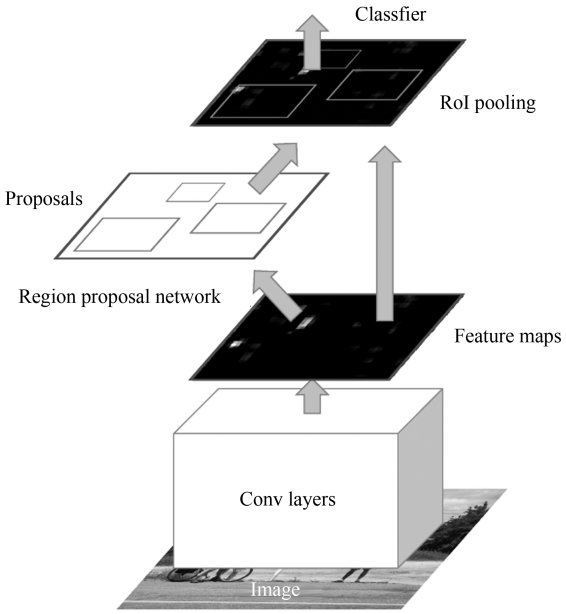

图 7　Faster R-CNN 的结构图[15]

Fig. 7　Flowchart of Faster R-CNN[15]

作为另一个例子, 智能车视觉系统测评也是一项困难任务. 从 2009 年开始, 在国家自然科学基金委的资助下, 每年举办一次 “中国智能车未来挑战赛”[50]. 通过在城市和乡村道路上测试智能车视觉系统的车道识别、障碍物规避、信号灯识别、交通标志识别等功能, 促进了中国智能车领域的发展. 但是, 这种实际场景测试只能覆盖很小一部分环境条件, 是不完备的测试, 无法保证视觉系统在实际应用时的有效性. 如果建立模拟实际场景的人工场景, “定制” 各种场景要素 (天气、光照、路况、交通标志等), 则能够建立更完备的测试数据集, 在计算机上测试智能车视觉算法的性能. 人工场景测试覆盖的环境范围更广, 并且成本更低, 可以作为实际场景测试的补充. 目前, 国家自然科学基金委已经设立相关项目, 并取得初步结果[51].

总体上, 我们可以面向具体应用, 利用人工场景做可控、可观、可重复的计算实验, 全面设计和评估视觉算法. 计算实验之所以重要, 是因为在复杂挑战的实际场景, 难以获得目标姿态、运动轨迹、语义分割、深度、光流等标注信息. 但是人工场景能够模拟复杂环境, 并且自动得到精确的标注信息, 使得以前不易进行甚至无法进行的实验通过计算实验得以顺利进行. 在 “学习与训练” 操作模式下, 结合大规模人工场景数据集和适当规模的实际场景数据集, 有监督训练机器学习模型, 优化参数学习和选择. 无论传统的统计学习模型, 还是目前流行的深度学习模型, 都可以利用人工场景数据集获得更好的泛化性, 更加胜任复杂环境下的视觉计算任务. 在 “实验与评估” 操作模式下, 利用人工场景数据集 (以及一定的实际场景数据集), 全面评价视觉算法在复杂环境下的性能. 控制人工场景的生成要素, 比较算法在各种环境下的性能, 生成 “算法 − 环境” 性能矩阵, 严格量化算法性能, 可以为算法改进提供客观依据.

3.3　平行执行的核心算法和关键技术

将视觉模型在实际场景与人工场景中平行执行, 使模型训练和评估在线化、长期化, 是平行视觉的最高阶段. 在复杂环境下, 视觉感知与理解是极其困难的, 不存在一劳永逸的解决方案, 只能在运行过程中不断调节和改善, 以平行执行的方式持续优化. 当系统运行时, 在物理和网络空间大数据的驱动下, 能够把人工场景构建在环内. 从实时图像中 (自动或者半自动) 获取场景关键要素, 包括静态物体、动态物体、天气、光照等, 结合 Web 和 Cyberspace 海量且丰富的虚拟对象模型, 在线 “培育” 各种有实际意义的人工场景. “有实际意义” 不是指人工场景必须在外观上 “复制” 或 “重建” 当前的实际场景, 而是指人工场景必须与实际场景有相通之处, 必须对模型训练和评估有借鉴意义. 在物联网和云计算技术的支持下, 虽然实际场景是唯一的, 但是与某个实际场景对应的人工场景可以有多个. 当然, 也可以多个实际场景共享多个人工场景. 因此, 实际与人工是一对多、多对多的关系.

在线构建的人工场景提供了 “无限” 的在线数据, 可以用来在线训练和评估视觉模型. 在线数据蕴含了实际场景的动态变化信息, 例如场景光照、天气等条件在不断变化. 在运行过程中, 视觉模型不应该一成不变, 必须通过计算实验, 随着场景变化逐渐调节和改善. 在 “学习与训练” 操作模式下, 如果是深

度学习模型, 可以在线累积人工场景数据, 同时随机选择离线的实际场景数据, 按照一定比例组成每一批训练数据, 有监督微调神经网络参数, 使模型自动适应实际场景的最新变化. 在 "实验与评估" 操作模式下, 利用在线的人工场景数据和实际场景数据, 定期评价模型性能. 如果模型性能下降较多, 则需要增加更多的训练数据以调节模型, 甚至替换成性能表现更好的模型. 总之, 平行执行将虚实互动常态化, 通过对人工场景的在线构建和利用, 持续优化视觉系统, 实现视觉计算的知识自动化.

4 结论与展望

本文将 ACP 理论推广到视觉计算领域, 提出平行视觉的基本框架和关键技术. 平行视觉在物理和网络空间大数据的驱动下, 结合计算机图形学、虚拟现实、机器学习、知识自动化等技术, 利用人工场景、计算实验、平行执行等理论和方法, 建立复杂环境下视觉感知与理解的理论和方法体系. 平行视觉利用人工场景来模拟和表示复杂挑战的实际场景, 使采集和标注大规模多样性数据集成为可能, 通过计算实验进行视觉算法的设计与评估, 最后借助平行执行来在线优化视觉系统.

平行视觉相关研究已经引起国际同行的高度重视. 在近几年召开的计算机视觉重要会议 (例如 CVPR、ECCV 等) 上, 将计算机图形学和虚拟现实技术用于解决复杂环境下的视觉计算问题, 在论文数量和关注程度上呈现出上升趋势. 随着虚拟现实技术的进一步发展, 构建的人工场景会更加逼真, 为平行视觉研究提供更可靠的基础支撑. 我们相信, 平行视觉将成为视觉计算领域一个重要的研究方向. 尤其是, 平行视觉与深度学习相结合, 将推动越来越多的智能视觉系统发展成熟并走向应用.

References

1 Wang Fei-Yue. Parallel system methods for management and control of complex systems. *Control and Decision*, 2004, **19**(5): 485−489, 514
(王飞跃. 平行系统方法与复杂系统的管理和控制. 控制与决策, 2004, **19**(5): 485−489, 514)

2 Wang F Y. Parallel control and management for intelligent transportation systems: concepts, architectures, and applications. *IEEE Transactions on Intelligent Transportation Systems*, 2010, **11**(3): 630−638

3 Wang Fei-Yue. Parallel control: a method for data-driven and computational control. *Acta Automatica Sinica*, 2013, **39**(4): 293−302
(王飞跃. 平行控制: 数据驱动的计算控制方法. 自动化学报, 2013, **39**(4): 293−302)

4 Wang K F, Liu Y Q, Gou C, Wang F Y. A multi-view learning approach to foreground detection for traffic surveillance applications. *IEEE Transactions on Vehicular Technology*, 2016, **65**(6): 4144−4158

5 Wang K F, Yao Y J. Video-based vehicle detection approach with data-driven adaptive neuro-fuzzy networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 2015, **29**(7): 1555015

6 Gou C, Wang K F, Yao Y J, Li Z X. Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines. *IEEE Transactions on Intelligent Transportation Systems*, 2016, **17**(4): 1096−1107

7 Liu Y Q, Wang K F, Shen D Y. Visual tracking based on dynamic coupled conditional random field model. *IEEE Transactions on Intelligent Transportation Systems*, 2016, **17**(3): 822−833

8 Goyette N, Jodoin P M, Porikli F, Konrad J, Ishwar P. A novel video dataset for change detection benchmarking. *IEEE Transactions on Image Processing*, 2014, **23**(11): 4663−4679

9 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

10 INRIA person dataset [Online], available: http://pascal.inrialpes.fr/data/human/, September 26, 2016.

11 Caltech pedestrian detection benchmark [Online], available: http://www.vision.caltech.edu/Image_Datasets/Caltech-Pedestrians/, September 26, 2016.

12 The KITTI vision benchmark suite [Online], available: http://www.cvlibs.net/datasets/kitti/, September 26, 2016.

13 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25 (NIPS 2012). Nevada: MIT Press, 2012.

14 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, **521**(7553): 436−444

15 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to be published

16 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016. 770−778

17 ImageNet [Online], available: http://www.image-net.org/, September 26, 2016.

18 The PASCAL visual object classes homepage [Online], available: http://host.robots.ox.ac.uk/pascal/VOC/, September 26, 2016.

19 COCO — Common objects in context [Online], available: http://mscoco.org/, September 26, 2016.

20 Torralba A, Efros A A. Unbiased look at dataset bias. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado, USA: IEEE, 2011. 1521−1528

21 Bainbridge W S. The scientific research potential of virtual worlds. *Science*, 2007, **317**(5837): 472−476

22 Miao Q H, Zhu F H, Lv Y S, Cheng C J, Chen C, Qiu X G. A game-engine-based platform for modeling and computing artificial transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 2011, **12**(2): 343−353

23 Sewall J, van den Berg J, Lin M, Manocha D. Virtualized traffic: reconstructing traffic flows from discrete spatiotemporal data. *IEEE Transactions on Visualization and Computer Graphics*, 2011, **17**(1): 26−37

24 Prendinger H, Gajananan K, Zaki A B, Fares A, Molenaar R, Urbano D, van Lint H, Gomaa W. Tokyo Virtual Living Lab: designing smart cities based on the 3D Internet. *IEEE Internet Computing*, 2013, **17**(6): 30−38

25 Karamouzas I, Overmars M. Simulating and evaluating the local behavior of small pedestrian groups. *IEEE Transactions on Visualization and Computer Graphics*, 2012, **18**(3): 394−406

26 Qureshi F, Terzopoulos D. Smart camera networks in virtual reality. *Proceedings of the IEEE*, 2008, **96**(10): 1640−1656

27 Starzyk W, Qureshi F Z. Software laboratory for camera networks research. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2013, **3**(2): 284−293

28 Sun B C, Saenko K. From virtual to reality: fast adaptation of virtual object detectors to real domains. In: Proceedings of the 2014 British Machine Vision Conference. Jubilee Campus: BMVC, 2014.

29 Hattori H, Boddeti V N, Kitani K, Kanade T. Learning scene-specific pedestrian detectors without real data. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, Massachusetts: IEEE, 2015. 3819−3827

30 Vázquez D, López A M, Marín J, Ponsa D, Gerónimo D. Virtual and real world adaptation for pedestrian detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(4): 797−809

31 Xu J L, Vázquez D, López A M, Marín J, Ponsa D. Learning a part-based pedestrian detector in a virtual world. *IEEE Transactions on Intelligent Transportation Systems*, 2014, **15**(5): 2121−2131

32 Gaidon A, Wang Q, Cabon Y, Vig E. Virtual worlds as proxy for multi-object tracking analysis. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016. 4340−4349

33 Handa A, Pătrăcean V, Badrinarayanan V, Stent S, Cipolla R. Understanding real world indoor scenes with synthetic data. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016. 4077−4085

34 Ros G, Sellart L, Materzynska J, Vazquez D, López A M. The SYNTHIA dataset: a large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV: IEEE, 2016. 3234−3243

35 Movshovitz-Attias Y, Kanade T, Sheikh Y. How useful is photo-realistic rendering for visual learning? arXiv: 1603.08152, 2016.

36 Haines T S F, Xiang T. Background subtraction with Dirichlet process mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(4): 670−683

37 Sobral A, Vacavant A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, 2014, **122**: 4−21

38 Morris B T, Trivedi M M. Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(11): 2287−2301

39 Butler D J, Wulff J, Stanley G B, Black M J. A naturalistic open source movie for optical flow evaluation. In: Proceedings of the 12th European Conference on Computer Vision (ECCV). Berlin Heidelberg: Springer-Verlag, 2012.

40 Kaneva B, Torralba A, Freeman W T. Evaluation of image features using a photorealistic virtual world. In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011. 2282−2289

41 Taylor G R, Chosak A J, Brewer P C. OVVV: using virtual worlds to design and evaluate surveillance systems. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis, MN, USA: IEEE, 2007. 1−8

42 Zitnick C L, Vedantam R, Parikh D. Adopting abstract images for semantic scene understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(4): 627−638

43 Veeravasarapu V S R, Hota R N, Rothkopf C, Visvanathan R. Model validation for vision systems via graphics simulation. arXiv: 1512.01401, 2015.

44 Veeravasarapu V S R, Hota R N, Rothkopf C, Visvanathan R. Simulations for validation of vision systems. arXiv: 1512.01030, 2015.

45 Qingdao "Integrated Multi-Mode" Parallel Transportation Operation Demo Project. Notice from National Development and Reform Commission. [Online], available: http://www.ndrc.gov.cn/zcfb/zcfbtz/201608/t20160805_814065.html, August 5, 2016
(国家发展改革委, 交通运输部. 青岛市 "多位一体" 平行交通运用示范. 国家发展改革委交通运输部关于印发《推进 "互联网+" 便捷交通促进智能交通发展的实施方案》的通知 [Online], http://www.ndrc.gov.cn/zcfb/zcfbtz/201608/t20160805_814065.html, August 5, 2016)

46 Yuan G, Zhang X, Yao Q M, Wang K F. Hierarchical and modular surveillance systems in ITS. *IEEE Intelligent Systems*, 2011, **26**(5): 10−15

47 Jones N. Computer science: the learning machines. *Nature*, 2014, **505**(7482): 146−148

48 Silver D, Huang A, Maddison C J, Guez A, Sifre L, van den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D. Mastering the game of Go with deep neural networks and tree search. *Nature*, 2016, **529**(7587): 484−489

49 Wang F Y, Zhang J J, Zheng X H, Wang X, Yuan Y, Dai X X, Zhang J, Yang L Q. Where does AlphaGo go: from Church-Turing Thesis to AlphaGo Thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, 2016, **3**(2): 113−120

50 Huang W L, Wen D, Geng J, Zheng N N. Task-specific performance evaluation of UGVs: case studies at the IVFC. *IEEE Transactions on Intelligent Transportation Systems*, 2014, **15**(5): 1969−1979

51 Li L, Huang W L, Liu Y, Zheng N N, Wang F Y. Intelligence testing for autonomous vehicles: a new approach. *IEEE Transactions on Intelligent Vehicles*, 2016, to be published

王坤峰 中国科学院自动化研究所复杂系统管理与控制国家重点实验室副研究员. 2008 年获得中国科学院研究生院博士学位. 主要研究方向为智能交通系统, 智能视觉计算, 机器学习.
E-mail: kunfeng.wang@ia.ac.cn
(**WANG Kun-Feng** Associate professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. He received his Ph. D. degree from the Graduate University of Chinese Academy of Sciences in 2008. His research interest covers intelligent transportation systems, intelligent vision computing, and machine learning.)

苟 超 中国科学院自动化研究所复杂系统管理与控制国家重点实验室博士研究生. 2012 年获得电子科技大学学士学位. 主要研究方向为智能交通系统, 图像处理, 模式识别.
E-mail: gouchao2012@ia.ac.cn
(**GOU Chao** Ph. D. candidate at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. He received his bachelor degree from the University of Electronic Science and Technology of China in 2012. His research interest covers intelligent transportation systems, image processing, and pattern recognition.)

王飞跃 中国科学院自动化研究所复杂系统管理与控制国家重点实验室研究员. 国防科学技术大学军事计算实验与平行系统技术研究中心主任. 主要研究方向为智能系统和复杂系统的建模、分析与控制. 本文通信作者.
E-mail: feiyue.wang@ia.ac.cn
(**WANG Fei-Yue** Professor at the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. Director of the Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology. His research interest covers modeling, analysis, and control of intelligent systems and complex systems. Corresponding author of this paper.)