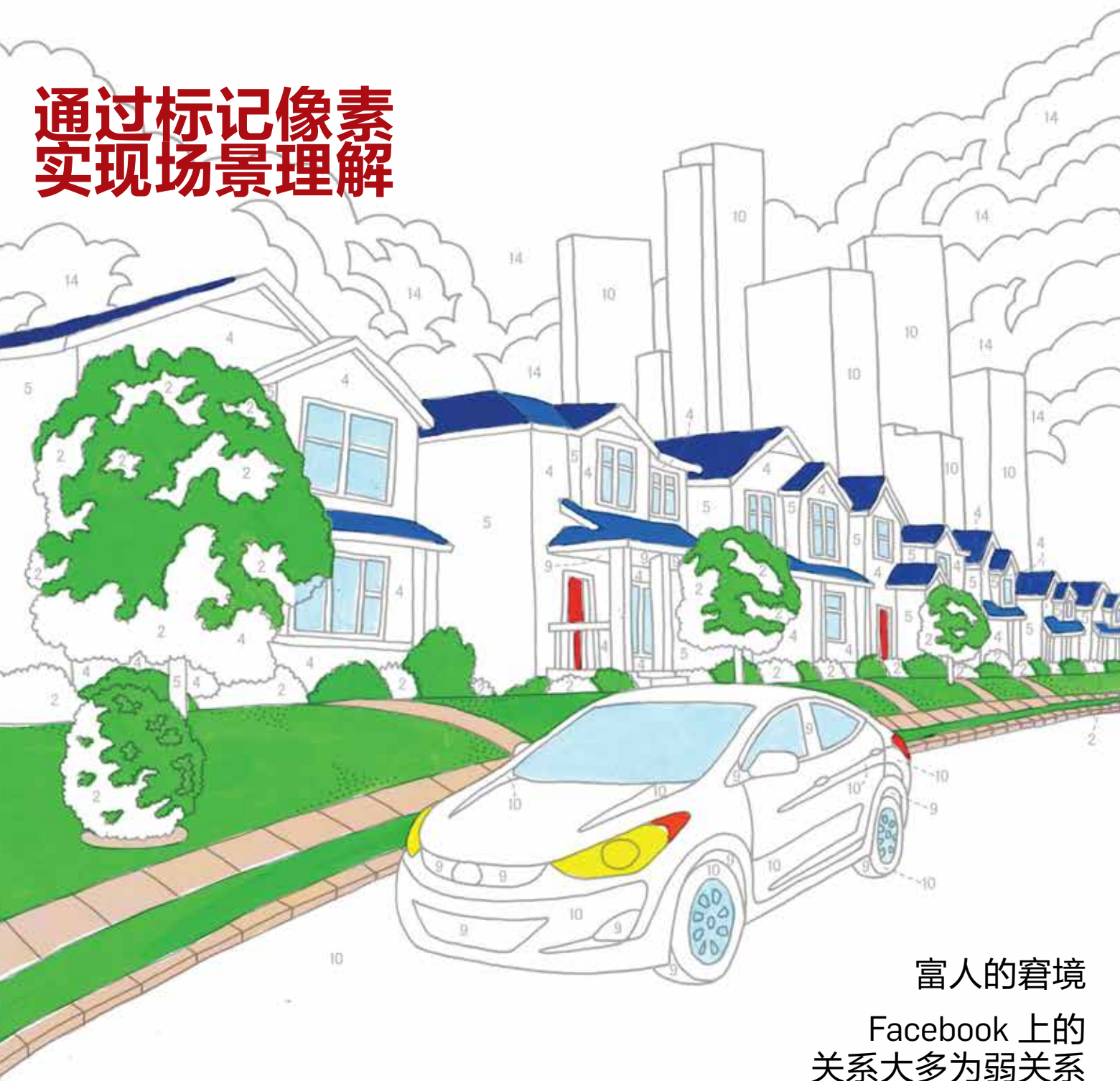


计算机协会通讯

CACM.ACM.ORG

2014年11月第57卷第11期

通过标记像素 实现场景理解



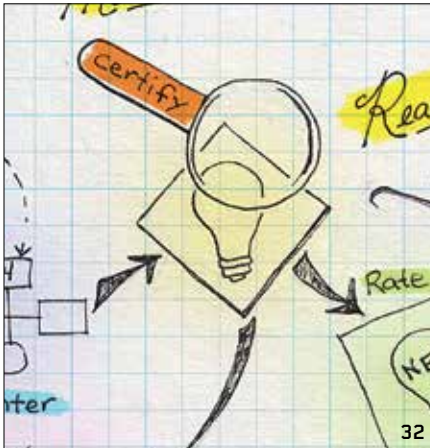
富人的窘境

Facebook 上的
关系大多为弱关系

织物与光之谜舞

并行编程效率
取得进展的 10 年

观点



32 观点

富人的窘境

关于开放式创新体系的批判性评论。

Mark Klein,

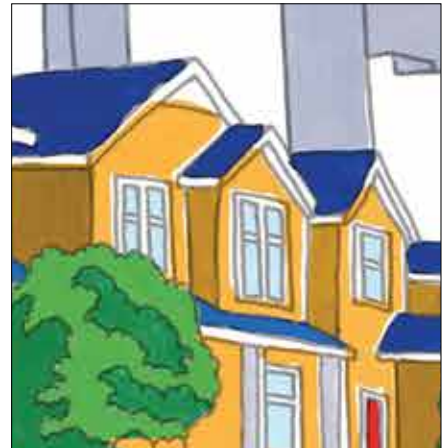
Gregorio Convertino

实践

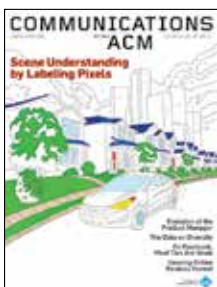
60 并行编程效率取得进展的 10 年
探讨 X10 的设计和优势。John T. Richards、Jonathan Brezin、
Calvin B. Swart 和

Christine A. Halverson

投稿文章

68 通过标记像素实现场景理解
用场景的语义和几何结构标注像素，
让计算机述其所见
Stephen Gould, Xuming He78 Facebook 上的关系大多为弱关系
强关系连接了相同社区中的个人；
弱关系连接的不同社区中的个人。
Pasquale De Meo,
Emilio Ferrara、Giacomo Fiumara
以及 Alessandro Provetti

研究亮点

97 技术视角
织物与光之谜舞
Szymon Rusinkiewicz98 使用微机 CT 造影构建织物的
体外观模型
Shuang Zhao, Wenzel Jakob,
Steve Marschner, Kavita Bala**关于封面：**

创造能够自动理解视觉场景的计算机，是一个困扰人工智能和计算机视觉领域的长期问题。本月的封面故事（第 68 页）提出了一个前景瑰丽的方法，通过用标签来注释一个场景中的每个像素说明相关的时间与地点。封面插图由 Andrew Joyce 提供。



Association for Computing Machinery
Advancing Computing as a Science & Profession

ACM计算机通讯(中文版)编审委员会

主席



陈文光
清华大学
cwg@tsinghua.edu.cn

并行计算和编程语言

陈文光教授现任清华大学计算机科学与技术系教授、副主任。

委员



陈海波
上海交通大学
haibo.chen@sjtu.edu.cn

操作系统和计算机体系结构

陈海波教授就职于上海交通大学软件学院。



崔斌
北京大学
bin.cui@pku.edu.cn

数据库

崔斌教授就职于北京大学信息科学技术学院, 并担任网络与信息系统研究副所长。



陈贵海
上海交通大学
gchen@cs.sjtu.edu.cn

上海交通大学计算机科学与工程系教授; 中国计算机学会开放系统专委会主任; 在并行与分布式计算领域有广泛的兴趣, 特别是各种网络系统, 例如无线传感器网络, 对等覆盖网络, 数据中心网络, 社交网络等。



李向阳
伊利诺理工学院
xli@cs.iit.edu

李向阳教授就职于伊利诺理工学院。他是中国国家自然科学基金海外杰出青年学者奖的获得者。



刘云浩
清华大学
yunhao@greenorbs.com

刘云浩教授现任清华大学长江特聘教授。他还担任ACM中国理事会主席。



山世光
计算技术研究所
sgshan@ict.ac.cn
计算机视觉和图案识别

山世光教授就职于中国科学院计算技术研究所(ICT)。



孙晓明
计算技术研究所
sunxiaoming@ict.ac.cn

理论

孙晓明教授就职于中国科学院计算技术研究所。



唐杰
清华大学
jietang@tsinghua.edu.cn
数据挖掘

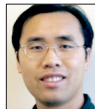
唐杰副教授就职于清华大学计算机科学与技术系。



田丰
中国科学院软件研究所
tianfeng@iscas.ac.cn

人机交互

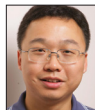
田丰教授就职于中国科学院软件研究所, 他还担任计算机协会中国人机交互学会主席。



谢涛
伊利诺伊大学厄巴纳-香槟分校
taoxie@illinois.edu

软件工程

谢涛副教授就职于美国伊利诺伊大学厄巴纳-香槟分校计算机科学系。



周昆
浙江大学
kunzhou@acm.org

计算机图形和虚拟现实

周昆教授是长江特聘教授, 浙江大学CAD&CG国家重点实验室主任。



诸葛建伟
清华大学
zhugejw@cernet.edu.cn

计算机安全

诸葛建伟副教授就职于清华大学网络科学与网络空间研究院。

ACM中国理事会

孙家广, 名誉主席
刘云浩, 主席
沈运申, 副主席, 分会
陈文光, 副主席, 出版物
王新兵, 副主席, 会议
万猛, 副主席, 宣传与公共关系
张铭, 常务理事
肖人毅, 常务理事
吕自成, 常务理事
秦志光, 常务理事
罗军舟, 常务理事
胡传平, 常务理事
胡斌, 常务理事
赵峰, 常务理事

ACM中国指导委员会

孙家广, 主席
李志民, 联席主席
姚期智
廖湘科
王珊
怀进鹏
梅宏
吕健
郑南宁
张尧学
林惠民

分会主席

上海分会 胡传平
南京分会 罗军舟
成都分会 秦志光
兰州分会 胡斌
重庆分会 廖晓峰
长沙分会 卢凯
广州分会 张军
济南分会 杨波
武汉分会 金海
大连分会 罗钟铉

ACM中国理事会办公室

中国北京清华大学
东主楼 11-236 室
邮编: 100084
电话: +86-10-62785025
电子邮件: acmchina@acm.org
联系人: 辛爽

ACM通讯

(ISSN 0001-0782) 由计算机协会
(2 Penn Plaza, Suite 701, New
York, NY 10121-0701) 按月发行。

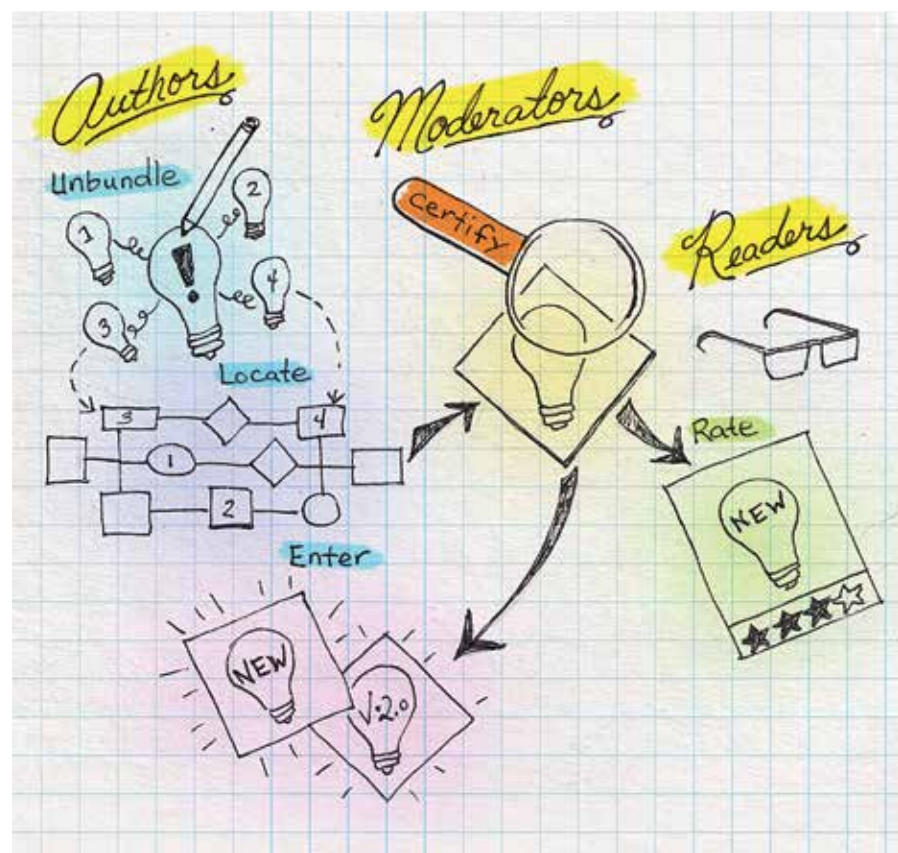


Association for
Computing Machinery

观点 富人的 窘境

关于开放式创新体系的批判性评论。

开放式创新体系是基于集群智慧的新兴成功事例。在该体系中，客户提出待解决的问题（例如，“我们征集新品饮料的创意”），并向大众提供上交解决方案及评价（或批评）他人解决方案的在线工具。许多开放式创新平台纷纷涌现（如 Ideascale、Spigit 和 Imaginatik），并已广泛应用于多种场景——从 IBM 到星巴克，从丹麦中央政府到白宫。最近的一项调查²发现，四分之一的企业计划在未来 12 个月内利用开放式创新体系，而这个数字还在不断增长。实践证明，这样的体系能够以极大的规模、极低的成本带来实质性贡献。例如，美国总统奥巴马在第一任任期的前几周内，邀请美国公民在 change.gov 网站上提问和投票，并承诺在重大新闻发布会上回答每个类别的前五个问题。这项举措吸引了超过 10 万名贡献者，他们共提交了超过 7 万个问题，并投出了 400 多万票。谷歌的“从十到百”（10 to the 100th）项目就如何开展谷歌慈善捐款征集到了超过 15 万条建议。在 2006 年 IBM 的“创意马拉松”（Idea Jam）活动中，15 万贡献者就 IBM 可能推出的产品和服务提



思维导图¹有助于实现更出色组织决策。

出了 46,000 个创意。这种大规模的参与度反过来也催生了惊人的新生现象，例如：

- ▶ **长尾效应**：比起能轻易采用的其他方式而言，大众能产生更多元化的创意，包括具有潜在突破性的“跳脱思维定势”的贡献。
- ▶ **创意协同**：通过将其他参与者

的创意重新加以组合和完善，众可以迅速提出大量新颖创意。大

- ▶ **许多双眼睛**：大众参与者通过彼此相互核查并指证错误，能够以极低的成本获得质量极高的结果。
- ▶ **大众的智慧**：相比个体的判断，大众的集体判断可能更为出色，表现往往胜过专家。

然而，开放式创新体系面临着严峻的挑战，而令人颇感矛盾的是，这种挑战在很大程度上又是由于成功吸引大量参与者引起的。在本书《观点》中，我们将讨论这些挑战，并提出一些前景光明的方向。

挑战

开放式创新体系在创意产生和创意评价两个环节都面临挑战。创意产生过程中的主要挑战包括：

- ▶ **“收割”成本**：开放式创新项目产生的创意语料库往往庞大、无序、极度过剩。“修剪”这样的清单来寻找最好的创意，可能是一项艰巨的任务。例如，谷歌的“从十到百”项目投入了 3,000 名员工来“修剪”他们收到的 15 万个创意，导致项目比原定计划推迟了九个月。IBM 的 100 名高级管理人员从全球各地飞往纽约，只是为了精简“创意马拉松”活动的结果。

- ▶ **杂乱无章的覆盖面**：开放式创新缺乏内在机制来确保提交的创意能够全面覆盖待解决问题的最关键方面，因此覆盖面非常随意，可能难以满足客户需求。

- ▶ **浅薄**：初浅开放式创新体系倾向于提出海量的相对初浅观点。我们认为，主要原因是目前的软件工具不能很好地支持和实现，无法实现协作性的创意发展和准确的奖励分配。

开放式创新体系也面临着众包式创意评价带来的挑战：通常情况下客户的需求与大众选择的方案之间存在差异。这种现象有多个原因：

- ▶ **初浅的评价**：由于用户之间几乎不检查和纠正别人的事实和推理，系统在利用彼此评价专长上为大众提供的支持非常有限。

- ▶ **评级局限**：当有成千上万的创意时，许多潜在有价值的创意可能不会被全面深入地进行评估，该体系很可能会迅速“锁定”于一个相

开放式创新体系在创意产生和创意评价两个环节都面临挑战。

对稳定且随意性很强的评价体系，胜出的创意有可能不如其他创意。

因此，开放式创新体系面临的严峻挑战是：如何确保从众多潜在的有价值的创意中，能够为客户选择最有价值的，并且又不会产生高昂的“收割”成本的创意。

光明前景

我们如何应对这些挑战？如何更充分地实现开放式创新体系的承诺？我们认为，改变现状，最重要的是对以下两个方面的改进。

更好的开放式创新进程。新的开放式创新进程，能够为大众做出最佳创意提供有力的指导，实现成员间的信息共享，简化创意的成型过程。

- ▶ **协作式创意界定**：要想帮助大众提出成熟的创意，要求系统有良好的激励机制和协作式创作架构。例如，可以要求参与者以思维导图¹的形式构建他们的创意树，包括：要解决的问题、针对这些问题的可能的解决方案，以及对每个可能解决方案的赞成和反对意见。每个参与者都要完成以上内容。随后，参与者可以利用每张地图上的最佳解决方案创意来撰写方案。奖励制度变得一目了然，因为每个方案的作者都是明确的。

- ▶ **新颖的评级机制**可以有助于大众根据客户推崇的标准，快速准确地评估各种创意。例如，可以使用

活动日程

11月16-19日

交互式桌面和界面，德国德累斯顿，赞助：SIGCHI，联系人：Raimund Dachselt，电子邮件：raimund.dachselt@tu-dresden.de

11月16-22日

第22届ACM SIGSOFT 软件工程基础研讨会，中国香港，赞助：SIGSOFT，联系人：Shing-Chi Cheung，电子邮件：sccheung@cse.ust.hk

11月30日-12月2日

第13届ACM 虚拟现实连续体及其行业应用国际会议，中国深圳，赞助：SIGGRAPH，联系人：Jin Xiaogang，电子邮件：jin@cad.zju.edu.cn

12月2-5日

新兴网络体验与技术会议，澳大利亚悉尼，联系人：Aruna Seneviratne，电子邮件：a.seneviratne@unsw.edu.au

12月8-12日

第15届国际中间件会议，法国波尔多，联系人：Laurent Reveillere，电子邮件：reveillere@labri.fr

12月13-17日

第47届年度IEEE/ACM 微体系结构国际研讨会，英国剑桥，赞助：SIGMICRO，联系人：Krisztian Flautner，电子邮件：krisztian.flautner@arm.com

1月8-10日

第9届泛在信息管理和通信国际会议，印度尼西亚巴厘，赞助：SIGAPP，联系人：Sukhan Lee，电子邮件：lsh@ece.skku.ac.kr

某种预测性市场，在这个市场中，参与者获得有限的预算来买卖股票，每种股票分别代表不同的创意，当他们“购买”的创意被客户成功实施后，就能得到回报。这样一来，就可以激励用户从客户的角度认真评价创意。思维导图在此处同样能够发挥作用，因为它可以通过创建支持性论点和反驳性论点的链条，帮助大众互相检查和利用彼此的推理。

▶ **创造力增强技术**（目前为止这些技术几乎完全用于面对面的团队环境）可以离线使用，从而为开放式创新项目提供支持，也能够调整适应大众尺度的在线场景。

▶ **在多个开放式创新系统中进行交替多回合创意构思和评价。**同时，可以要求大众的创意必须基于前一轮筛选中胜出的创意，这样创意的产生就更有可能侧重于客户需求。

更深层次的计算机支持。群（人）和云（计算机）具备协同能力。大众对创意的构思、理解和评价方式都是计算机无法比拟的，但只是最适合执行那些小而快、不需要大量背景的任务。相比之下，计算机擅长快速分析海量数据，从全局角度揭示发生（和未发生）的事。要想结合两者的优势，需要弥合众包者使用的自然语言与计算机使用的形式语言之间的语义鸿沟。总有一天，这个目标可以通过先进的算法来实现。这种算法能够让计算机深层次理解自然语言。不过，要实现这一点，似乎还有很长的路要走。现在，我们的目标是寻求方法实现众包者能够：（1）工作量最简化，（2）让计算机执行大量操作。例如通过以下方式：

▶ **语义标记：**众包者可以利用语义线索（例如创意和论点边界、主题关键词）来为自然语言的创意语料库添加注释。这可以是一个混合主导过程，即计算机提议可能的标记，然后由众包者进行纠正，并且

随着众包者和云之间的语义差距日渐缩小，我们可以为开放式创新创建强大而全新的计算机支持形式。

使用机器学习技术来根据这一人类反馈持续改善计算机算法。

▶ **设计工具**让用户可以将自己的创意表现为基于“领域特定原语”（domain-specific primitives）构建的半形式化模型，而不仅仅是自然语言文本。大众的设计工具（例如，谷歌的 SketchUp）为流行的设计与创意平台融合的工具，但尚未纳入开放式创新平台。

随着众包者和云之间的语义差距日渐缩小，我们可以为开放式创新创建强大而全新的计算机支持形式：

▶ **分析工具，**利用半形式化的创意表达方式，有助于评价创意的优劣；

▶ **语义压缩算法，**删除重复创意，将相关的创意集中在一起，以压缩创意语料库；以及

▶ **利用可视化工具，**对众包者已做工作进行总结，方便用户明确差距和发展前景，通过重点引导，为今后的工作提供有指导意义的信息。

战斗的号角

综上所述，开放式创新体系，具有集众所长的潜质。它能解决从商业到政府部门等，各行各业问题的能力。然而，这种潜力远未得到充分实现，这很大程度上是因为我们无

法有效应对此类系统所产生的海量用户多层次的创意。要在这方面取得进步，需要许多学科的通力合作，包括计算机科学、认知科学、社会心理学、计算语言学和经济学。您愿意跟我们共同努力迎接挑战吗？

□

延伸阅读

Bailey, B.P. and Horvitz, E. What's your idea? A case study of a grassroots innovation pipeline within a large software company. In *Proceedings of CHI 2010*, ACM Press, NY, 2010.

Bason, C. *Leading Public Sector Innovation: Co-creating for a Better Society*. Policy Press, 2010.

Bjelland, O.M. and Chapman Wood, R. An inside view of IBM's innovation jam. *MIT Sloan Management Review* 50, 1 (2008), 32-40.

Chesbrough, H., Vanhaverbeke, W., and West, J., Eds. *Open Innovation: Researching a New Paradigm*. Oxford University Press, Oxford, U.K., 2006.

Gulley, N. Patterns of innovation: A web-based MATLAB programming contest. *Human Factors in Computing Systems* (2001), 337-338.

Jouret, G. Inside Cisco's Search for the Next Big Idea. *Harvard Business Review* 87, 9 (2009), 43-45.

Lakhani, K.R. and Jeppesen, L.B. Getting unusual suspects to solve R&D puzzles. *Harvard Business Review* 85, 5 (2007), 30-32.

von Hippel, E. *Democratizing Innovation*. MIT Press, 2005.

参考资料

1. Klein, M. and Iandoli, L. Supporting collaborative deliberation using a large-scale argumentation system: The MIT collaboratorium. *Directions and Implications of Advanced Computing; Conference on Online Deliberation (DIAC-2008/OD2008)*. University of California, Berkeley, 2008.
2. Thompson, V. *IDC MarketScape: Worldwide Innovation Management Solutions 2013 Vendor Analysis*, (2013); http://idcdocserv.com/240823_spigit.

Mark Klein (m_klein@mit.edu) 是麻省理工学院集体智慧中心的首席科学研究员，也是麻省理工学院科学和人工智能实验室、新英格兰复杂系统研究所以及瑞士苏黎世大学动态和分布式信息系统小组的研究员。

Gregoria Convertino (gconvertino@informatica.com) 是加州雷德伍德城 Informatica 公司的高级用户研究员。

译文责任编辑：李向阳

版权归属于作者。

探讨 X10 的设计和优势。

作者 JOHN T. RICHARDS、JONATHAN BREZIN、CALVIN B. SWART
和 CHRISTINE A. HALVERSON

并行编程效率 取得进展的 10 年

2002 年，美国国防部高级研究计划局 (DARPA) 启动了一项高效能计算系统 (HPCS) 的重大创新计划。推动这些计划的原因是因为人们相信编写、调试、调优和维护 peta 级 (译注: 1 petaflops = 10^{15} flops, 即每秒千万亿次浮点运算) 软件的难度阻碍了人们对下一代并行机的利用。

作为此计划的一部分，DARPA 鼓励在新的编程语言、运行时和工具方面有所建树。人们相信，只要使并行构造的表达变得更容易、使运行时模型与仍在开发的异构处理器体系结构相匹配，以及提供强大的集成开发工具，程序员效率就会提高。这个猜想很合理，但是我们不想停留在猜想，而是实际测量效率提高了多少。

虽然测量程序员效率还没有既定的方法，但是有一点很清楚，那就是效率指标必须采用比率的形式，即：实现的编程成果与取得这些成果的代价之比。这里，成果的定义是成功创建一组可在工作站的两个处理器核上正确运行的并程序。这虽然距离达到 peta 级规模还很遥远，但是由于新的并行软件通常是先在两个内核上开始的（然后在更多数量的处理器上进行扩展和调优），所以我们将其视为合理的近似。此外，编写几乎任何并行应用程序代码（无论规模有多小）的人应该会对用两个核得到的成果感兴趣。定义了“成果”后，那么“代价”的定义就更加简单，因为可以合理地将其近似为创建这组并程序所需的时间。

这次研究的目的是测量从 2002 年 (HPCS 创新计划开始于该年) 到 2012 年的大部分时间里程序员的效率 (按前面的定义)。比较的重点主要是放在比较两种并行编程方式上，即：以 C/MPI (消息传递接口) 为典型的 SPMD (单程序流多数据流) 模型，以及受到 X10 (<http://x10-lang.org>) 等新语言支持的 APGAS (异步分区全局地址空间) 模型，虽然我们同时也研究了环境和工具的差异。请注意，比较不是在曾经的 C/MPI 与现在的 X10 之间作比较，而是将 2002 的情况与现在的情况做历史比较。其实，C++ 及其异常以及 MPI-2 及其单边通信协议很可能提高程序员效率，并且其本身就值得研究。

根据我们的目标，我们力求为 C/MPI 用户尽可能接近地复制 2002 年的编程环境。这包括 gdb 调试



器，以及一组典型的命令行工具。对于 X10，我们使用带 X10 插件的 Eclipse (<http://www.eclipse.org>)，因为它是在 2010 年出现的，也就是本研究开始的时间。X10 是作为 HPCS 创新计划的一部分开发的。它将 Scala (<http://www.scala-lang.org>) 等较新型语言的简洁性，与能够很好地映射到现代并行机 PGAS 模型的并发编程模型融为一体。十年后，该语言仍在发展，但是在本

研究开始时，其基本模型还是稳定的。重要的是，X10 调试器是在本研究完成之后才出现的，所以当前 X10 用户可以期待比这里所报道的更大的效率提高。

以前的研究

本次研究不是尝试测量 X10 编程效率的第一次研究。Ebcioglu 等人⁵以及 Danis 和 Halverson³描述了这样一场实验，实验中一群本科生（并

行编程新手）接受了 C/MPI、UPC (Unified Parallel C) 或 X10（非常早期的版本）的培训，然后尝试并行化 Bader 等人²所述的 SSCA 1（可伸缩合成紧凑应用程序 1）字符串匹配算法。其平均完成时间大约为 C/MPI 510 分钟、UPC 550 分钟，以及 X10 290 分钟。去掉测试期间代码执行时间的差异后，这几个时间大约为 C/MPI 360 分钟、UPC 390 分钟，以及 X10 180 分钟。然而，

双内核上首次成功并行运行所需的天数以及完整程序的代码行数。

程序	X10 天数	X10 行数	C/MPI 天数	C/MPI 行数
SSCA 1	9.0	1625	20	1225
SSCA 2	7.9	1150	28	3000
消费者 - 生产者	5.0	1275	24	1750
非平衡树搜索	7.5	1125	30	2200
Floyd 算法	3.8	520	12	1400
离散傅立叶变换	5.6	500	15	670
总计	38.8	6195	129	10,245

要解释这种大约两倍的效率提升有点复杂，因为缺少明确的代码成功完成的标准。

在随后的研究中，Halverson 和 Danis⁶ 将 Eclipse 编程环境添加到 X10 程序员可用的工具中。这次研究的参与者更有经验，他们在以前的课程中已经得到了一些并行编程培训。这次同样得到了 X10 和 Eclipse 效率高于 C/MPI 的结果，但是由于在 7 个 C/MPI 参与者中只有一个人完成了任务（407 分钟），而在 7 个 X10 参与者中，有 5 人完成了任务，平均用时 321 分钟，因此这造成要评估效率究竟高出多少变得复杂。

在这两次早期研究中，只编写了一个程序（SSCA 1 的一部分）、只测试了相对新手的并行程序员，并且任务只限于并行化在可运行的串行代码中提供给测试参与者的易并行算法。所有这三个缺陷将在本文报告的评估中加以讨论。

实验方式

在一项长期研究中，我们测量了开发 6 个并行程序的时间，每个程序用两种语言，一共 12 个程序。每个程序都从问题陈述和 / 或算法描述开始开发，而没有任何预先存在的代码示例可供参考。第一次同时使用双核工作站上的两个内核成功并行运行时，即告计时完成。

六个程序中两个是在 DARPA HPCS 计划的第 2 阶段，SSCA 1 和 SSCA 2 中定义的，并且在 Bader 等

人²的文章中有所介绍。SSCA 1 是在前面的两个参考研究中使用的字符串匹配问题。但是在本次研究中，用每一种语言开发了 SSCA 1 的两个版本：一个是易并行版本，当目标字符串远远短于搜索字符串时，该版本效果很好；另一个是更复杂的反对角版本，此版本更适合两个字符串长度大致相等的情况。SSCA 2 是这次评估新引进的，我们用每一种语言编码了全部四个程序核心。

除了这两个 SSCA 外，我们还根据 Berkeley Motifs¹ 的分析，又定义了四个问题。我们将全部介绍这六个问题。

开发这 12 个程序历时近一年。虽然让多名熟练的程序员来执行此任务可能更理想，但是这会大大超出可用的资金。实际上，研究经费只允许一个程序员（我们效率评估团队的一员和本报告的其中一名作者）作为测试参与者。他是一位数学博士，经过训练，他从 1979 年起就专业从事 C 编程，曾编写过 IBM 第一个 C 编译器的前端，从 2007 年起，开始编写多核系统的 MPI 程序。他同时在 2007 年开始 X10 编程，开发了 Halverson 和 Danis⁶ 的研究测试床，并撰写了 X10 教程的几个章节。虽然很容易将从一个程序员身上得出的结果看成是偶然，但是观察到的效率提高跟从事更简单任务的新手程序员身上发现的结果差不多，而且鉴于该程序员更精通 C，且在研究时尚缺

乏 X10 调试器，因此这个结果很可能还相当保守。

程序。这次研究定义了六个编程问题，每个问题代表一类并行应用程序。它们合起来涵盖了很大范围的典型小并行程序：

- **SSCA 1（第一内核）。**这涉及一个由染色体和蛋白质研究所引发的字符串匹配问题。它要求找到一对大写字母字符串的子串之间的最佳近似匹配。

- **SSCA 2（所有四个程序核心）。**这个问题涉及一个稀疏图的有效表示，这个图的边集事先已知将随机分布在一组可用的处理器上。第一个程序核心用于构造图，图的格式可用作所有后续程序核心的输入。第二个程序核心从图中按权重提取边，并形成有最大权重的边的列表。第三个程序核心提取一系列子图，这些子图是沿着从一组初始顶点开始的指定长度的路径形成的。最后一个程序核心计算网络介数指标，该指标识别图中最短路径上极为重要的顶点。

- **消费者 - 生产者。**此问题使用一个进程作为服务器，服务器管理一个多客户端进程的队列，这些进程随机加入和离开队列。由于所有操作的目标都是同一个缓冲区，因此首要编程问题是最大程度减少争用。由于将元素添加到队列和从队列中移出元素发生在队列的两端，因此这成了避免争用的主要机会；如果队列足够大，就可以安全地同时执行两种操作。

- **非平衡树搜索。**与前面的问题（即，通过中心服务器进程管理协调）相对，这个问题是一组对等进程各自管理自己的工作队列，当需求上升时，与其他进程分担工作。整个任务是树的广度优先搜索，且已知这棵树的叶距离根的深度极其不同。每个处理器管理各自未访问树节点的队列，但是空闲的对等处理器可能要求其提供一些节点来供其追查。

► **Floyd 算法**。给定一个有权有向无环图（权重非负），计算所有成对顶点之间最短路径的权重和。

► **离散傅立叶变换**。实现的算法是最早的 Cooley-Tukey 快速傅立叶变换。输入和输出都是 2^n 复数数组。至少在 $n=3$ 且 $2^n=8$ 之前，传统的串行算法击败了 Cooley-Tukey，也就是说，将该数组分布到 $2^n/8$ 个进程可能有利。

观察到的效率提升

附表总结了这 12 个程序中的每个程序首次并行运行成功的天数，以及所写代码的行数。X10 的六个程序需要总共 39 天的开发时间。C/MPI 的六个程序需要总共 129 天的开发时间。因此，因语言（其次是环境）差异而产生的总体效率提高超过了 3 倍。在编写 X10 的 39 天里，一共编写了 6,195 行代码，总体上每天 159 行代码。在结合 MPI 编写 C 的 129 天里，总共编写了 10,245 行代码，总体上每天 79 行代码。虽然我们没有测量这些程序的性能，但是 Saraswat 等人⁷的研究调查了并行性能非常出色的非平衡树搜索程序的 X10 实现，这个程序跟我们的程序非常相似。

促进效率的因素

如何解释这次研究发现的 3 倍性能提高？虽然部分提高是因为 Eclipse 和 X10 插件提供的集成工具，但最主要的原因应归因于 X10 语言的功能。这些功能包括 activity 和 place 提供的任务和划分、async 和 finish 提供的灵活任务支持，以及巧妙集成的异常处理机制。这些功能合在一起，简化了并行性的表达，并且可以使程序结构更贴近地反映原有问题结构。此外，自动垃圾回收省去了管理内存的苦差事，而面向对象简化了处理复杂结构的任务，同时避免了误算内存引

用。下面几节提供的示例将说明这些功能如何促进效率。

表达并行性。X10 程序先从一个位置 (place) 执行的一个活动 (activity) 开始。活动是一个顺序的执行线程，位置是将处理能力与一部分分区全局内存相结合的系统片。任何活动可以根据需要在其当前正在运行的位置，或者任何其他方便的位置发起新的活动。

本地线程。举个简单例子，SSCA 1 需要从某种外设源读取一对字符串。如果字符串很长（因为此问题意在扩展到任意长度的字符

串，因此它们可能很长），那么并发读取两者比较理想。如果要从同一个处理器访问这两个流，那么您需要的 X10 代码如图 1 所示。

可以在 X10 中实现简单的分叉-汇集 (fork-join) 并行性，使用 async 语句进行分叉，将一组活动包含在 finish 块中实现汇集。包含完整计算的 try 块捕获可能在文件 I/O 或活动启动 / 关闭时发生的任何错误。

对 C 来说，根本问题是如何获得与原有线程共享内存的第二个线程。图 2 显示了可以使用 POSIX 线

图 1. 并行读取两个流所需的 X10 代码。

```
try {
    val seq1:Array[Byte](1); // seq1 和 seq2 都是字节数组
    val seq2:Array[Byte](1); // 按一个整数索引
    finish { // 控制权只有在两个活动都完成时才离开
        async{ seq1 = readASequence(path1); } // 派生活动
        seq2 = readASequence(path2); // 并行运行
    }
    compare(seq1, seq2); // 此处两个序列都可用
}
catch(e:Exception) { handleTheProblem(e); }
```

图 2. 并行读取两个流所需的 C 代码。

```
typedef struct {
    size_t size; /* 数据数组的字节数 */
    int rc; /* i/o 返回码，通常为 0 */
    char buffer[]; /* 字符串或错误消息 */
} ByteArray;

static void do_the_read(void *arg) {
    char *path = (char *)arg;
    pthread_exit(read_sequence(path));
}

...

int rc;
if (my_id == 0) { /* 我是 MPI rank 为 0 的处理器吗？ */
    pthread_t thread; /* 是的，那么我来主管 */
    ByteArray *seq1, *seq2;
    rc = pthread_create(&thread, NULL, do_the_read, path1);
    if (rc != 0) perror("Thread creation failed");
    else {
        seq2 = read_sequence(path2);
        rc = pthread_join(thread, (void *)&seq1);
        if (rc || seq1->rc || seq2->rc) {
            rc = report_the_problem(rc, seq1, seq2);
        }
    }
}
MPI_Bcast(&rc, 1, MPI_INT, 0, MPI_COMM_WORLD);
if (rc == 0) return compare(seq1, seq2);
else return rc;
```

程在 C 中实现的同样的代码。

代码中的 `my_id==0` (典型的 C/MPI 用法) 这样的测试语句表现出了 SPMD 模型在描述能力上的相对不足。如果发生了 I/O 错误, 处理器 0 必须让其他处理器知道出现了问题。这也是为什么在读入字符串后, 无论是否出现问题, 都必须广播的原因。函数 `do_the_read` 封装了第二个线程的任务。即使它执行与根线程完全一样的读取, 也必须单独进行。

远程线程。上面的例子非常简单, 因此很容易低估其重要性。为了更好地理解, 请考虑稍微不同的变化。假设, 我们并不知道处理器 0 应该负责处理大部分工作, 而是在运行时才知道数据位于哪个处理器上。如果数据是来自网络或者来自一组数据库, 那么很可能出现这种情况。代码应该如何改变? 图 3 显示了 X10 代码。

X10 程序中的位置 (place) 跟 MPI 进程一样, 都有一个唯一整数 ID。在图 3 中, 两个序列是从 ID 为 `seq1Id` 和 `seq2Id` 的两个位置读取的。这里的读取操作与原来的读取操作的唯一不同之处是, 必须提供 `at` 子句来表明每个读取发生在哪里。其余的所有代码保持不变, 包括错误处理, 即使现在可能涉及多个处理器。这是 X10 异常处理模型的结果, 这种异常处理旨在反映程序的活动树。它允许该树中的任

何活动处理该活动以下的子树中任何部分抛出的异常。(虽然程序员可以随时建立 X10 异常的子类, 并插入位置信息, 但是 X10 异常并不自动包含抛出它的位置。此外, 异常对发生该异常的活动的子级或平级没有影响。)虽然在 2002 年以前, C++ 就针对串行代码引入了 `try/catch` 样式的异常处理, 但是没有一个是线程框架将其延续到跟 X10 一样简单的多进程集中, 在 X10 中, 单线程和多线程的情况无需改动即可得到支持。

存取序列化。单服务器队列是一种常见的模式, 其中由一组执行者串行写入的项目流由可能不同的另一组执行者读取。通常的解决方案是在一个数组中缓冲这些项目, 直至其被消耗。项目添加到数组末端, 但是从数组开头处移出。

当有多个执行者时, 这种情况下出现争用的可能性显而易见。从数组移出 (Remove) 的操作必须序列化成一个接一个前后相继, 添加 (add) 操作也必须如此。另一方面, 除非数组快空了, 否则 `add` 和 `remove` 无需序列化前后相继, 因为它们影响数组不同的两端。

如果允许 `add` 与 `remove` 并行运行是安全的, 那么如何才能“挤出”最后一点性能呢? 序列化两个 `add` 或两个 `remove` 只需要轻量级原子版本的 C `postfix ++`。如果数组快空了, 则所有操作都必须序列化,

这需要原子性的代码块来保证线程一旦在该块中激活, 就成为该块中的唯一线程, 直至其退出该块。

在 X10 中, `AtomicInteger` 类提供对单个整数值变量的原子更新。由于这使用了硬件支持, 因此该操作比保护整个块更加轻量。语句块前面的关键字 `atomic` 序列化对块的存取。因此 X10 提供了编写 650 行可相当自然地执行读取操作的代码所需的条件。到 2002 年之前, 还没有与 `Atomic-Integer` 或原子块相当的标准化 C API。(请参阅 <http://gcc.gnu.org/onlinedocs/gcc-4.1.1/gcc/Atomic-Builtins.html>。此 API 实际上就是 C11 中标准化的内容。)原子块和原子更新所需要的序列化可使用 `MPI_Recv` 等同步 MPI 调用来实现, 但是最终得到的代码更长 (在我们的实现中有 1,100 行), 而且更难看懂。

对那些想知道为什么需要 1,100 行代码的人来说, 其实只有一小部分代码是用来支持队列操作的。这是客户端-服务器应用程序。服务器必须启动、必须停止, 客户端必须发出请求, 必须有访问控制 (不是每个人都可以使用每个数组) 等等。而使得 X10 版本更加简短的原因是错误处理更加简单, 而且不需要编写客户端和服务 API。

对等协调。在前面的示例中, 协调是通过中心服务器进程管理的。如果一组对等进程各自管理自己的队列, 并在需求上升时, 与其他进程分担项目, 那么这会发生什么情况呢? 这时, 对部分活动进程来说, 每个进程既充当客户端, 又充当服务器。我们编写的示例程序执行树的广度优先搜索, 并且事先已知树相当不平衡。每个处理器管理自己的未访问节点队列, 但是, 为了更平均地分散工作负载, 可能要求处理器将某些未处理的节点交给其他处理器处理。

图 3. 使用只在运行时才知道的处理器进行读取的 X10 代码。

```
try {
    val seq1:Array[Byte](1); // seq1 和 seq2 都是字节数组
    val seq2:Array[Byte](1); // 按一个整数索引
    val seq1Home = Place.place(seq1Id); // id 指定 Place 对象
    val seq2Home = Place.place(seq2Id); // 具体化处理器
    finish {
        async { seq1 = at(seq1Home) readASequence(path1); }
        seq2 = at(seq2Home) readASequence(path2);
    }
    compare(seq1, seq2);
}
catch(e:Exception) { handleTheProblem(e); }
```

这里最大的问题是终止：一个进程如何知道什么时候不仅自己做完了工作，而且所有其他进程也做完了工作？处理此问题的算法（至少在较低数量级的规模）可以追溯到 30 多年前的 Dijkstra 等人⁴的算法，并且该算法从那时起经过了多次精化。SPMD 解决方案就是基于这些算法。X10 解决方案完全不一样，它依靠按需线程。

我们先来看看 X10。一个 X10 活动开始操作时，会要求活动集中的每个对等方开始搜索树的一部分。如果其中一个对等方没有工作，它会获得一个“邻近”对等方的列表，它获准向这些对等方要求更多工作。为了做到这点，它将对自身的引用发送给每一个相邻对等方，然后终止自己的当前活动。但是，发出这些请求的对象仍然存在。因此，如果空闲对等方的其中一个邻居有多余的工作，它可以在空闲对等方的位置上派生一个新活动，在那里空闲对等方可以继续工作。如果空闲对等方要确保只接受了一个邻居的工作，则还要涉及一点握手，否则整个过程结束。由于根活动在一个 finish 块中作为 asyncs 派生了多个对等方活动，因此它只能在所有这些 asyncs 都不再处于活动状态时，才能退出该块，从而完成作业。

C 解决方案完全不同。每个处理器上没有对等对象准备接收远程过程调用，即使有，也没有跟 X10 finish 块相似的东西。处理空闲对等方的朴素方法是让其在同步接收时挂起，但是谁来负责知道何时向该对等方发出“全都完成”消息？这正是 Dijkstra 等人通过轮询所有对等方的聪明方法来解决的问题。

在 C 解决方案中，每个处理器必须执行三个活动：主要工作（即，搜索和处理树中属于它的那一部分）；侦听终止控制消息；以及侦

C 缺乏内省也是一个严重问题。在 X10 之类面向对象的语言中，对象知道自己有多大，存在哪些活跃引用。

听其邻居要求分担工作的请求。因为使用两个非阻塞接收来侦听控制和共享通信，因此每个处理器实际上有三个并发活动：主要活动和两个挂起的接收。每个处理器的主循环负责及时处理传入的控制和共享请求。主要复杂之处在于确定谁向谁发出什么控制信息。

管理内存和处理错误。前面讨论的问题属于我们程序的自然线程模型与 X10 APGAS 和 MPI SPMD 并行模型之间的契合。在编写代码的串行部分时，还有几个众所周知的与 C 相关的缺点影响我们的效率。X10 中不存在这些缺点。平均来说，每 1,000 行 C 代码就会遇到约六个或更多这些有名的问题，总体影响非常显著。很多问题需要额外的处理，因为它们出现的地方并不在需要纠正的地方附近。

内存泄露。X10 与很多最新语言一样，也有自动垃圾回收功能。显式管理存储这一需求长期以来一直被视为妨碍 C 编码效率的严重障碍之一。您可能认为，在大约 1,200 行相对直观的字符串匹配代码这样简单的问题中（例如 SSCA 1），内存泄露不是问题——但这确实是问题，尤其是如果这段代码要长时间运行（或要并入频繁调用的库例程中）的话。这种情况下需要查找并消除内存泄露，这样的任务很容易需要一到两天才能在我们的代码中找到分配了内存但却从不释放的位置。

使内存引用正确。C 缺乏内省也是一个严重问题。在 X10 之类面向对象的语言中，对象知道自己有多大，存在哪些活跃引用。在 C 中，这些计算必须手动完成。这不是说，使用 X10 就不会误算数组范围。误算的可能性仍然存在，但是 X10 会在运行时，在发生数组越界存取的那一刻，捕获越界存取。即使老练的 C 程序员也不大可能花时间让每一次数组存取都万无一失。由于检测到有错存取的地方通常远离需要

纠正代码的地方，所以这些错误不易查找和修复。

错误处理。C 缺乏便利的异常机制迫使程序员写得更啰嗦。这一问题在 Floyd 算法中浮现出来，例如，在我们的程序员需要泛型输入流来读取数值的 ASCII 流的时候。Floyd 的 API 有一个入口，这个入口对流进行分词处理，将分出的词转换为相应的数值类型，并确保值合法。很明显，流可能遇到很多问题。问题是，如何处理这些错误。

对于 X10 中的错误，X10 会抛出异常，异常的类型可确定遇到的问题。应用程序可避免提供检测一般错误（如低级函数发现的意外文件结尾）的特殊代码，因为低级函数也可通过抛出异常来通报有错误。因此，应用程序可专心通报内容中的语义错误。

对大多数临时性代码来说，错误处理不算是严重问题，但对于生产代码，对于复杂的通信模式（如 Dijkstra 等人的终止算法）而言，这肯定是。C 的通报机制最适合专家使用。但是 C 的问题在多线程 SPMD 领域藏得更深。请考虑流为了将找到的分词转换成整数而调用的标准库例程 `strtoll`。下面是对 `strtoll` 的错误表示的一段讨论，摘自 `strtoll` 的“手册”页：

“`strtoll`、`strtoll`... 函数返回转换的结果，除非结果值下溢或上溢。如果无法执行转换，则返回 0，且全局变量 `errno` 设置为 `EINVAL`（最后一个功能并非在所有平台上都可移植）。如果发生了上溢或下溢，则 `errno` 设置为 `ERANGE`...”

请考虑 C 应用程序为了处理各种可能的错误而需要的代码。代码在调用 `strtoll` 之前，应确保将 `errno` 清零吗？毕竟，`errno` 可能因为以前完全无关的问题而变为非零。此外，对于检查 `errno` 来了解是否发生错误的代码，光检查 `errno` 是否非零还不够，因为错误

可能是 I/O 错误、解析错误或范围错误。而且您也不能确保 `errno` 是线程安全的——它并非在所有系统上都线程安全。那么该怎么办呢？应该在应用程序中的什么地方清除 `errno`（它的值可是全局的）？需要让其他哪些进程意识到问题，如何让它们意识到问题？

总结

C 和 MPI 在并行编程社区大行其道自有其很合理的道理。它们的设计优雅、干净、有出色的文档，而且已经仔细实现。在经验丰富、训练有素的专业高手手中，它们可提供在其他地方完全得不到的控制级别。C 和 MPI 都不是站着不动的目标：尽管它们现在已经是成熟的技术，但是两者仍在不断改进。

但是，C/MPI 利大于弊的时候多吗？通过我们的全部三次研究，尤其是这最后一次研究，我们看到，使用更高级的语言和编程模型可以获得显著的利益——第一次成功并行运行的开发时间快了 2 到 6 倍。对于需要保留数年的大型程序来说，这样的效率提升可能更大。

现在很多研究和院校机构已开始对 X10 和 APGAS 展开探究。虽然这门语言不一定进入主流应用，但是令其如此高效的品质很可能在并行社区得到长期认可：灵活的线程处理、自动垃圾回收、运行时类型决定的错误检查、分区全局内存以及根异常处理全都非常有用。希望我们的实验能促进那些希望提高并行程序员效率的人认证研究 X10 的设计及其优势。

致谢

本文得到美国国防部高级研究计划局的大力支持，协议编号 HR0011-07-9-0002。作者感谢 Catalina Danis、Peter Malkin 和 John Thomas，以及 IBM 的 HPCS 效率评估团队的成员，感谢他们为本项

研究所作的宝贵贡献。

queue.acm.org 上的相关文章

Unlocking Concurrency

Ali-Reza Adl-Tabatabai, Christos Kozyrakis

and Bratin Saha

<http://queue.acm.org/detail.cfm?id=1189288>

The Ideal HPC Programming Language

Eugene Loh

<http://queue.acm.org/detail.cfm?id=1820518>

Software Transactional Memory:

Why Is It Only a Research Toy?

Calin Cascaval, Colin Blundell, Maged Michael,

Harold W. Cain, Peng Wu, Stefanie Chiras, and

Siddhartha Chatterjee

<http://queue.acm.org/detail.cfm?id=1454466>

参考资料

- Asanovic, K., Bodik, R., Catanzaro, B.C., Gebis, J., Husbands, P., Keutzer, K., Patterson, D., Plishker, W., Shalf, J., Williams, S. and Yelick, K. The landscape of parallel computing research: a view from Berkeley. Technical Report No. UCB/EECS-2006-183. Electrical Engineering and Computer Sciences, University of California at Berkeley, 2006. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.pdf>.
- Bader, D., Madduri, K., Gilbert, J., Shah, V., Kepner, J., Meuse, T. and Krishnamurthy, A. Designing scalable synthetic compact applications for benchmarking high productivity computing systems; <http://www.cse.psu.edu/~madduri/papers/SSCA-CTWatch06.pdf>.
- Danis, C. and Halverson, C. The value derived from the observational component in an integrated methodology for the study of HPC programmer productivity. In *Proceedings of the Third Workshop on Productivity and Performance in High-End Computing*, (2006), 11–21.
- Dijkstra, E., Feijen, W. and van Gasteren, A. Derivation of a termination detection algorithm for distributed computations. *Information Processing Letters* 16, 5 (1983), 217–219.
- Ebcioğlu, K., Sarkar, V., El-Ghazawi, T. and Urbanic, J. An experiment in measuring the productivity of three parallel programming languages. In *Proceedings of the Third Workshop on Productivity and Performance in High-End Computing*, (2006), 30–36.
- Halverson, C. and Danis, C. Towards an ecologically valid study of programmer behavior for scientific computing. In *Proceedings of the First Workshop on Software Engineering for Computational Science and Engineering*, (2008).
- Saraswat, V.A., Kambadur, P., Kodali, S., Grove, D. and Krishnamoorthy, S. Lifeline-based global load balancing. In *Proceedings of the 16th ACM Symposium on Principles and Practices of Parallel Programming*, (2011), 201–212.

John Richards 是 IBM Watson Group 的研究管理者，并担任苏格兰邓迪大学计算机学院的荣誉教授。

Jonathan Brezin 是一名数学家，在上世纪 60 和 70 年代，曾在明尼苏达大学和北卡罗来纳大学任职。之后他加入了 IBM Research，并于今年从那里退休。

Cal Swart 于 1982 年加入 IBM Research，目前是 IBM Watson Group 的高级技术研究人员。他是 Watson Life Research 的成员，研究认知计算的新应用。

Christine Halverson 是硅谷的独立咨询师。以前她曾在 IBM Research 工作，在那里她当了 5 年 DARPA HPCS 计划研究并行程序员。

译文责任编辑：陈文光

用场景的语义和几何结构标注像素，
让计算机述其所见

STEPHEN GOULD, XUMING HE

通过标记像素 实现场景理解

长期以来，通过计算机编程让计算机自动解释图片的内容一直都是人工智能和计算机视觉领域的挑战。在早期计算机视觉研究领域中有的一则广为人知的故事，其中就出现了这一难题。当时，MIT（麻省理工学院）的一名本科生被要求花一个暑假的时间让计算机来描述它从摄像机拍摄的图片中“看到”的物体。³⁵ 约 50 年之后，研究人员仍然在努力解决同一问题。

一个场景可以用多种方法描述，会包括对象的细节、区域、几何结构、位置、活动、甚至一些不可见的属性（比如日期和时间）。例如，典型的城区场景（见图 1）可以通过确定前景的汽车对象，和背景的草地，天空和道路区域的位置来描述。或者，也可把图像总结为街景。我们希望计算机能够推理出场景的所有这些方面，并同时

提供粗略的图像级标签和详细的像素级标注，用于描述场景的语义和几何结构。早期的计算机视觉系统试图通过单一的统一模型来一起描述场景的所有方面，进而实现这一点。然而，问题的难度使得研究人员很快放弃了这种统一的方法。而且，直到最近，场景理解的研究仍在沿着不同的技术路线展开。

在其中的一个路线上，研究人员试图使用少量的标签提供场景的高层概要或分类（比如城市和森林），而不明确地识别其中的对象。另一个路线称为“对象检测”，旨在通过在对象周围放置一个边界框来定位场景中的离散对象（比如汽车或行人）。现在数码相机和智能手机中使用的人脸检测算法执行了这一任务。然而，这些方法并没有给出对象的详细轮廓，也不能推断出图像的整体内容。

或许另一种技术路线最接近场景理解这一长期目标，它旨在为整张图片生成像素级的标注。用于场景理解的此类像素标记或语义分割方法是本文关注的重点。对象和背景类别被分割为离散的不重合区

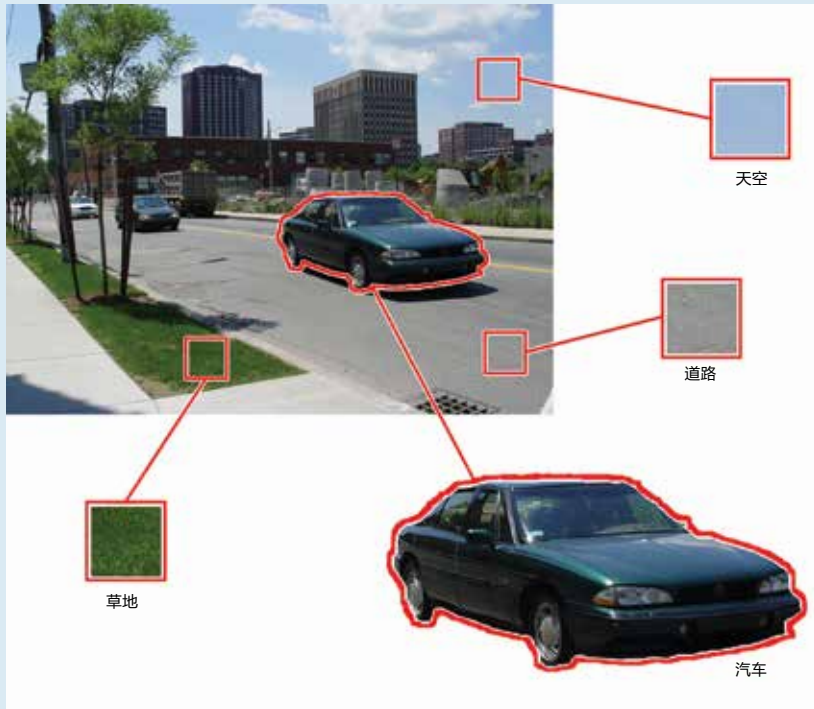
» 重要见解

- 图像理解一直以来都是人工智能领域的挑战之一。最近这方面的进步催生了机器人感知、监控与环境监测、基于内容的图片搜索和社交媒体摘要中的诸多应用。
- 图像理解的方法之一是使用类别标记标注图像中的每一个像素。这一过程使用了名为 CRF 的概率模型，它能够处理不确定性，并能把上下文信息传播到整个图片。
- 随着机器学习技术的日益进步，机器性能的逐步提高以及数据数量的不断增长，我们越来越接近机器能像人一样看见和理解世界的目标。



图 1. 从前景对象（“物体”）和背景区域（“素材”）方面理解典型的城区场景。

背景区域可以较容易的通过局部颜色和纹理线索来识别；但是，识别汽车等前景对象需要复杂的形状和外观模型。



域，然后为每个区域（等同于区域内的所有像素）设置一个标记。除了为每个像素标上类别标记外，还可以用唯一的标识符标注每个对象的不同实例。例如，这样做之后，两辆临近的轿车可以被当成不粘连的对象。这种多类别 / 多实例的方法代表了当代场景理解研究的最前沿。层级分割，即所谓的“场景解析”³⁷，可通过把对象分成部件来生成更为精细的场景视图；比如，汽车可以拆解成车轮、车身面板和车窗。

这些像素标记方法使用了预先定义类别标记集合，其中规定了模型可识别的对象类别和场景类型。标记可以是语义方面的（比如草地、道路、天空和汽车），或是几何方面的（比如水平的、垂直的和倾斜的），并可针对不同的场景类型进行微调；例如，与室外的场景相比，室内的场景描述需要截然

不同的语义和几何结构的类别。扩大可识别的类别的数量并提高其多样性是当代研究的核心推动力。

像素标记的大多数方法使用了名为条件马尔可夫随机场（CRF）的概率模型，它提供了一个形式框架用于对场景的视觉外观和底层语义（或几何）标记之间的复杂关系进行编码。不仅如此，这一形式体系还允许纳入有效的推理算法，并支持从数据中学习模型参数。在一些最近的研究中（比如 Heitz 等人，¹⁵ Hoiem，¹⁸ Li 等人，²⁸ 和 Yao 等人⁴⁰），研究人员已经寻求将场景理解的不同技术路线再次整合成单一的一致模型，其中纳入了高层级的场景标记、对象分割和几何推理。例如，这种统一的视角允许通过关联场景的语义和几何方面自然地表现约束（例如对象支持和尺度）。这些当前最高水平的的整体模型基

于像素标记方法进行构建，并使用 CRF 框架整合场景的各个方面。

在此，我们概述了场景理解使用的 CRF 模型的各种变体，说明了它们如何利用与现实世界场景有关的各种假设（比如汽车通常在道路上出现）。然而需要注意的是，尽管会不时出现一些革新，但场景理解仍然是一个远未解决的开放问题。我们还提供了基准 CRF 模型处理标准场景理解数据集时的结果样例，用以展示现有场景理解模型的能力和缺陷。

像素标记

进行像素标记时，会从预先定义类别集合（比如草地、树木、道路、汽车和人类）中选出一个类别标记来标记图像中的每个像素。此处假设，每个像素属于单一的感兴趣类别，而且该类别可以被无歧义地识别出来。

方法之一是单独地分类每个像素，无视图像中其他像素的类别。然而，正如本文所述，单独处理每个像素可能会产生非常不一致的结果。当前最高水平的场景理解算法采用了更加复杂的方法，其通过在多像素上定义随机场统一考虑了它们的类别标记。^{13,34} 最近几年，由于使用了高效的推理和学习算法，此类方法获得了很大的成功。^{3,22,31}

条件马尔可夫随机场。 CRF 首先在自然语言处理领域引入，但随后便应用于很多机器学习任务中。²⁵ 它们主要的好处是提供了一种原理性的概率框架，把相关输出变量（比如图像中像素的标记）之间的关系描述为观测特征（比如像素颜色）的函数。因此，在整合多视觉线索，融合相关场景理解问题时，它们是理想的选择。不仅如此，CRF 给出了一种紧凑的表示，并提供了高效地（近似）推理和学习算法。

在形式上，设 $\mathbf{y} = (y_1, \dots, y_n)$ 为离散随机变量（输出变量）组成的向量。我们的关注点是基于某些观测特征 \mathbf{x} 预测出这些变量。我们假设每个变量会从预先定义的有限标记集合 $y_i \in L$ 中选取一个标记。在进行像素标记时，变量 y_i 指为图像中的第 i 个像素指定的标记，而 n 指图像中的总像素数。通常情况下，特征 \mathbf{x} 通过与单独像素或图像区域关联的实值向量表示。例如， \mathbf{x} 可以编码颜色或纹理线索，后面我们会进行讨论。

通过能量函数 $E(\mathbf{y}; \mathbf{x})$ ，CRF 模型通过如下公式定义了给定观测特征时输出变量的概率分布：

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\{-E(\mathbf{y}; \mathbf{x})\} \quad 1.$$

其中 $Z(\mathbf{x})$ 是所谓的配分函数，它保证了概率分布是规范化的(和为1)。注意，由于配分函数的计算涉及对 \mathbf{y} 的赋值进行求和，而 \mathbf{y} 的赋值数是指数级别的，所以这种计算一般来说是无法进行的。幸运的是，如文中所述，推断最可能标记时，可以不必计算配分函数。

通过把能量函数 $E(\mathbf{y}; \mathbf{x})$ 分解为更小的团势能 (clique potentials) 之和，可以紧凑地表示 CRF。

$$E(\mathbf{y}; \mathbf{x}) = \sum_c \psi_c(\mathbf{y}_c; \mathbf{x}). \quad 2.$$

式中，每个团势能 $\psi_c(\mathbf{y}_c; \mathbf{x})$ 为在随机变量子集上定义的实值函数。我们使用简写 \mathbf{y}_c 指团中变量的子集(或“团势能”的定义域)。^a粗略的说，对于团势能范围内各变量的每次联合赋值，团势能包含了这些赋值的数值得分(忽略模型中的所有其他变量)。团势能为其定义域内各变量的每种联合赋值编码了一个数值得分。概率影响会依靠共享变量的或有重合定义域的团势能在模型中扩散。

a 在形式上， c 是幂集 $\{1, \dots, n\}$ 中元素的稀疏子集的索引。

团势能内变量的数量定义了模型的阶数。高阶模型的每个团中可能包含极其多的变量。然而，如果缺少合适的结构，这些模型会导致难以驾驭的推理问题。常用的像素标记模型只包含一元项和成对的项。

$$E(\mathbf{y}; \mathbf{x}) = \sum_{i=1}^n \psi_i^U(y_i; \mathbf{x}) + \sum_{ij \in \mathcal{E}} \psi_{ij}^P(y_i, y_j; \mathbf{x}). \quad 3.$$

上式中的一元项与图像的每个像素相关联，而成对项则定义在像素对集合 \mathcal{E} 之上。该集合通常是稀疏的，只包含图像中位置临近的像素对。图 2 是一个定义在四连通邻域上的网格结构 CRF 示例。在此类模型中，每个变量的标记不仅会受到局部特征的影响，同时也会受到周围变量的标记的影响。这种影响可分别通过一元项和成对项获得。详细来说，成对项允许对平滑性假设进行编码；换言之，照片中的某个像素很可能与其相邻像素属于同一对象。最近的一些研究考虑了全连通图，

即使用所有的像素对之间的成对项。这样的成对项支持系统捕捉大范围像素之间的互动。但是，只有在成对项具有涉及高斯核的某种特定形式时，这种问题才容易处理。²³

研究人员通常只对可能性最大的场景解释感兴趣；从概率的角度讲，这被称为最大后验概率 (MAP) 推理，即需要求解：

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmax}} P(\mathbf{y}; \mathbf{x}). \quad 4.$$

因为配分函数 $Z(\mathbf{x})$ 不会因 \mathbf{y} 的赋值不同而发生变化，可以通过求解等效能量最小化问题得出可能性最大的场景解释

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmin}} E(\mathbf{y}; \mathbf{x}). \quad 5.$$

针对像素标记问题建立的 CRF，研究人员已经开发了快速算法来获取近似的能量最小化。^{3,21,22} 一种最为普遍的方法称为“图割”，^{3,21} 它是一种移动判定 (move-making) 算

图 2. 九个随机变量 $\{y_1, \dots, y_9\}$ 上的网格结构化成对 CRF 的图形表示，其中每个变量通过一个节点表示，变量之间的直接成对关联通过边表示，观测特征 $\{x_1, \dots, x_9\}$ 则通过阴影节点表示。

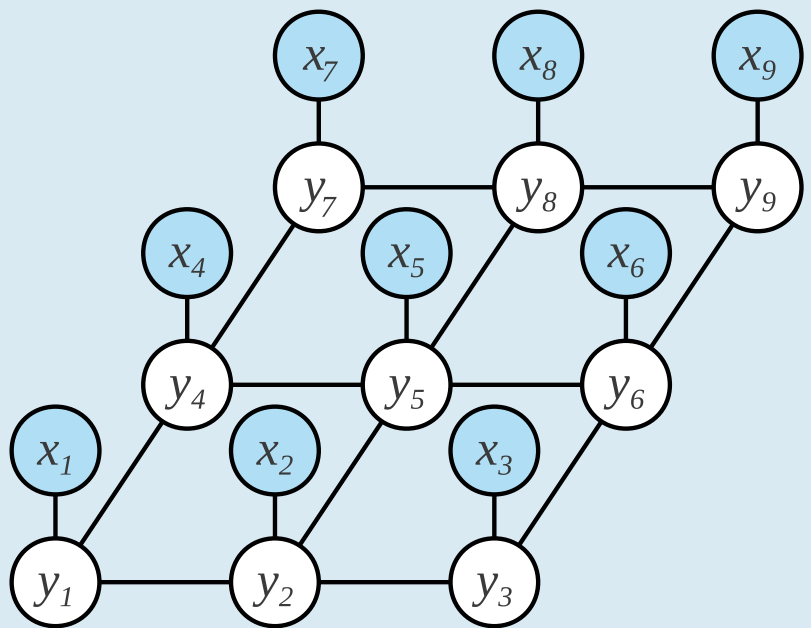
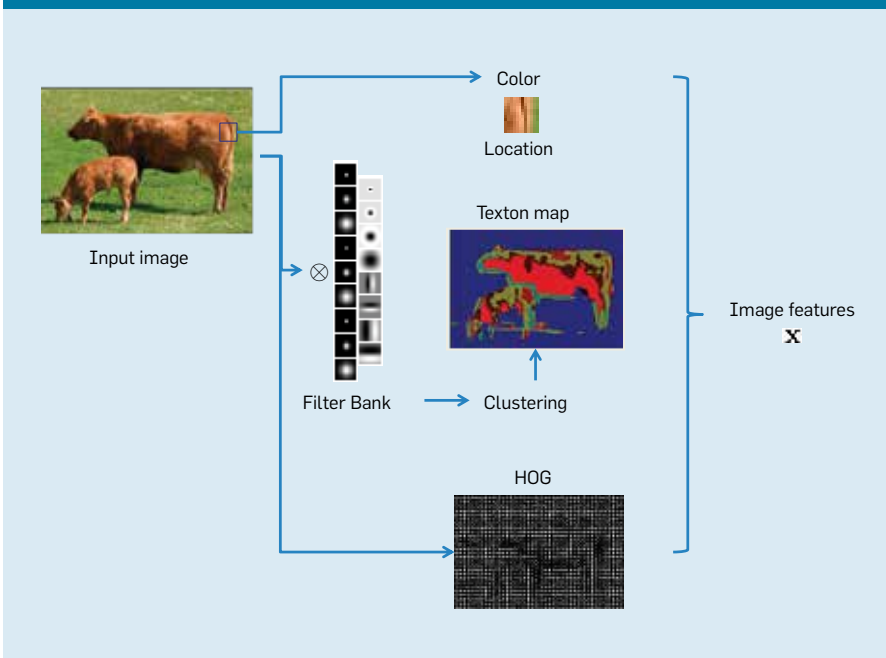


图 3. 在 CRF 模型的势能中使用的典型图像特征，包括颜色、图像位置、滤波器输出和 HOG；滤波器输出可用于生成“纹理基元特征”，或纹理原型。



法，首先给各个变量赋初值，然后迭代地求解一系列二值优化问题，逐步改进当前的解。

团势能 $\psi_c(\mathbf{y}_c; \mathbf{x})$ 可用多种方式确定，但其通常会包含可以控制势能值的参数，并把势能值作为各变量和观测特征的函数；例如，势能可定义为特征值的线性组合，并通过权向量进行参数化。或者，势能也可被定义为决策树分类器的输出，通过节点的分裂和叶子的概率进行参数化。设计场景理解的 CRF 模型时，主要的难点在于学习这些参数。由于无法计算配分函数，所以不能应用最大似然法。研究人员已经尝试了多种近似学习技术。其中比较流行的有“使用交叉验证的分段学习”³⁴和“伪似然学习”。¹² 最近几年，最大间隔学习方法已经获得了成功。³¹ 不过，最有效的 CRF 参数学习策略仍然是一个悬而未决的研究问题。

特征。 CRF 模型的优点之一是，可以在团势能中使用多种图像特征，并把它们当成观测值或每张图像的固定特征处理。把图像特征当成观测值，不仅为场景表示带来

了灵活性，还简化了模型结构，因为模型不需要显式地捕捉这些特征的概率分布。针对感兴趣的对象的视觉属性，研究人员会对不同类型的图像特征进行裁剪。

计算机视觉的研究人员充满感情地把对象类型归为两个大类：“物体”和“素材”。从局部外观中识别背景类别（素材）较容易；但识别对象（物体）却需要更多的线索（比如形状）。图 1 反映了这种差别，其中使用了三个背景类别的例子——草地、道路和天空——以及一个前景的对象——汽车。

局部的外观特征为颜色和纹理线索。两个常用的颜色空间为 RGB 和 CIE-Lab，因为 RGB 简单，而 CIE-Lab 的颜色较接近人类的感知。³⁵ 对于与区域相关联的图像特征，通常采用颜色直方图或汇总统计量来描述它们的总体外观。纹理线索旨在捕捉图像中重复的局部模式，通常基于滤波器组及其输出的分布抽取。一个广泛使用的例子是纹理基元（即小的纹理原型）特征^{30,34}，其首先把滤波器组的输出聚类为纹理基元码本，然后用纹理基元码字

的直方图表示图像块；图 3 列举出了几种常用的图像特征。

然而，与背景类别相比，前景对象的局部外观特征的变化要大得多。实现它们的可靠识别，需要更多鲜明的、稳定的特征。这些特征需要超越孤立像素这一局限，对结构信息进行编码；例如，对象形状的信息可通过方向梯度（HOG）直方图特征表示。⁵ 另外，后面我们会讨论通过自顶向下的处理取得其他高层特征。它们也可用于表示对象的属性。不仅如此，全局的外观特征（比如图像的颜色均值）也可作为特定的对象提供上下文；例如，绵羊倾向于在浅绿色的图像中出现，而汽车则倾向于浅灰色的图像中出现。

图像特征可被联接起来组成每个像素的特征向量 \mathbf{x} ，然后纳入 CRF 模型的势能函数。具体来说，在把图像线索映射到标记时，发挥了关键作用的一元势能可被定义如下：

$$\psi_i^c(y_i = k; \mathbf{x}) = \mathbf{w}_k^T \phi_i(\mathbf{x}) \quad 6.$$

其中 \mathbf{w}_k 为标记类别 k 的权重系数， $\phi_i(\mathbf{x})$ 为像素 i 的特征的（非）线性映射。权重 \mathbf{w}_k 通过使用结构化预测学习算法等处理训练数据来确定。³¹

上下文的作用。 在人类自身的对象识别和场景理解中，上下文发挥了重要的作用。以图 4 中的城市场景为例；我们马上识别出了街道及其中的每辆汽车，虽然汽车的高度不到 32 个像素，这多少有点让人惊奇。做出这种识别的原因是，虽然汽车的局部证据较弱，但根据场景的空间布局得出的强上下文线索对此进行了补偿。

上下文信息可以多种方式纳入场景理解模型。最简单的上下文形式为特征和类别标记的统计相关性；例如，与草地或树木相比，蓝

色的像素更像天空或水。与此类似，与天空或水相比，绿色的像素更像草地或树木。这也正是通过 CRF 模型中的一元势能所获取的上下文类型。不过，我们也能够对更为复杂的上下文假设进行编码。

对象类型的同现（比如汽车通常与道路同时出现）也可被纳入场景理解模型中。^{32,40} 此时，两个对象的相对位置并不关键；唯一重要的是它们通常一起出现。通过引入标明每个对象类别存在与否的二元变量 z_k ，我们可以把此类上下文纳入 CRF。把这些对象变量 z_k 与像素变量 y_i 联系起来的成对势能强制保持这种一致性；换言之，如果某个对象存在，那么某些像素必须被标记为该对象，反之亦然。然后，两个对象变量 z_k 和 z_l 之间的成对势能便能捕捉同现的倾向。不过，请注意，由此得出的 CRF 图不再是规则的像素网格。更重要的是，这种情况下高效的图割技术无法用于推理，所以各方法必须求助于较慢的推理算法；具体的例子参见 Yao 等人的论文。⁴⁰

类别同现是对上下文的粗略衡量，无法捕捉对象之间的相对位置。所以，这会导致对象之间相对位置不一致的错误检测（比如浮在树林中的汽车）。对象的空间位置与场景中的相邻背景¹⁶ 或对象的相对位置关联¹² 后，可通过设计势能函数把这种关联信息纳入模型，自动纠正此类错误。不幸的是，建模对象的相对位置需要直接利用非局部推理，这会推理变得非常棘手。全局实体（比如室外场景中的地平线位置¹⁹ 或室内场景中的箱形结构布局^{14,39}）可用于间接关联对象，简化推理。在本文中，我们在 CRF 模型中引入额外的变量（比如用于确定地平线位置的 v_{hz} ）模型（公式 2），并通过成对势能将它们与像素标记变量 y_i 关联起来。

像素与超像素的对比 因为图像在计算机中表示为像素的矩形阵

列，所以在场景理解时，把标记与单独的像素关联是一种自然选择。不仅如此，对于常规的图像结构，使用相邻像素的成对项构建 CRF 相当方便。不过，像素本身是成像过程的人工产物，并不一定反映场景的底层结构或复杂性；例如，用于场景理解时，一兆像素的照片与八兆像素的照片传递的信息基本相同，虽然后者的像素是前者的八倍，因而需要更大的模型。像素是类别的嘈杂指标；例如，在图 5 中，尽管两个相邻的像素属于同一个物体，它们的颜色却可能区别很大。

避免不必要的大量变量的潜在方法之一是，用超像素或拥有一致外观的小型连续区域为图像建模。此时，方程 3 中 y_i 值为第 i 个超像素分配的标记， n 为超像素的数量，详见 Fulkerson 等人的论文。⁹ 通过聚集颜色相近的相邻像素来形成众多的小块区域，可以生成图像的超像素表示。^{4,8,33} 由于超像素通常比感兴趣的对象要小得多，这种过程被称为“过分割”。传统的过分割方法通常相当慢，而且还会导致不符合对象边界的极不规则区域。最近的若干研究工作已经在尝试生成形状更为规则¹，更接近场景中真实边界的超像素。

很多算法都提供了控制超像素数量的手段。图 5 包含了把图像分割成超像素的三种不同过分割方

式。第一个（图 5b）生成的超像素太少，图像中的很多细节都丢失了。而最后一个（图 5d）则保留了图像中的所有基本结构，但把描述场景所需的实体数量减少了到 46 分之一，从 68,160 个像素降到了 1,476 个超像素。

除了降低模型的规模外，超像素还为特征计算提供了空间支持^{9,17}，让这些特征更不容易受到噪声的影响。然而，与像素相比，使用超像素描述图像也存在一些缺陷。首先，超像素不符合规则的相邻结构，这使得程序员很难衡量每个成对项在 CRF 模型中的影响。其次，超像素把区域边界作为预处理步骤，而这些边界可能与真实的对象边界不一致。

二者——像素和超像素——的最佳结合可以通过使用像素表示图像，但使用超像素提供其他的特征或强制施加高阶一致性限制实现。²⁰ 例如，模型可对超像素中的像素类别标记不一致的情况进行惩罚。这种惩罚可通过高阶 Potts 势能在 CRF 模型中实现。

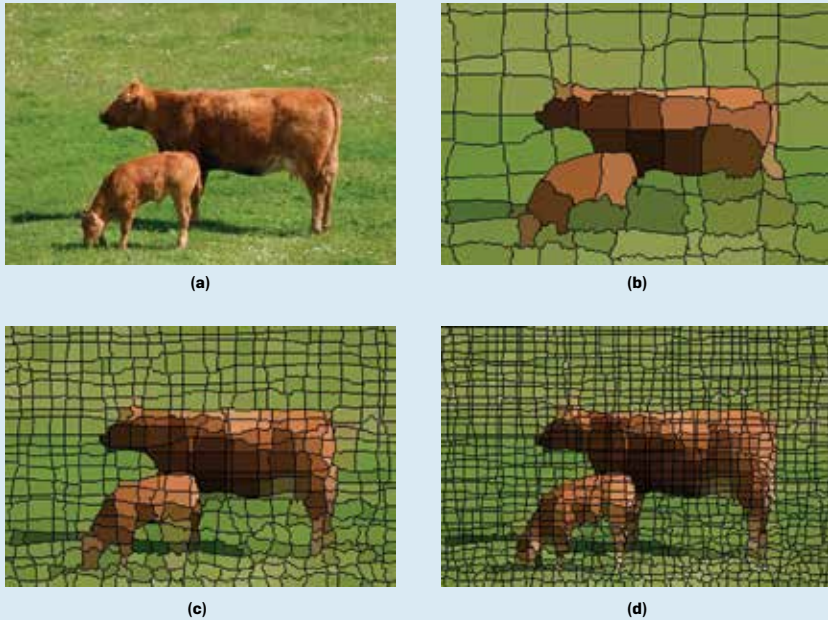
$$\psi_c^H(\mathbf{y}_c) = \begin{cases} 0 & \text{if } \exists \ell \in \mathcal{L} \text{ s.t. } y_i = \ell \text{ for all } i \in c \\ \lambda & \text{otherwise} \end{cases} \quad 7.$$

其中，如果团 c 中所有变量的标记不同，则需要付出 $\lambda > 0$ 的代价。采用这种方式后，算法会优先使用

图 4. 用于说明对象检测时上下文重要性的低分辨率图片示例；把选出的对象孤立出来时，其识别相当难；但在街景的上下文中，可以很容易地确定它为汽车。



图 5. 图像以及把该图像过分割成超像素的多种方法；超像素的数量控制了模型复杂度和表示准确度之间的平衡。



超像素定义的边界，但是，如果存在足够的证据支持相反的判断，这种假设可被推翻。

超像素还可在迭代推理程序（比如移动判定算法^{10,11}）或数据驱动的马尔可夫链蒙特卡罗算法内使用。³⁸ 此时，在当前场景解释过程中，超像素可以通过建议改变其中的对象边界来引导推理过程。

自顶向下与自底向上的对比 CRF 模型提供了统一的框架来整合图像线索和关于场景-标记配置的先验假设。用来定义一元团势能或形成超像素表示的图像线索通常被称为“自底向上的信息”。这种自底向上的线索描述了像素之间的相似性，可用于为标记任务生成信息丰富的建议（比如较大的同质区域）。我们把类别依赖的一元势能看成自底向上过程的一部分，虽然它们也包含了标记信息。关于标记的最常见先验假设定义在相邻（超）像素上（通过 CRF 中的成对项），代表了区域标记上的软平滑约束。虽然这种约束是在区域标记上施加的有效先验，但是它缺乏表达特定对象信

息的能力（比如全局形状和姿态）。我们把后面的这些属性上施加的先验称为“自顶向下的信息”，因为它们编码的知识层级超越了像素和简单的图像区域。

对象形状可能是最重要的对象级信息类型。形状先验通常用刚性的或可变形的掩膜（mask）表示，对于每个像素，该掩膜会设置其属于该掩膜所对应的前景对象的成本^{2,27}（见图 6）。可变形的掩膜可捕捉与对象类别相适的姿态变化（比如人和动物），而且通常通过基于部件的模型实现，其中掩膜的单独部分可相对另一部分移动。由于形状为前景区域的全局属性，形状先验自然会导致较大的团，因此也需要更复杂的推断算法。在实践中，一些掩膜表示分解为局部一元或成对项的汇总；例如，给定一个匹配的形状掩膜，通过修改处于对应的一元势能内的代价，位于掩膜内部的像素会得到激励接受特定的标记。

纳入对象先验的常用策略之一是使用对象检测生成对象实例建议

（比如边界框），并定义这些区域上的对象形状势能。^{11,24} 由此得出的能量函数通常包含高阶项，如方程 2 中的团 c 定义为对象支撑域内的像素，而该项偏爱对这些像素进行一致的标记。

高阶能量函数可通过广义的图割推理算法²⁰ 优化，或通过添加辅助变量转换为成对模型。或者，自底向上和自顶向下信息均可在移动判定算法的设计中使用，其迭代地对变量赋值提出可让能量减少的变化；例如，对象检测的输出可以生成强大的自顶向下移动，克服使用纯自底向上驱动的分割所产生的碎片化对象标记结果。¹¹

如前文所述，通过融合场景分割与多场景分析任务（比如场景分类、对象检测和深度 / 布局估计^{15,18,40}），某些最近的方法使用整体视角来实现场景理解。在此类框架中，几个层级的自顶向下信息（包括几何结构、候选对象的边界框以及场景分类）可被纳入一个单一的 CRF 能量函数。具体来说，这类模型引入了变量来表示这些量，而它们与像素标记的关系则通过对或高阶势能函数建模。这样的话，总体问题可被分解为一些更小的任务，轮流进行求解，或被当成统一的目标联合优化。

数据集和软件

场景理解算法的很多软件实现均可免费获取。其中有两个有名的例子，一个是 Darwin 软件框架^b，另一个是自动标记环境（ALE）^c。它们提供了基于前文所述的像素级 CRF 模型实现场景理解的基础框架。最基础的模型由一元和成对势能组成。一元势能通过使用局部和全局外观特征构建。成对的团势能通过与各像素四连通或八连通的相邻节点定义，且通常包括一个对对比度敏感

b <http://drwn.anu.edu.au>

c <http://cms.brookes.ac.uk/staff/PhilipTorr/ale.htm>

的平滑先验³⁴，以便防止颜色相近的相邻像素被赋予不同的标记。

不仅如此，很多良好标记的数据集是现成的，很多研究人员使用它们来开发和比较场景理解算法。为了展示上述软件可以达到的效果，下面我们给出使用 Darwin 软件框架处理两个标准数据集时的结果：

斯坦福背景数据集 (SBD)¹⁰ 由 715 张描绘乡村、城市和港口场景的照片组成。图像用两种不同的标记集进行标记：第一种捕捉语义类别，包括七个背景类别（天空、树木、道路、草地、水、建筑物和山）以及单一的前景对象类别；第二种则捕捉场景的几何结构（天空、垂线和水平线）。系统会为每个图像像素分配语义和几何两个标记；

微软剑桥研究院 (MSRC) 数据集³⁴，由 591 张图片组成。像素标记为 23 个不同的类别之一。然而，由于马和山的类别的出现次数非常少，它们通常不被使用。在训练和评价时，不属于剩余 21 个类别中任意类别的像素会被忽略。该数据集的缺点之一是，实准标记非常粗糙，而且在靠近对象边界时往往不正确。然而，该数据集包含多种多样的图像，并且被广泛使用。

随着场景理解研究日渐成熟，规模更大、图像多样性更佳的数据集，对于应用已有的场景解算法以及启发新算法都变得更为重要。PASCAL 视觉对象类 (VOC) 数据集⁶ 是一个非常大的图像数据集，其中的图像标注了对象边界框和 20 种不同的（前景）对象类别的像素级分割掩膜。它包含了约 20,000 张图片，他们被划分到多个挑战任务中，其中的训练、验证和评价图像集已经预先规定。场景理解研究人员感兴趣的另一个大型数据集是 SIFT Flow 数据集²⁹，它是从 LabelMe 图片库 (<http://labelme.csail.mit.edu>) 中选出的室外图片

子集，包含了 2,688 张图片，标记了 33 种不同的对象和背景类别。如果研究人员希望在这两种数据集上均取得良好的效果，则他们需要融合上文描述的多种技术。

场景理解算法的准确性可以通过很多手段评价，包括复杂的边界质量度量 and 交比并 (Jaccard) 的分数。最简单的方法是计算该模型在“隔离的”或单独的图片集上正确标记的像素的百分比。这些图片集又被称为“测试集”或“评价集”。作为评价机器学习算法的标准做法，这些图片不应该在训练模型参数时出现。在形式上，我们可以写成

$$acc = \frac{\sum_{i=1}^n [\hat{y}_i = y_i^*]}{n} \quad 7.$$

其中 \hat{y}_i 为算法预测的像素 i 的标记， y_i^* 是像素 i 的实准标记， $[\cdot]$ 为指标函数，当其变量为真时取值 1，否则取值 0。另一种评估度量可以更好地说明模型处理罕见类别的性能，这种度量为类别平均准确性，定义为

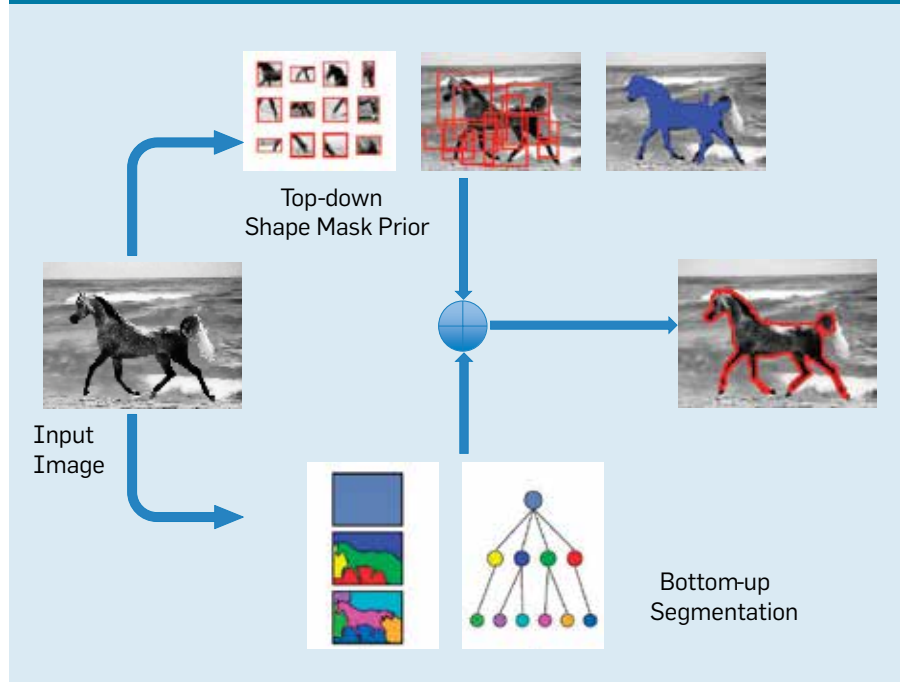
$$acc^{class-avg} = \frac{1}{|L|} \sum_{\ell \in L} \frac{\sum_{i=1}^n [\hat{y}_i = \ell] \wedge (y_i^* = \ell)}{\sum_{i=1}^n [y_i^* = \ell]} \quad 7.$$

方程 8 和方程 9 定义了不同的准确性衡量标准，它们在统计学上通常被分别称为“微观平均”和“宏观平均”。

对 MSRC 和斯坦福背景 (Stanford Background) 数据集进行语义分类时，最新水平的性能分别约为 86% 和 77% 的像素准确度；类别的平均准确度通常要低 5%-10%。在较大的数据集上时，如果没有使用自顶向下和上下文线索，性能可能会相当差，特别是对于那些出现次数较少的类别。

图 7 通过展示 MSRC 数据集中一幅样例图片的结果说明了场景理解模型在不同方面的效果。如图所示，单独对像素进行分类时（左侧的结果栏），产生了噪声非常大的结果。添加成对平滑项后，有助于去除噪声（右侧）。然而，当特征相当弱时（顶排），算法不能正确地得到图像中的对象类别，即便使用局部特征识别背景时可能相当容

图 6. 在 CRF 框架中融合自底向上和自顶向下信息。² 自底向上的过程形成局部假设，并把它们映射到语义标记上。自顶向下的过程代表了在对象实例上施加了基于部件的形状约束。



易。正如我们讨论的那样，更强的特征（包括局部和全局线索）与成对平滑项结合后，会产生正确的标记结果（底部右侧）。

语义分割的这些例子指出了场景理解算法中更普遍的趋势。总体来说，如果更复杂的特征包含了上下文信息（比如像素位置以及全局和基于形状的特征），则它们的性能情况要比局部外观特征好得多。不仅如此，与独立的像素分类相比，带有成对平滑先验的 CRF 模型表现更好，但是当独立分类器使用的特征的复杂度提高时，这种好处减少了。这种取舍的结果是可以预见的，因为这些特征不仅让基线的性能提高，还能让特征包含上下文信息，而这可替代平滑先验。

从 Darwin 软件框架中取得的

定性结果(见图 8)突出了下列几点；如图所示，预测的准确度通常相当好，而且模型能够很好地识别对象类别之间的边界。不过，前景对象的标记偶尔会渗入背景。在 MSRC 的结果中，这种渗透更为明显，而且这可部分归因于该数据集中的实准标记相当粗糙。在使用超像素的模型中，这些边界错误还可能源自不准确的过分割。

在图 8 中，一个有趣的结果是鸭子的标记 (MSRC, 左列第三行)。此处，水的类别正确标出，但是两只鸭子的类别不正确。模型错误地把白鸭标成了水，因为模型不仅混淆了鸭子的局部外观和水的局部外观，而且相当强的平滑假设倾向于把鸭子标记成与周围背景相同的类别。第二只鸭子被错误地标记为船。

虽然对人类来说这可能看起来很荒唐，对算法来说这是个合理的错误，因为上下文是正确的（船也跟水同时出现），而且训练模型时，只有少量的图片里有鸭子。

结语

我们从像素标记任务的视角探索了场景理解问题，其中包含了场景理解算法面临的一些挑战。本文还简略介绍了处理这些问题的当前趋势。在这些趋势中活跃的各种研究以及越来越多的可用高质量数据集反映了场景理解的研究现状；例如，更好的低层特征表示现在可以从大量的数据中自动学习得出，而不再人工设计。²⁶ 研究人员也正在研究中层视觉线索（也称为“属性”）来克服因为训练数据稀少而造成的某些局限；例如，知道对象拥有羽毛后，可以缩小可用于该对象的标记的范围。⁷

不仅如此，基于结构化预测模型的改进学习算法³¹意味着可以同时调整大量的参数。这一成果不仅会产生更优的参数，还允许使用更丰富的模型（比如带有参数化高阶项的模型）。其他目前正被研究的模型是混合模型（比如网格结构化 CRF），它融合了其他的形状约束模型【例如受限玻尔兹曼机 (RBM) 和多尺度模式】。此时，CRF 捕

图 7. MSRC 数据集中某图片的语义分割样例

图中列出了原始图像（左侧）以及不同场景理解模型所得出的颜色编码像素标记（右侧）。各模型的特征（比较局部外观与局部和全局外观）和模型复杂度（比较独立的像素分类与使用成对项的 CRF 模型）存在差异；相关的颜色图例参见图 8。

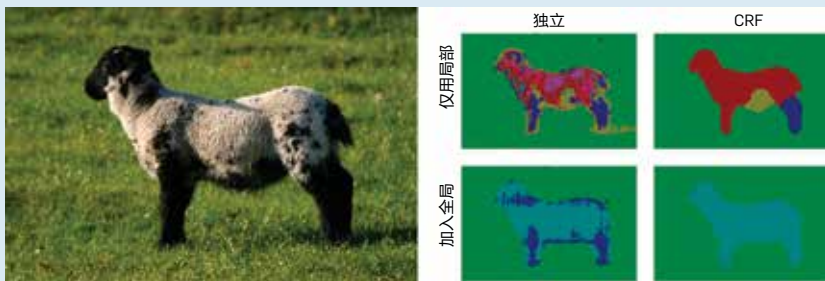
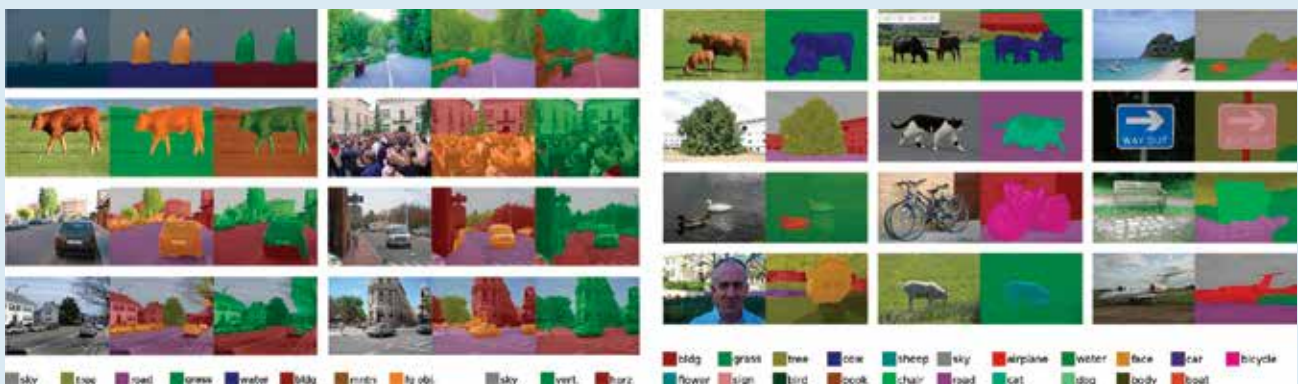


图 8. 使用 Darwin 软件库处理两个标准场景理解数据集后得出的代表性结果；图中给出了原图以及覆盖了预测出的类别标记的图像；彩色时效果最佳。



捉了外观和平滑性，而 **RBM** 及其变体则促进了对对象形状的全局一致性。这些模型虽然才刚刚起步，但已经展现了广阔的前景。

另一个激动人心的方向是，使用非参数标签迁移技术可以在数据集规模和对象类别的多样性方面获得更高的可扩展性。^{29,36} 这些技术虽然（多少）克服了 **CRF** 模型中隐含的封闭世界标记集假设，但引入了其他的复杂性（比如更消耗资源的测试时计算），并需要消除语言歧义（比如，“水”和“河流”是否在语义上等价；更重要的是，它们是否指同一个对象）。

人们可能会问，是否有必要标记图像中的每个像素。实际上，对于某些场景理解任务而言（比如人脸检测），粗略的边界框可能足够了。然而，为了得到场景的详细描述，在早期计算机视觉研究人员设想的各个方向中，像素级的标记似乎都无法避免。这一研究前沿的进步（比如本文中描述的那些进步）及其与一致的整体模型的融合让我们越来越接近计算机真的能够描述它看到的场景这一目标。

参考资料

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 11 (Nov. 2012), 2274–2282.
- Borenstein, E., Sharon, E., and Ullman, S. Combining top-down and bottom-up segmentation. In *Proceedings of the IEEE Workshop on Perceptual Organization in Computer Vision at the IEEE Conference on Computer Vision and Pattern Recognition* (Washington, D.C., June 27–July 2). IEEE Computer Society Press, 2004, 46–46.
- Boykov, Y., Veksler, O., and Zabih, R. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 11 (Nov. 2001), 1222–1239.
- Comaniciu, D. and Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 5 (May 2002), 603–619.
- Dalal, N. and Triggs, B. Histograms of oriented gradients for human detection. In *Proceedings of the Conference on Computer Vision and Pattern Recognition* (San Diego, CA, June 20–25). IEEE Computer Society Press, 2005, 886–893.
- Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., and Zisserman, A. The Pascal visual object classes challenge. *International Journal of Computer Vision* 88, 2 (June 2010), 303–338.
- Farhadi, A., Endres, I., Hoiem, D., and Forsyth, D. Describing objects by their attributes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Miami, FL, June 20–25). IEEE Computer Society Press, 2009, 1778–1785.

- Felzenszwalb, P.F. and Huttenlocher, D.P. Efficient graph-based image segmentation. *International Journal of Computer Vision* 59, 2 (Sept. 2004), 167–181.
- Fulkerson, B., Vedaldi, A., and Soatto, S. Class segmentation and object localization with superpixel neighborhoods. In *Proceedings of the 12th International Conference on Computer Vision* (Kyoto, Japan, Sept. 29–Oct. 2). IEEE Computer Society Press, 2009, 670–677.
- Gould, S., Fulton, R., and Koller, D. Decomposing a scene into geometric and semantically consistent regions. In *Proceedings of the 12th International Conference on Computer Vision* (Kyoto, Japan, Sept. 29–Oct. 2). IEEE Computer Society Press, 2009, 1–8.
- Gould, S., Gao, T., and Koller, D. Region-based segmentation and object detection. In *Advances in Neural Information Processing Systems 22* (Vancouver, B.C., Canada, Dec. 6–11). Curran Associates, Inc., 2009, 655–663.
- Gould, S., Rodgers, J., Cohen, D., Elidan, G., and Koller, D. Multi-class segmentation with relative location prior. *International Journal of Computer Vision* 80, 3 (Dec. 2008), 300–316.
- He, X., Zemel, R.S., and Carreira-Perpinan, M. Multiscale conditional random fields for image labeling. In *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Washington, D.C., June 27–July 2). IEEE Computer Society Press, 2004, 695–702.
- Hedau, V., Hoiem, D., and Forsyth, D. Recovering the spatial layout of cluttered rooms. In *Proceedings of the International Conference on Computer Vision* (Kyoto, Japan, Sept. 29–Oct. 2). IEEE Computer Society Press, 2009, 1849–1856.
- Heitz, G., Gould, S., Saxena, A., and Koller, D. Cascaded classification models: Combining models for holistic scene understanding. In *Advances in Neural Information Processing Systems 21* (Vancouver, B.C., Canada, Dec. 8–13). Curran Associates, Inc., 2008, 641–648.
- Heitz, G. and Koller, D. Learning spatial context: Using stuff to find things. In *Proceedings of the European Conference on Computer Vision* (Marseille, France, Oct. 12–18). Springer, Berlin, Heidelberg, 2008, 30–43.
- Hoiem, D., Efros, A.A., and Hebert, M. Recovering surface layout from an image. *International Journal of Computer Vision* 75, 1 (Oct. 2007), 151–172.
- Hoiem, D., Efros, A.A., and Hebert, M. Closing the loop on scene interpretation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Anchorage, AK, June 23–28). IEEE Computer Society Press, 2008, 1–8.
- Hoiem, D., Efros, A.A., and Hebert, M. Putting objects in perspective. *International Journal of Computer Vision* 80, 1 (Oct. 2008), 3–15.
- Kohli, P., Ladicky, L., and Torr, P.H. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision* 82, 3 (May 2009), 302–324.
- Kolmogorov, V. and Zabih, R. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 2 (Feb. 2004), 147–159.
- Komodakis, N., Paragios, N., and Tziritas, G. MRF optimization via dual decomposition: Message-passing revisited. In *Proceedings of the International Conference on Computer Vision* (Rio de Janeiro, Oct. 14–21). IEEE Computer Society Press, 2007, 1–8.
- Krahenbuhl, P. and Koltun, V. Efficient inference in fully connected CRFs with gaussian edge potentials. In *Advances in Neural Information Processing Systems 24* (Granada, Spain, Dec. 12–17). Curran Associates, Inc., 2011, 109–117.
- Ladicky, L., Russell, C., Kohli, P., and Torr, P.H. Graph cut-based inference with co-occurrence statistics. In *Proceedings of the 11th European Conference on Computer Vision* (Crete, Greece, Sept. 5–11). Springer, Berlin, Heidelberg, 2010, 239–253.
- Lafferty, J.D., McCallum, A., and Pereira, F.C.N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning* (Williamstown, MA, June 28–July 1). Morgan Kaufmann, San Francisco, 2001, 282–289.
- Le, Q.V., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G.S., Dean, J., and Ng, A.Y. Building high-level features using large scale unsupervised learning. In *Proceedings of the International Conference on Machine Learning* (Edinburgh, Scotland, June 26–July 1). Morgan Kaufmann, San Francisco, 2012.
- Levin, A. and Weiss, Y. Learning to combine bottom-up and top-down segmentation. *International Journal of Computer Vision* 81, 1 (Sept. 2008), 105–118.
- Li, L.-J., Socher, R., and Fei-Fei, L. Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Miami, FL, June 20–25). IEEE Computer Society Press, 2009, 2036–2043.
- Liu, C., Yuen, J., and Torralba, A. Nonparametric scene parsing via label transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 12 (Dec. 2011), 2368–2382.
- Malik, J., Belongie, S., Shi, J., and Leung, T. Textons, contours and regions: Cue integration in image segmentation. In *Proceedings of the International Conference on Computer Vision* (Corfu, Greece, Sept. 20–25). IEEE Computer Society Press, 1999, 918–925.
- Nowozin, S. and Lampert, C.W. Structured learning and prediction in computer vision. *Foundations and Trends in Computer Graphics and Vision* 6, 3–4 (May 2011), 185–365.
- Rabinovich, A., Vedaldi, A., Galleguillos, C., Wiewiora, E., and Belongie, S. Objects in context. In *Proceedings of the International Conference on Computer Vision* (Rio de Janeiro, Oct. 14–21). IEEE Computer Society Press, 2007, 1–8.
- Ren, X. and Malik, J. Learning a classification model for segmentation. In *Proceedings of the International Conference on Computer Vision* (Nice, France, Oct. 13–16). IEEE Computer Society Press, 2003, 10–17.
- Shotton, J., Winn, J., Rother, C., and Criminisi, A. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *Proceedings of the European Conference on Computer Vision* (Graz, Austria, May 7–13). Springer, Berlin, Heidelberg, 2006, 1–15.
- Szeliski, R. *Computer Vision: Algorithms and Applications*. Springer, Berlin, Heidelberg, 2011.
- Tighe, J. and Lazebnik, S. SuperParsing: Scalable nonparametric image parsing with superpixels. In *Proceedings of the European Conference on Computer Vision* (Crete, Greece, Sept. 5–11). Springer, Berlin, Heidelberg, 2010, 352–365.
- Tu, Z., Chen, X., Yuille, A.L., and Zhu, S.-C. Image parsing: Unifying segmentation, detection and recognition. *International Journal of Computer Vision* 63, 2 (July 2005), 113–140.
- Tu, Z. and Zhu, S.-C. Image segmentation by data-driven Markov chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 5 (May 2002), 657–673.
- Wang, H., Gould, S., and Koller, D. Discriminative learning with latent variables for cluttered indoor scene understanding. In *Proceedings of the 11th European Conference on Computer Vision* (Crete, Greece, Sept. 5–11). Springer, Berlin, Heidelberg, 2010, 497–510.
- Yao, Y., Fidler, S., and Urtasun, R. Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition* (Providence, RI, June 16–21). IEEE Computer Society Press, 2012, 702–709.

Stephen Gould (Stephen Gould) 是堪培拉澳大利亚国立大学工程与计算机科学学院计算机科学研究所的研究员 (高级讲师) , 兼任堪培拉澳大利亚国立信息通信技术研究中心机器学习研究组的客座研究员。

Xuming He (xuming.he@nicta.com.au) 是堪培拉澳大利亚国立信息通信技术研究中心计算机视觉研究组的高级研究员 , 兼任堪培拉澳大利亚国立大学工程与计算机科学学院工程研究院兼职研究员 (高级讲师) 。

译文责任编辑: 山世光

**强关系连接了相同社区中的个人；
弱关系连接的不同社区中的个人。**

作者：PASQUALE DE MEO、EMILIO FERRARA、
GIACOMO FIUMARA 以及 ALESSANDRO PROVETTI

Facebook 上的 关系大多为弱 关系

无处不在的社交科技网络为开发人员和用户带来了概念性和技术性挑战。评估用户关系的紧密程度以及用户关系对信息交流和传播的促进作用已成为一项核心研究课题。在社科领域，人际关系的这些方面已经在马克格兰诺维特 (Mark Granovetter) “弱关系的力量”这一理论框架下得到了广泛研究¹³。一些研究讨论了该理论是否适用于 Facebook 等在线社交网络，并指出利用互动数据可以预测用户关系的强度。这些研究所采用的方法必须处理用户生成数据。而出于隐私考虑，此类数据通常并不公开。

我们提出另一种只依赖社交网络拓扑结构（如 Facebook 上的好友关系）的强弱关系定义方法。在线社交网络 (OSN) 往往可以划分位不同社区，因此，我们建议将连接不同社区用户的边归为弱关系，将连接相同社区用户的边归为强关系。我们使用代表 Facebook 部分社交图谱的大型网络来检验这一定义，并研究了强弱关系对信息传播过程的影响。我们的研究表明，OSN 中的个人会自发组建起联系紧密的社区，而弱关系可以产生凝聚力并优化信息传播的范围。

要想分析并理解 Facebook 等 OSN，必须具备如由 Kleinberg 提出的理论基础。¹⁵ 然而，考虑到 OSN 常规用户的可获取数据的规模、分布和组织方式（如隐私和监控规则），对真实的 OSN 进行研究无疑是计算机科学领域的一项巨大挑战。¹

在这种背景下，对 OSN 的大型子集进行分析，便可生成一系列在统计学上可靠的测量值，而它们正是理解 OSN 结构与演化的基础。实际上，综合性指标在数据管理、隐私管理和在线营销方面具有重要价值。评估用户关系的紧密程度并确定用户关系对信息传播的促进作用是一项难以解决的问题。此前，此类用户连接和社交沟通问题已在社科领域加以研究，并且主要以格兰诺维特的理论为框架。¹³

弱关系是指社交图谱上相距较远区域的个人之间的关系，也是来自不同国家、语言体系、年龄段或共同经历群体的个人之间最常见的关系。弱关系是跨长社交距离向各阶层的群体传播信息的有力方式。强关系是相互信任或熟识的个人之间的关系，如家庭纽带和亲密友谊。

格兰诺维特提出的弱关系是否也以其设想的方式存在于 Facebook 等 OSN 中，或者表现为来自

» 重要见解

- 在线社交网络可以促进由相隔较远的社区组成的好友网络的发展；切断弱关系会对整个网络的连接状况造成巨大影响。
- 我们重新定义了强弱关系，以反映 Facebook 等大型在线社交网络的结构和演化。
- 当在线社交网络内的用户自发组成社区时，一些有趣的属性就会显现出来。



图片由 ALEKSANDAR MIJATOVIC 提供

社交图谱不同区域的个人之间的关系? 这一问题难以回答, 原因至少包括以下两方面: **Facebook** 主要依靠记录好友关系组建起来。这意味着, 好友关系体现(并整合)了人际关系的多种等级和细微差别, 而这些等级和差别难以通过分析在线数据来分类并描述。此外, **Facebook** 的规模和复杂性不断增

加, 好友关系网络日趋密集而不是稀疏。¹

随着各 OSN 的互连程度越来越高, 检验格兰诺维特理论面临着可扩展性方面的挑战。**Gilbert** 和 **Karahalios**⁹ 的早期研究采用了监督式方法, 让一组 **Facebook** 用户评估他们的好友关系的强度。面对如今的规模庞大的 OSN, 以

Gilbert 和 **Karahalios**⁹ 的方法对格兰诺维特理论开展大规模研究恐怕是难以付诸实践的。在其他方法中, **Bakshy** 等人² 的方法尤其引人注目, 他们访问了 **Facebook** 的用户活动数据, 并将关系强度作为关于用户互动类型和频率的函数进行计算。然而, 该方法必须设定区分强弱关系的中止阈值, 该阈值的

任何调整都会严重影响到弱关系的判定。

我们在本文中提出了一种定义弱关系的新方法，该方法以大型 OSN 的分析为依据，并考虑了相关的计算难题。我们首先识别网络内的社区，然后将连接不同社区用户的边归为弱关系，将连接相同社区用户的边归为强关系。按照我们的定义来确定弱关系具有快捷和可靠的优点，因为用于查找大型网络中的社区的当前算法十分高效⁸，而且我们无需规定任何阈值。我们针对由 Gjoka 等人¹⁰ 收集并发布的 Facebook 好友关系公开数据集开展了广泛的实验性分析，并将其与随机生成的图谱进行零模型比较。我们使用了著名的社区检测算法：Louvain Method (LM)³ 和 Infomap。²⁰

我们对实验结果进行分析，并提炼出以下观点：

结果不受社区检测算法的影响。利用 LM 和 Infomap 分别找出的弱关系大致吻合，这说明我们的弱关系定义基本上不受所选社区检测算法的影响；

弱关系的数量多于强关系。社区检测算法找到的弱关系的数量多于强关系；

弱关系出现更频繁。弱关系出现在小型社区中的频率更高；以及

弱关系有助于信息传播。按照我们的方法所确定的弱关系在网络信息传播中发挥关键作用，它们的移除将减少可通过信息传播而到达的网络的数量。

弱关系和强关系

马克·格兰诺维特提出了关于关系强度的经典定义：¹³ “关系强度取决于交往时间、情感度、亲密度（互信度）以及关系中的互惠行为这些因素，可能是这些因素的线性组合。” 该定义指出了社交关系的重要特征，即关系强度、联系频率以及互惠行为等。格兰诺维特认为，应根据与社

我们建议将连接不同社区用户的边归为弱关系，将连接相同社区用户的边归为强关系。

交网络结构相关的拓扑信息来判定强弱关系。因此，他提出了“关系桥梁”的概念：“关系桥梁是人际关系网络中的沟通两个人的唯一人脉。一般而言，每个人都拥有众多联系人，甲的任何联系人的信息或影响力若想传播到乙的任意一个联系人，都要通过甲与乙的关系，那么甲与乙之间存在关系桥梁。” 请注意，所有关系桥梁都是弱关系。但遗憾的是，关系桥梁这一定义在 OSN 中具有局限性并且无法适用。根据著名的“小世界效应”（一对顶点间存在最短的连接路径）和“无标度网络”（维系整个网络的枢纽是存在的）理论，一条被移除后可造成终端顶点连接完全断开的边是不太可能存在的。

为确保格兰诺维特的定义适用于 OSN，我们重新定义了一种“捷径桥梁”，它也是一条连接一对顶点的边，被移除后可导致这对顶点之间的距离变长。这里距离是指连接这两个顶点的最短路径的长度。这一定义尽管在其他方面看似合理，但实际上更具争议性。首先，它建立在最短路径概念的基础上，但遗憾的是，即使对中等规模的网络来说，确定各对顶点的最短路径在计算上也是不可行的。其次，即使对距离进行重新定义，采用“加权路径”（按分配至每个使用中的连接的权重之和进行计算），这对 OSN 来说也是不具备计算可行性的。我们的目标是探索一种新颖且具备计算可行性的弱关系定义，使之适用于大型 OSN 的分析。我们并没有使用基于两个用户之间的互动强度的关系型定义，而是提出了一种基于社区的定义。我们这样定义弱关系：将网络划分为不同社区后，即可获得所谓的“社区结构”，那些连接来自不同社区顶点的边就是弱关系。此后，我们将社区内的边归为强关系。

弱关系的一个重要特征就是，属于关系桥梁（按照格兰诺维特的

观点)的弱关系可以建立更多且更短的路径。从社区结构的角度来看,移除弱关系会比移除强关系造成更大破坏。这甚至也解释了弱关系(定义方式略微不同)为何能够高效传播信息。^{5,21}

我们的定义的优点。我们对弱关系的定义拥有四个吸引人的特征:

比格兰诺维特的定义更细化。尽管由弱关系连接的顶点属于不同的社区,但这对顶点之间的边未必是“关系桥梁”,移除这条边可能不会断开顶点之间的连接,而且在实际中,断开情况几乎不会出现;

以拓扑信息为依据。拓扑信息是划分强弱关系的唯一依据。Zhao等人²¹仅使用局部的拓扑信息,即相邻的两个终端顶点的拓扑信息,而我们的定义以全球信息为处理对象,因为该定义基于对整个网络的划分;

二元化。我们的定义将网络中的每条边标记为强弱关系之一。因此,我们无需调整任何阈值,也不必将低于该阈值的边归类为弱关系,否则我们将无法根据强度来比较两条边;以及

准确发现社区。我们的实验结果表明,我们对弱关系的定义是可靠的,社区发现结果不会受到所选的社区检测算法的影响。⁸

相关方法

一些研究项目考察了非拓扑信息中的弱关系的强度。为使用非拓扑方法来确定关系的强弱等级,研究人员必须选择一些可测变量,并根据这些变量推导出连接两个用户的关系的强度。在这种情况下,弱关系表现为一种熟人关系,即它们连接的是互动频率低、彼此不会造成重大影响的熟人。对于“离线”社交网络,Marsden和Campbell¹⁷列举了一些可作为弱关系指标的变量(如沟通互惠行为,以及存在至少一个共同朋友)。

在社交Web的背景下,Gilbert和Karahalios⁹以及Petroczi等人¹⁹分别对拥有35名参与者和56名参与者的小型OSN进行了研究,他们通过向参与者直接提问来评估数个关系强度指标(如亲密度/密切度、互惠度、社交性/交际性),并以此为依据向各类关系分配权重。Gilbert和Karahalios⁹将Marsden-Campbell的方法应用到Facebook,并列出了74项变量作为潜在的关系强度预测因子。随后,他们将关系强度建模为变量的线性组合,并采用常规最小二乘回归法的变体来计算权重。为验证该模型的有效性,Gilbert和Karahalios⁹招募了35名用户,并要求他们对Facebook上的好友进行评级。Gilbert和Karahalios的计算结果的准确率高达85%左右,但是他们很难在大型OSN片段上取得相同的业绩。要使用这种方法,就必须收集大量关于用户行为的信息。考虑到隐私问题和专属数据使用受限问题,所需的数据大多不会面向学术研究开放。

Zhao等人²¹建立了关系强度的计算公式,权重 w_{ij} 是分配给连接顶点 i 和 j 的边的权重:

$$w_{ij} = \frac{c_{ij}}{k_i + k_j - 2 - c_{ij}}$$

其中 k_i 和 k_j 是 i 和 j 的等级,而 c_{ij} 是共同熟人的数量。Zhao等人²¹的数值模拟结果显示,如果用较低的 w 值来逐步移除连接,信息覆盖率就会急剧下降。

Zhao等人的研究结果印证了这一假设,即弱关系是信息传播的关键。Bakshey等人²在2012年发表了一篇关于Facebook中的弱关系的论文,并在文中探讨了四项用于衡量连接强度的参数,它们分别是:互发私信的频率、在彼此的帖子上公开留言的频率、共同出现在

照片中的次数以及共同在他人的帖子上留言的次数。

Bakshy等人²没有像我们那样将Facebook社交连接的拓扑层面纳入考虑,至少他们没有明确地这样做过。此外,他们选用的分析方法需要使用专属信息和用户活动记录,所以我们的方法不容易相互比较。我们的研究更接近Blondel等人²和Zhao等人²¹的研究,主要体现在我们都将弱关系视为有利于信息传播的有效连接,并且在Facebook上对弱关系进行检验(如Blondel等人的实验²)。

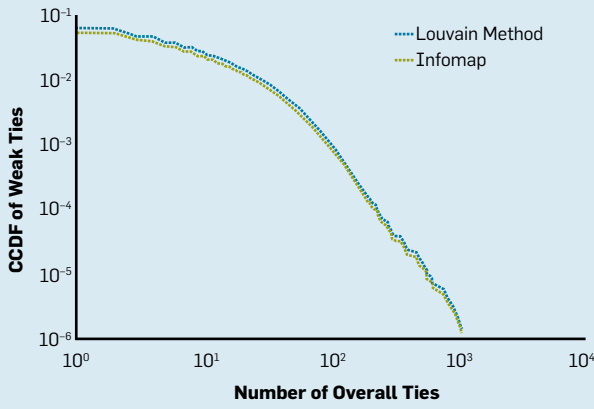
然而,有一项区别是不容忽视的:Blondel等人和Zhao等人均为关系分配了分数,并根据一个阈值来分类强弱关系。而我们划分强弱关系的依据是,该关系是否连接来自不同社区的顶点。这种分类方法是二元化的,无需使用分数和难以妥善确定并调整的阈值。Grabowicz等人¹²采用一种类似的方法,他们使用关于网络拓扑的信息来确定弱关系。然而,他们以Twitter为重点研究对象,Twitter可以建模为用户关系大多不对称的定向网络。Twitter的“转发”和“提醒”功能对弱关系的确定有重要影响,因此我们的研究难以相互比较。

结果

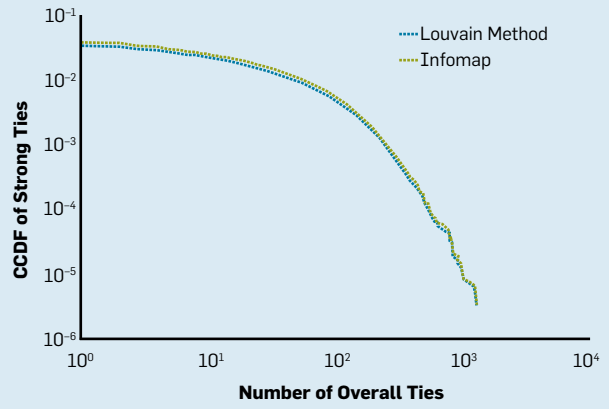
我们的试验结果反映了我们的弱关系定义的利与弊。在最初的社区检测阶段,我们采用两种流行算法:LM³和Infomap。²⁰我们考虑了两种测试平台,并将它们与Gjoka等人¹⁰收集的Facebook网络片段(包含957,000个用户和5,840万个好友关系)^a以及由Erdős-Rényi随机图谱(包含多达2048个顶点)构成的“零模型”网络进行比较;我们将任意一对顶点间存在边的可能性在0.05到

a 关于数据集的完整说明,请访问<http://informatica.unime.it/weak-ties/>并阅读补充材料。

图 1. 关于 Facebook 数据集中的 (a) 弱关系和 (b) 强关系的互补累积分布函数。

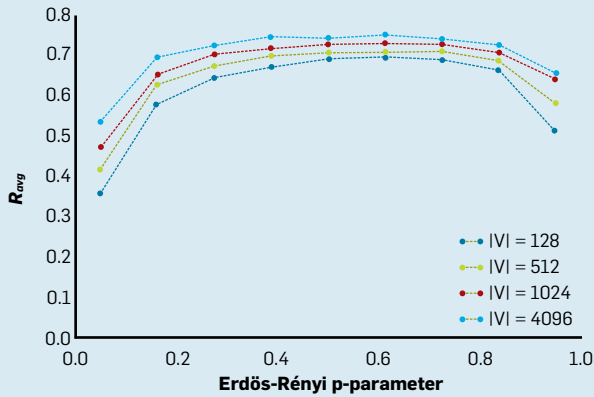


(a)

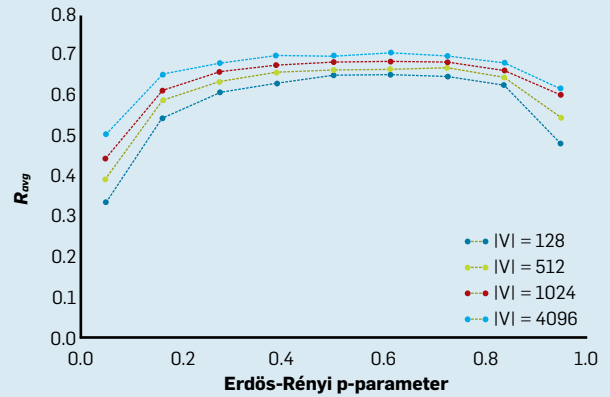


(b)

图 2: 将 (a) LM 和 (b) Infomap 应用于 Erdős-Rényi 随机图谱所得到的 R_{avg} 比值 (以 $IVI = 128, 512, 1,024, 4,096$ 为例)。

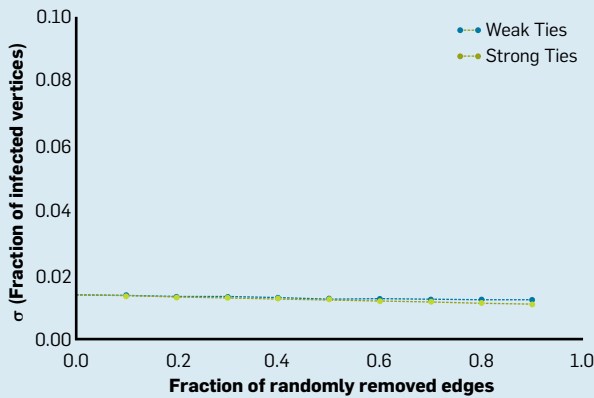


(a)

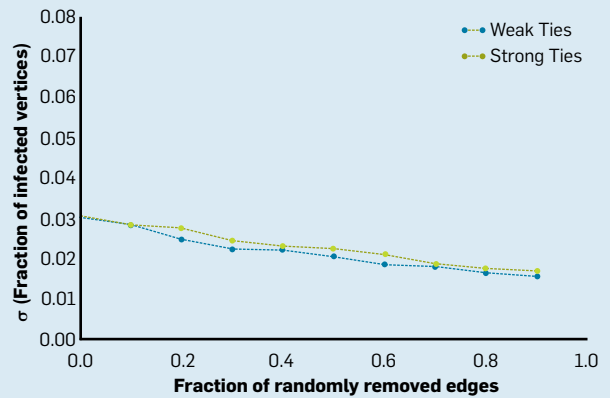


(b)

图 3. 不同 τ 值的图谱覆盖率 (a) 当 τ 在 0 - 0.9 变化时, Erdős-Rényi 随机图谱的覆盖率 ($IVI = 512, plink = 0.05, pinf = 0.01$); (b) 当 τ 在 0 - 0.9 变化时, 生成的 Facebook 数据集的覆盖率。图 (a) 和图 (b) 记录了 LM 的结果。



(a)



(b)

0.95 之间调动，每次的调幅为 0.05。

定义的可靠性。作为一项初期实验，我们研究了弱关系定义的可靠性，以判断其是否会受到社区查找方法的影响。如果使用不同的社区检测方法可能产生不同的结果，即便各结果之间的差别很轻微，这也说明我们的强弱关系分类方法受制于社区检测方法。我们对 Facebook 样本应用了 LM 和 Infomap 算法，并使用归一化互信息 (NMI) 来比较这两种算法所发现的社区结构。我们得到的 NMI 约为 0.9，这表明两种算法的差异水平相当低。^b

我们的结论是，我们的弱关系定义在应用于 Facebook 时是可靠的，并且不受特定社区检测算法的影响。但需要注意的是，由 LM 和 Infomap 发现的社区存在着微妙差别，但仍然具有 NMI 高的特点。接下来，我们将以图表形式详细地比较使用 LM 和 Infomap 所得到的结果，以便更好地理解采用特定社区检测算法的实际后果。

Facebook 上的强弱关系。在本节，我们考虑了强弱关系在 Facebook 上的分布。我们计算了弱关系（与强关系相对）的“互补累积分布函数” (CCDF)，该函数代表找到一个拥有 k 等级以上的弱关系（与强关系相对）的顶点的几率（参见图 1）。

强关系的 CCDF 较弱关系的 CCDF 下降更快。当 $k = 4$ 时，我们看到一个临界点，超过该点以后，弱关系在数量上迅速超过了强关系；也就是说，与弱关系相比，强关系在更高等级顶点的出现频率更低。

我们的实验结果与格兰诺维特的弱关系“构想”基本一致。社会学理论（如三元闭包、¹³ 认知平衡、

我们的研究结果表明，OSN 中的个人会自发组建起联系紧密的社区，而弱关系可以产生凝聚力并优化信息传播的范围。

¹⁴ 趋同性¹⁸) 指出，个人往往愿意聚集在小型社区。

因此，我们可以确定，小型群体内的人际关系更加紧密；如果个人分属相距较远的社区，那么他们之间的关系强度就会下降。按照格兰诺维特的观点，大多数连接都是联系次数少、互动频率低的弱关系。

随机图谱中的强弱关系。在本节，我们对 Erdős-Rényi 随机图谱中的强弱关系分布进行了对比分析；我们将核实从 Facebook 数据中得出的结论能否在图谱中依然有效。虽然图谱的结构可以提前获知，但图谱在构造上缺少明显的社区结构。我们计算了弱关系数量与关系总数的比值 R_{avg} ；图 2 描绘了不同 $|V|$ 值的 R_{avg} 以及任意两定点间存在边的可能性 p_{link} 在 0.05 到 0.95 之间的均匀变化。

我们可以从中得出以下结论：

弱关系的数量多于强关系。

R_{avg} 总是大于 0.6，这说明弱关系的数量多于强关系，在随机生成的图谱中也不例外；

稳定性和独立性。 R_{avg} 保持相对稳定，并且不受 p_{link} 影响； G 的稀疏程度对弱关系数量的影响十分有限；以及

影响有限。 $|V|$ 对 R_{avg} 的影响有限；当 $|V|$ 从 128 增加到 4,096 时（增长了 32 倍）， R_{avg} 仅上升了 17.14%。

弱关系在信息传播中的作用。

我们还研究了弱关系对信息传播流程的影响，并阐明了我们的定义与格兰诺维特的定义之间的联系。我们认为，弱关系是将原本互不相连的个人连接起来的特定关系，而这些个人分属社交图谱中相距较远的地区。

格兰诺维特所定义的弱关系应在信息传播中发挥重要作用。我们定义的弱关系又是如何传播信息的呢？

我们使用独立级联模型 (ICM)¹¹ 来模拟信息在网络上的传播，并再

^b 关于 NMI 的定义，请参阅补充材料的脚注 a；如果 NMI 介于 0-1 之间，并且接近 1，那么两种算法找到社区基本吻合。

次比较了 Facebook 数据集和 Erdős-Rényi 的随机图谱。我们在 ICM 中均匀地随机选取顶点 v_0 ，并从该顶点向周围发送信息，信息成功传播的概率（传染概率）为 p_{inf} 。

每个“受感染”的结点重复上述过程继续将消息传播给邻居。Leskovec 等人¹⁶指出，合理的 p_{inf} 值为 0.01、0.02 和 0.03。

为生成具有统计学意义的结果，我们将顶点 v_0 的初始值选定为 250 次；我们在每次选择 v_0 时，都会模拟信息的传播，并测算信息覆盖率 σ ，即收到信息的顶点（被传染的顶点）数量与顶点总数的比值。我们逐步（随机地）删除一部分弱关系 τ ，并不断重复该实验。

在模拟中， τ 在 0.1 - 0.9 的范围内变化，我们计算了每个值的相应覆盖率。我们针对弱关系和强关系分别执行了上述程序。图 3 和补充材料 (<http://informatica.unime.it/weak-ties/>) 中显示了将 LM 和 Infomap 分别应用于 Erdős-Rényi's 随机图谱所得到的 σ 值。我们在这里考虑了 $|V| = 512$ 个顶点、 $p_{\text{link}} = 0.05$ 、 $p_{\text{link}} = 0.03$ 的图谱。

请注意，逐渐移除弱连接可导致 σ 值降低；按平均值计算， σ 值的最大降幅为 11.98%，平均降幅为 5.71%，标准偏差值为 3.4%。我们发现，移除强连接也会造成 σ 值降低，但降幅并不明显。据观察，如果 $|V|$ 增加，覆盖率 σ 也会增加。图 3b 和补充材料中记录了与 Facebook 数据集相关的 σ 值。

如果 τ 为固定值，那么与移除强关系相比，移除弱关系会造成 σ 出现更明显的下降。根据我们的观察， σ 值的最大降幅为 14.31% 左右，平均降幅为 6.26%，标准偏差值为 3.1%。

这些数值总是大于 Erdős-Rényi 的随机图谱中的数值，这证明移除弱关系对真实 OSN 的影响要大于对不具备有意义社区结构的网络的影响。我们由此得出以下结论：

我们提出的弱关系定义把握了格兰诺维特观点的实质，即移除弱关系比移除强关系更显著地降低 / 妨碍信息的流通。

总结

我们根据网络自身的社区结构提出了适用于 Facebook 等 OSN 的全新弱关系定义。我们针对包含 957,000 名用户的 Facebook 样本和随机生成的图谱进行了实验，并以此强调了弱关系的作用和重要性。我们使用关于社区规模和密度的函数来描述弱关系的整体统计分布。我们研究了强弱关系在信息传播过程中的作用，研究结果表明，我们的适用于 OSN 的弱关系定义与马克·格兰诺维特的原始定义存在一定关联。

尽管多个最新研究项目都侧重于 Facebook 的社交图谱、¹⁴ 社交结构、⁷ 弱关系本身，² 但我们的基于社区的弱关系定义更适用于 Facebook 和类似的大型（密集）OSN。

今后，我们希望跟进以下两个项目：第一个项目的研究了网络加权策略的适用性，以便根据给定的基本原理（如每个连接都能传播信息）来计算关系强度；我们拟采用一种适用于 OSN 的全新的加权边方法，该方法是由我们在 De Meo 等人⁶ 的论文中设计的。第二个项目旨在分析与 Facebook 用户相关的地理数据；由于不同图谱会发生合并（如社会图谱和地理图谱），我们希望进一步了解物理和虚拟距离的作用。

鸣谢

感谢 Minas Gjoka 向我们提供了 Facebook 数据集，也感谢匿名审阅人提出的建设性意见。 □

参考资料

1. Backstrom, L., Boldi, B., Rosa, M., Ugander, J., and Vigna, S. Four degrees of separation. In *Proceedings of the ACM Web Science Conference* (Evanston, IL, June). ACM Press, New York, 2012, 33–42.
2. Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. The role of social networks in information diffusion. In *Proceedings of the World Wide Web Conference* (Lyon, France). ACM Press, New York, 2012, 519–528.

3. Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* (Oct. 2008).
4. Catanese, S., De Meo, P., Ferrara, E., Fiumara, G., and Provetti, A. Crawling Facebook for social network analysis purposes. In *Proceedings of the International Conference on Web Intelligence, Mining and Semantics* (Sogndal, Norway, May 25–27). ACM Press, New York, 52:1–52:8.
5. Centola, D. The spread of behavior in an online social network experiment. *Science* 329 (Sept. 2010), 1194–1197.
6. De Meo, P., Ferrara, E., Fiumara, G., and Ricciardello, A. A novel measure of edge centrality in social networks. *Knowledge-Based Systems* 30 (June 2012), 136–150.
7. Ferrara, E. A Large-scale community structure analysis in Facebook. *EPJ Data Science* 1, 1 (Nov. 2012), 1–30.
8. Fortunato, S. Community detection in graphs. *Physics Reports* 486 (Feb. 2010), 75–174.
9. Gilbert, E. and Karahalios, K. Predicting tie strength with social media. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (Boston, Apr.). ACM Press, New York, 2009, 211–220.
10. Gjoka, M., Kurant, M., Butts, C.T., and Markopoulou, A. Practical recommendations on crawling online social networks. *IEEE Journal on Selected Areas in Communications* 29, 9 (Oct. 2011), 1,872–1,892.
11. Goldenberg, J., Libai, B., and Muller, E. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* 12, 3 (2001), 211–223.
12. Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M., and Eguiluz, V.M. Social features of online networks: The strength of intermediary ties in online social media. *PLOS ONE* 7, 1 (Jan. 2012).
13. Granovetter, M.S. The strength of weak ties. *American Journal of Sociology* 78, 6 (May 1973), 1360–1380.
14. Heider, F. *The Psychology of Interpersonal Relations*. Lawrence Erlbaum, New York, 1982.
15. Kleinberg, J. The convergence of social and technological networks. *Commun. ACM* 51, 11 (Nov. 2008), 66–72.
16. Leskovec, J., Adamic, L., and Huberman, B. The dynamics of viral marketing. *ACM Transactions on the Web* 1, 1 (May 2007).
17. Marsden, P.V. and Campbell, K.E. Measuring tie strength. *Social Forces* 63, 2 (Dec. 1984), 482–501.
18. McPherson, M., Smith-Lovin, L., and Cook, J.M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27, 1 (Aug. 2001), 415–444.
19. Petroczi, A., Nepusz, T., and Bacs, F. Measuring tie-strength in virtual social networks. *Connections* 27, 2 (2006), 39–52.
20. Rosvall, M. and Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105, 4 (Jan. 2008), 1,118–1,123.
21. Zhao, J., Wu, J., and Xu, K. Weak ties: Subtle role of information diffusion in online social networks. *Physical Review E* 82, 1 (July 2010).

Pasquale De Meo (pdemeo@unime.it) 是意大利墨西拿大学古现代文明系的助理教授。

Emilio Ferrara (ferrarae@indiana.edu) 是美国印第安纳大学布卢明顿分校信息和计算机学院的博士后研究生。

Giacomo Fiumara (gfiumara@unime.it) 是意大利墨西拿大学数学和信息系的助理教授。

Alessandro Provetti (ale@unime.it) 是意大利墨西拿大学数学和信息系的助理教授。

译文责任编辑：唐杰



技术视角 织物与光之谜舞

Szymon Rusinkiewicz

除了颜色之外，还有哪种性质能让不同的物料呈现出特有的外观？在过去的几十年中，这一基础问题孕育了可真正称作科学的下列计算机图形学分支：研究如何理解复杂材料与光的互相作用。与其他科学一样，这种理解是在结合复杂的数据模型和精细的实验工作后而得到的。

光碰到任何物质时，会被反射、散射或吸收。生成精确和可信的计算机图形效果图时，需要可以重现这种相互作用的算法。在最简单的系统中，光从表面反射的模型远远达不到真实的程度，它生成的所有表面均带有塑料感，这是计算机图形学的一个普遍问题。解决这一问题，尤其是解决织物的问题时，需要在概念上进行一系列的跳跃，也会创造更复杂的模型，使用更复杂的算法。

首先，需要抛弃织物是表面这一概念。虽然我们可能会谈论布料的“表面”，但是在对光进行真实的仿真时，需要把布料当成厚度较小但一定大于零的立体。这一立体由纤维填满，其中所有的纤维都能够散射或吸收光线。因此，我们必须抛弃那种认为光线在各种表面上反射的渲染系统，转向认为光线在立体内部散射的系统。虽然体渲染已被普遍用于为半透明物体（如云、弹珠、甚至人的皮肤）建模，但是直到最近几年，像后文作者这样的研究人员才明白用这种方式为布料建模的重要性。体仿真产生的外观

不仅远看“像真的”，在超近距的特写中也像真的。

一旦我们接受了织物可以用纤维集合建模后，便需要在空间中对所有的纤维进行实际的定位。在有些情况下，纤维是有序的，比如在编织或针织的布。的。多年以来，我们在图形学中唯一的选择是编写程序仿真这些排列，这需要同时微调算法及其庞杂的参数以适应各种新织物的情况。

后文使用了实际的织物样品和微机断层扫描造影，创造性的方式解决了这一问题。自二十世纪九十年代中期以来，在计算机图形学中，有种趋势变得相当重要：使用新的造影和捕获技术【范围涵盖数码相机至三维扫描仪，甚至是工业用MRI(磁共振成像)机器】来获取真实世界中的复杂事物，而这些复杂的

作者给出了一个最新水平的范例，展示了如何利用算法、数学以及测量帮助我们理解织物与光的相互作用。

事物很难用分析性手段或简单的算法进行建模。

最后，我们必须为单独的每条纤维的光散射和吸收方式建模。在此，作者结合了理论和实验。他们提出了与底层物理机制（例如，能量守恒）相符的分析模型，其中包含若干未知的自由参数。接下来，他们对这些参数进行选择，确保效果图（包括完全的光散射体仿真和测定的纤维排列）与实际织物的照片统计数据相匹配。当然，这种过程也支持执行适当的实验步骤，其中可用与某张照片匹配的多个参数生成预测值，然后把这些预测值与另一张照片的情况相比较。

后文的作者给出了一个最新水平的范例，展示了如何利用算法（光散射体仿真算法）、数学（决定立体中单点处发生的散射方式）以及测量（微机CT获取纤维排列）帮助我们理解织物与光的相互作用。本文不仅概括了理解外观的现代方法，而且还介绍了新的理论方法和实验手段，可用于仿真多种不同类型的材料。

Szymon Rusinkiewicz (smr@princeton.edu) 是新泽西州普林斯顿镇普林斯顿大学计算机科学系教授，同时兼任普林斯顿计算机图形学小组成员。

译文责任编辑：周昆

版权归属于作者。

使用微机 CT 造影 构建织物的体外观模型

Shuang Zhao, Wenzel Jakob, Steve Marschner, Kavita Bala

摘要

布是日常生活中不可缺少的物品之一；因此，在过去的几十年中，布的可视化渲染一直是图形学的一个重要研究领域。布的外观繁复多样，其中重要的原因之一是它拥有复杂的三维结构。用于构建这种三维结构的立体算法可以准确地仿真光与布的相互作用，生成极为逼真的布图像。但是，很难创造布的三维模型：为每种类型的物料编写特殊的程序相当繁重，而且很需要编程时间和直觉。不仅如此，由于缺乏重要的视觉特征，比如自然产生的不规则性，生成的模型看起来异乎寻常的“完美”。

本文提出了一个新的方法来获取立体模型。这一方法基于从 X 射线计算机断层摄影 (CT) 扫描中获取的密度数据和从不受控光照下的多张照片中获取的外观数据。为了给物料建模，需要进行 CT 扫描，获取标量密度立体。这种三维数据拥有微米分辨率的布料结构细节信息，但缺乏所有的光学信息。所以，我们把这些密度数据与布料样品的参考照片结合起来，推断出它的光学属性。我们证明了这种方法较容易生成极为精细的体外观模型，而且在较大的尺度上，多种明显不同的织物的（如绸缎和天鹅绒）独特纹理和高光也会自动显现——所有这些仅仅基于获取到准确的中尺度几何数据。

1. 引言

布是日常生活中最基本的物料之一；因此，在过去的几十年中，创造酷似照片的布料效果图一直是计算机图形学中一个活跃的研究领域，其应用领域包括虚拟原型设计、娱乐（电影和游戏）以及零售业。

布的外观繁复多样，其中重要的原因之一是它拥有复杂的三维结构，这造成了复杂的纹理和反射率。不仅如此，这种结构是不规则的，很难建模，但在视觉上却呈现重要的随机性。自二十世纪九十年代开始，研究人员便开始探索体渲染技术。这一技术可以准确地表现此类结构，并能直接模拟光与布之间的相互作用。这一方法解决了常用的基于表面的模型的局限。由于基于表面的模型把布当成厚度无穷小的薄片，产生的视觉效果无法令人满意。^{8, 12, 18} 布料的外观范围很广，但均拥有共同的基本结构，

即由多条又长又亮的纤维组成。布料的这种又厚又带绒毛的本质非常适于构建立体模型。另外，最近的进展⁷提高了体散射的通用性，足以让我们渲染出完全基于物理的体外观模型，可用于布料、毛皮和其他较厚且不像表面的物料。然而，仍然存在根本性的问题：创建这些体模型本身。先前的工作主要依赖于程序方法（专用程序）构建这些立体模型，但是这种方法的通用性有限：对于每种新物料，需要投入相当大的创造性工作来设计这些程序。不仅如此，由于缺乏真实布匹呈现出的细微的不规则性，生成的模型看起来异乎寻常的“完美”。

本文探讨了一个完全不同的方法，这个方法用于构建体外观模型，并特别关注了布料。因为很难为布料的详细几何结构构建出良好的模型，我们使用了立体成像来直接测量结构，然后在使用参考照片填充光学属性。第二步是通过解一个统计地对照片和基于物理的渲染结果进行匹配的逆问题来实现的。

业界开发了很多立体成像技术，包括计算机断层摄影 (CT)、磁共振和超声。但与照片不同，得到的数据与物料的光学外观并无直接关系：只于它的结构有关。所以，这些图片的体渲染可用于说明隐藏的内部几何结构，但却无法直接用于渲染出真实感图片。例如，对编织的棉布进行微机 CT 扫描后，可得出交错的纱线及其组成纤维的详细视图，精确展现各条纤维的方向和各条纱线的位置，但是却没有它们与光的相互作用的信息：无法确定织物的颜色是黑色，或白色，或是中间的任何颜色。

在本文中，我们证明，只要极少的其他信息便可把 CT 数据扩展成为真实的外观模型。在渲染特写视图时，了解三维结构的价值显而易见，因为这些细节是直接可见的。但同等重要的是，物料中纤维的形状和排列方式也决定了材料的总体外观（镜面高光形状和性质）以及在不同光照和视野下可视纹理的变化。结合正确的渲染技术后，简单的纤维局

本文最初发表于 *ACM Trans. Graph.* 30, 4 (2011)。

部反射模型能自动预测明显不同的各类物料（如天鹅绒和绸缎）的特有外观，而且只要了解材料物的三维结构便可实现。

本文的贡献在于其说明了如何结合并利用物料照片的外观信息来丰富通过对少量织物样本进行 CT 扫描而获取的结构信息，进而构建可信的、一致的光学属性；在使用基于物理的体渲染器进行渲染时，这种体外观模型可以生成逼真的外观。本文从头至尾描述了我们的体外观建模过程，并通过为外观明显不同的布料构建模型进行了展示。使用的布料范围广泛，色泽从无光泽到光亮，质地从粗糙到光滑，我们的方法均能获取它们特有的高光、纹理和绒毛感。

2. 相关研究

表面外观建模、布反射率建模和布结构建模存在多个相关领域，我们对其中的真实感体渲染和建模研究进行了归类。

外观建模。因为针对标准表面的模型不足以处理复杂的厚料，研究人员和业内人士转而依靠基于图像的渲染方法，比如双向纹理函数（BTF），它基本上包含了在所有可能的光照和视角方向下该表面照片的详尽集合。^{4,5}对于用其他方式很难处理的材料来说，虽然 BTF 可以创造出逼真的结果，但是基于图像的方法需要大量的存储空间，其分辨率也往往不足以获取高光光泽度，而且通常也无法获取或预测投射角，这样便无法获得逼真的轮廓和边缘。

早期的体外观模型有两个重要的例子，一个是 Kajiya 和 Kay⁸ 的毛皮渲染，一个是 Perlin 和 Hoffert¹² 的“超纹理”。虽然此后使用离散的曲线来渲染头发和毛皮变得更为常见，但是上述模型的结果证明了把立体模型应用于复杂且几乎不可清晰分辨的细节时所产生的价值。Lumislice 阐述^{3,18}了一种类似的方法，专注于构建针织品模型并进行渲染。Magda 和 Kriegman¹¹则描述了一种获取立体纹理的方法，其结合了立体法向量场、局部的反射率函数和占位信息。

所有这些方法均需要大量的建模工作。最近，Jakob 等人⁷介绍了一个原理性的公式来渲染各向异性的，有方向的立体媒质，为更加基于物理的体外观模型开创了更多的可能性。

布的反射率模型。长期以来，布料一直都是图形学中复杂光学行为的源头之一。Westin 等人¹⁷通过射线追踪中构造模型构建了布匹的反射率特征，而我们的系统中布匹高光的出现方式与其存在着某种关联。Ashikhmin 等人²则使用人工设计的微面元分布渲染了天鹅绒和绸缎。Adabala 等人¹提出了一种基于微面元模型的编织布渲染方法，Irawan 和 Marschner⁶则基于对各种织物的纤维切向分析阐述了一种精确的模型，并使用测出的反射率进行了验证。相对于当时最先进的技术而言，上述方法中的每一种方法都取得了相当好的外观，但它们都是人工设计的模型，为单独的物料或特定种类的物料特别开发。

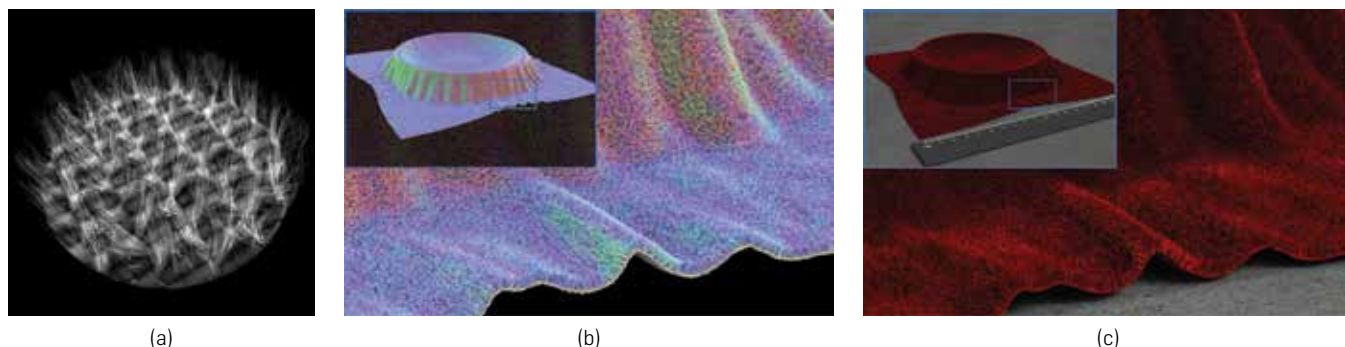
由于我们的方法基于完全通用的系统，其中只用了充满纤维的立体作为底层的假设，所以我们能处理各种织物或类似织物的物料，几乎没有原则性的限制。不仅如此，通过导入现实世界中的立体细节，在特写，轮廓，边缘和转角中，我们均可取得相当好的外观，而在相同的情况下，表面的模型看起来异常的光滑和平坦。

布料结构。针对布料几何结构的研究已进行了数十年之久。^{9,13}最近，在几个应用中，研究人员已经使用了 X-射线体层摄影^{10,15}来仔细观察织物的结构。¹⁶X-射线体层摄影使用了同步加速器装置或者快速改进的微机 CT 扫描仪。这些研究专注于抽取与物料的机械性质有关的几何信息，我们用到了其创造的某些分析工具¹⁵。

3. 概述

我们系统的目标是创造布料的真实感体外观模型。我们需要生成三维立体的样本来描述物料在各体元处的光学属性。随后，当我们使用基于物理的渲染

图 1. 我们使用 CT 成像构建了复杂物料（如天鹅绒）的体外观模型：（a）CT 数据给出了小体积的标量密度；（b）我们抽取了纤维方向（用假彩色表示）并平铺出较大的表面；（c）我们匹配外观参数和照片，以创造出完整的外观模型。重现了天鹅绒的精美细节和独特高光。



系统进行渲染时，系统可以真实地重现真实布料的外观（图 1）。

因为布料由纤维组成，所以我们需要可以处理纤维各向异性的体散射模型；为了实现这一目的，我们选择了 Jakob 等人⁷提出的修订模型（详见第 4 部分）。该模型需要光密度、反照率和两个相位函数参数：方向向量和镜面叶瓣的宽度。直观地说，光密度描述了光在布料内部散射的频率；反照率和相位函数则分别说明了光被吸收的部分以及各散射处光方向的改变情况。

我们的技术首先用微机 CT 扫描一小部分物料，展现了在一平方厘米范围内单独纤维层级的详细情况。只要花费适中的费用（几百美元），便可立即从若干设施处获取这种扫描结果，而且业界也正在研究推出合适的桌面 CT 扫描仪。在按顺序执行的三个阶段中（图 2），我们处理并丰富了这些数据，最后得到了定义所需散射模型参数的一个立体。这个立体使用了从 CT 数据中推导得出的密度和方向场，再加上三个全局参数：反照率、叶瓣宽度和缩放密度场所需的密度乘数。

第一个阶段（第 5 部分）处理了密度体，用取向信息加以丰富，并通过三维取向的纤维对数据进行卷积来探测取向结构，再通过设定阈值分离有用的结果与噪声，进而去除噪声。这一阶段生成了密度和方向场。

只有在确定了全局的光学参数后，才能渲染出这一立体。第二个阶段（第 6 部分）使用了在已知光照（但非受控）条件下物料的一张照片，然后通过匹配已渲染的立体的纹理与照片中的纹理把光学属性与第一个阶段中获得的取向立体关联起来。

得出的立体模型在渲染小样本时效果很好；第三个阶段使用了这一小块布料，然后把它映射到大面积的布料表面上，其中使用了随机平铺来复制布料，使用了壳状映射¹⁴来扭曲布料。

由此得出的效果图（第 7 部分）说明，在直接利用了中尺度的几何信息后，这种独特的外观建模方法可以生成极好的各种外观，范围从几何信息本身可见的小尺度一直延伸到有向散射属性从测定三维结构中

自然产生的大尺度。对于仿真难度较高的物料，比如天鹅绒和绸缎，即使我们没有使用可以区分它们的光散射测量值，使用我们相当微小的立体散射模型后，仍可预测它们的独特外观，因为我们拥有准确的几何信息。

4. 纤维散射模型

我们使用 Jakob 等人⁷引入的各向异性的辐射传递方程（RTE）为光线传输建模，其中说明，在参与的媒介内，

$$(\boldsymbol{\omega} \cdot \nabla)L(\boldsymbol{\omega}) + \sigma_t(\boldsymbol{\omega})L(\boldsymbol{\omega}) = \sigma_s(\boldsymbol{\omega}) \int_{S^2} f_p(\boldsymbol{\omega}' \rightarrow \boldsymbol{\omega})L(\boldsymbol{\omega}')d\boldsymbol{\omega}'$$

其中 σ_s 和 $\sigma_t : S^2 \rightarrow \mathbb{R}$ 是各向异性散射和消光系数，而 f_p 为相位函数。为了便于理解，上式省略了空间依赖性。

上式可理解为是对通用化的各向同性 RTE，其增加了对各种取向不同的媒介“影响”的支持。例如，由于在前后一致对准的纤维表面上，平行照射的光线比垂直照射的光线碰到的阻碍小，需要使用 $\sigma_t(\boldsymbol{\omega})$ 的方向依赖性来为这种效果建模。

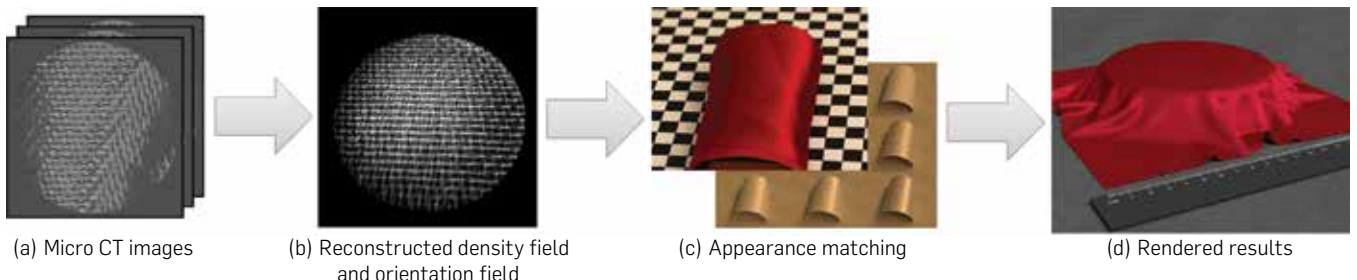
为了确定需要解决的问题，我们必须选择一个兼容的散射模型，可提供内部一致的 σ_t 、 σ_s 和 f_p 定义。为了达到这种目的，我们使用了在同一研究中提出的微片模型。这是类似与微面元模型的立体模型，代表了不同种类的立体散射互动，使用了用于描述空间各点处（未解析）理想镜片的定向薄片分布 $D(\mathbf{m})$ 。与微面元模型相似，随后的相位功能需要求出位于入射方向和出射方向之间的中间方向处的 $D(\mathbf{m})$ 。为了完整起见，我们重新把模型的定义描述如下：

$$\sigma_t(\boldsymbol{\omega}) = a\rho \int_{S^2} |\boldsymbol{\omega} \cdot \mathbf{m}| D(\mathbf{m}) d\mathbf{m},$$

$$\sigma_s(\boldsymbol{\omega}) = \alpha\sigma_t(\boldsymbol{\omega}),$$

$$f_p(\boldsymbol{\omega}' \rightarrow \boldsymbol{\omega}) = \frac{a\rho\alpha}{4\sigma_s(\boldsymbol{\omega})} (D(h(\boldsymbol{\omega}, -\boldsymbol{\omega}')) + D(-h(\boldsymbol{\omega}, -\boldsymbol{\omega}')))$$

图 2. 我们的体外观建模过程：(a) 获取 CT 图像；(b) 创建密度场和方向场；以及 (c) 通过匹配照片与渲染得出的图像的统计数据，为立体模型的光学参数赋值。(d) 使用我们所获取的体外观和几何模型渲染出较大的模型。



上式中, ρ 指颗粒密度, a 是单一薄片的面积, α 是关联的反照率, $h(\omega, \omega') := (\omega + \omega') / (\omega + \omega')$ 。注意, 上面各式已经进行了简化, 其中假设薄片的反照率与散射角无关。这不仅大大减少了我们的搜索空间, 而且仍能让我们得出一个可描述与各纤维材料之间的散射互动的模型, 效果也相当好。

4.1. 薄片分布

在捕捉 Jakob 等人⁷提出的同一种关键特征时, 我们提出了较易整合的薄片分布。我们使用了下列密度函数, 其规定了以大圆为中心的截断高斯, 该圆垂直于局部纤维取向 ω :-

$$D(\omega) = \frac{1}{(2\pi)^{3/2} \gamma \operatorname{erf}\left(\frac{1}{\sqrt{2}\gamma}\right)} \exp\left(-\frac{(\bar{\omega} \cdot \omega)^2}{2\gamma^2}\right)$$

式中, 标准差 γ 确定了纤维的粗糙度。更准确的来说, 渲染所需的参数为

- ω -, 局部纤维取向;
- γ , 薄片分布的标准差;
- α , 薄片的单次散射反照率,
- a 和 ρ 分别为微片的面积和密度。它们的乘积大致对应于传统各向同性立体渲染中的互动系数 σ_o 。因此, 我们把它们设为处理后的CT密度的倍数, 也就说 $ap(\mathbf{x}) := d \cdot \text{CT}(\mathbf{x})$, 其中 d 为比例常数。

第 5 部分讨论了获取 $\text{CT}(\mathbf{x})$ 和 $\omega(\mathbf{x})$ 时的步骤; 在第 6 部分, 我们阐述了找出 α , γ 和 d 的方法。

5. CT 图像处理

微机 CT (计算机断层摄影) 设备使用 X 射线 CT 方法来探查小型至微观结构, 而且越多越多的人能够使用它。这种成像模式适于很多种物料, 可从中抽取其中小样品进行扫描。

在本部分中, 我们描绘了从 CT 密度体中抽取纤维方向的过程, 其中使用了特殊的纤维检测过滤器。随后, 我们解释了获取适于渲染的取向和密度场时遵循的处理步骤。

5.1. 重建方向场

CT 图片提供了体素化的密度场, 但无方向信息。因为在我们的光学模型中, 相位函数需要方向, 所以需要为每个非空体元重建方向。我们的方法使用取向纤维来检测纤维, 基于 Shinohara 等人¹⁵使用的类似纤维定位 CT 数据中的各条纤维。

为了检测纤维在位置 \mathbf{p} 处的方向 \mathbf{d} , Shinohara 提出了沿轴 \mathbf{d} 的方向圆柱对称的纤维, 它由距轴距离不同的高斯组成: $q(\mathbf{d}; \mathbf{p}) := -2 \exp(-ur^2) + \exp(-wr^2)$,

其中 $r = \|\mathbf{p} - (\mathbf{p} \cdot \mathbf{d})\mathbf{d}\|$, 为距纤维轴心的距离, 参数 u 和 w (通常 $u < w$) 基于样本中的纤维尺寸, 根据经验进行调节 (见图 3)。

我们为原始的 CT 立体设定了阈值 ϵ_d , 由此得出了二值立体 b : 对于任意的 \mathbf{x} , 如果 $\text{raw}(\mathbf{x}) \geq \epsilon_d$, 则把 $b(\mathbf{x})$ 设为 0, 否则设为 1。然后, 对于每个固定的方向集, b 与过滤器 q 进行卷积: $J(\mathbf{x}, \mathbf{d}) := \sum_{\mathbf{p} \in V} q(\mathbf{d}; \mathbf{p}) b(\mathbf{x} + \mathbf{p})$, 其中 V 为边缘长度 h 的材积。

如图 3 所示, 当 \mathbf{d} 等于纤维的方向时, 函数 J 达到了最大值。所以, 通过为每个体元 \mathbf{x} 找出让 $J(\mathbf{x}, \mathbf{d})$ 最大化的 \mathbf{d}' , 并设 $\omega(\mathbf{x}) = \mathbf{d}'$, 计算出方向场。在我们的实现中, 我们从 $32 \times 32 \times 6$ 立方图中选出的方向 $\{\mathbf{d}_j\}$ 集合, 然后预先计算出上面的 q 。然后, 对于每个非空体元 \mathbf{x} , 我们设 $\omega(\mathbf{x}) = \mathbf{d}_j$, 其中 $j = \arg \max_j J(\mathbf{x}, \mathbf{d}_j)$ 。

5.2. 对 CT 图像进行去噪

CT 图像, 特别是像我们的布料样本那样的低密度物料的图像, 通常都包含大量的噪声。所以去除噪声是获取优质数据用于渲染的关键。因为布料结构总是有取向的, 而且噪声的各向同性通常比较明显, 所以 J 的值在去除噪声时相当有用。

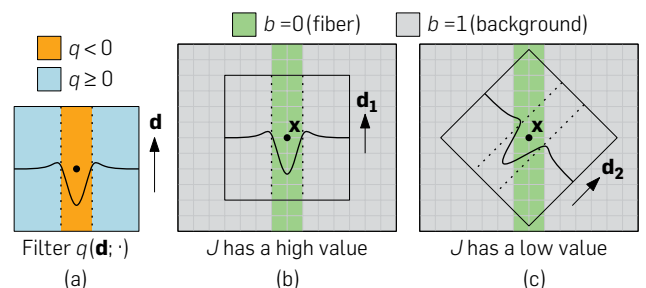
在我们的系统中, 我们使用了两个阈值去噪。第一个阈值 ϵ_d 是体元值本身的阈值, 用于去除微弱的背景噪声, 否则背景噪声会让模型变得模糊。设定这种阈值后, 会生成二值立体 b 。第二个阈值 ϵ_j 为 J 值的阈值, 用于去除各向同性的噪声。由于这种噪声的密度值太高, 无法用第一种阈值去除。我们设:

$$\text{CT}(\mathbf{x}) := \begin{cases} \text{CT}_{\text{raw}}(\mathbf{x}) & \text{CT}_{\text{raw}}(\mathbf{x}) \geq \epsilon_d \wedge J(\mathbf{x}, \bar{\omega}(\mathbf{x})) \geq \epsilon_j; \\ 0 & \text{otherwise.} \end{cases}$$

5.3. 数据复制

由于我们的样本非常小, 所以需要为渲染复制立体数据。我们研究了 Zhao 等人¹⁹提出的基于样例的综合法。Zhao 提供了复杂的工具来实现这点, 但本文却使用了截然不同的方法。本文中我们考虑了两种简单的随机平铺方法, 用从我们的模型中抽取出的多片立体数据铺在各个表面上, 而不引入常规的转向结构。在这两

图 3. 计算二维中的函数 J : (a) 过滤器形状 q ; (b) 当 q 与纤维对齐时; (c) 当 q 未对齐时。



种方法中，我们简单地用从立体中复制的矩形片阵列铺在表面上，而不考虑片边缘的连续性。

对于规律不可见的物料而言（如天鹅绒和毛毡），我们从以立体为中心的矩形区域中复制出表面上的各片。为了提供变化的局部结构，对于每一块片，它的源矩形会按照不同的随机角度旋转。对于带有编织结构的物料，比如丝绸和华达呢，我们使用了类似的方法，但是使用了源片的随机平移，而不是旋转。各样品的编织花样通过人工识别，同时标出一块包含整数次重复的矩形区域。接下来，从子矩形中挑选出各块（更小的）表面片，其中包含了与编织花样匹配的部分。结果可得一种平铺方式，可以重现正确的编织花样，又避免了明显重复的纹理。然后，我们使用壳状映射把这种平铺后的数据映射到任意的表面上。¹⁴

6. 外观匹配

CT 数据经处理后，会为立体产生空间上不同的密度和方向。但是，模型的光学外观参数仍需确定。因为 CT 扫描并未给我们提供物料的光学属性，我们使用了物料的照片来计算外观参数。

为了让问题可控，我们假设该立体包含相同的材料，只有密度和方向不同。这一假设适于使用单一类型的纤维制造的织物，也包含了很多重要的样例。未来，我们会研究由不同材料的纱线编织的织物。所以，外观参数在整个立体范围内一模一样，而我们必须确定这些参数。这些参数包括：薄片分布 γ 的标准差（与纤维的粗糙度对应），散射反照率 α （与物料的颜色对应）和密度比例 d （与不透明度对应）。图 4 说明了这些参数的效果。

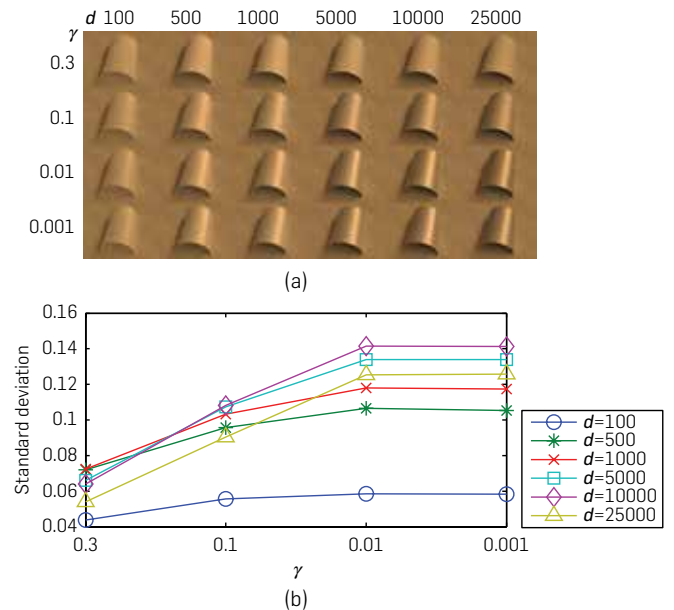
为了匹配物料的光学属性，我们必须使用样品的照片。一个方法是使用扫描的样品的照片，根据扫描情况对相机进行校准，然后关联图像中的像素与立体中的射线。这种校准和采集的工作相当繁重；高分辨率的扫描结果在实际使用时会造成很多困难。另外，我们发现，确定我们寻找的少量参数值时，不需要这种粒度的细节。与此相反，我们假设，不同的小片上的织物在统计上是相似的。因此，我们的方法是从统计学上把渲染得出的图像纹理与相同物料的不同部分的照片进行匹配。照片是在不受控但已知的光照条件下拍摄的。

接下来我们将描述我们对光学参数与照片进行匹配时所使用的度量，然后我们会讨论我们的匹配算法。

6.1. 匹配使用的度量

外观匹配的过程并不是直接把照片的颜色映射到立体上，因为立体模型描述的是局部的散射属性，而外观由全局立体多重散射过程决定。我们的方法是使用基于物理的渲染器反复渲染立体，并调节光学参数，以

图 4.(a)使用绸缎立体平铺得出的圆柱体效果图，其中使用了固定的反照率，变化的叶瓣宽度 γ 和密度乘数 d 。(b) 绸缎样品的对应像素值的标准差：较尖的叶瓣提供了更亮的外观，但会引起更大的标准差。 d 的作用更为复杂。



便让所渲染的图像的某些纹理统计数据与照片的统计数据相匹配。

我们匹配了两种简单的统计量：像素值均值和像素值的标准差，它们根据照片的对应区域以及大致相同的几何形状的渲染结果计算得出。这种方法可有效地对匹配区域内的图像亮度和纹理对比度进行了匹配。我们也尝试了其他的方法，但是发现均值和标准差的度量更简单，也相当鲁棒。因此，从照片流入立体模型的唯一信息是单个矩形中像素的均值和标准差。

外观匹配模型涉及选择几何形状、照相机位置、光照和匹配区域。这些步骤本质上是手动选择的，同时我们采用了选择设置参数的原则，通过设置展现布料外观独特特征。例如，我们确定了一种配置方式，使用后可以在绸缎上出现高光。除此之外，我们没有采取特别的措施来安排匹配外观时使用的输入，而且得出的结果似乎对细节不敏感。

6.2. 优化过程

如图 4 所示，对于这两种测量值，密度乘数都起着相当复杂的作用。既然我们的前向过程（本质上是蒙特卡罗路径跟踪）相当繁重，我们选择通过渲染此类矩阵在我们的实现中预先确定密度乘数。解决密度乘数后，简化了逆问题并引出了可行的解决方案。我们发现，算法对密度乘数的选择并不是特别敏感；我们的结果使用了两种主要的设置，它们相差了十倍（见表 1）。

使用固定的密度乘数后，我们使用迭代算法求解了反照率的值 (α ，分别估计红、绿、蓝的值)、花瓣宽度 (γ ，单个标量值)。注意，像素值的均值和标准差分别随 α 和 γ 的变化单调变化。因此，可按如下方式使用二分搜索显著提高性能：首先，设定 γ 最初的猜测值，然后搜索 α 来匹配像素值均值。接下来，固定 α ，搜索 γ 来匹配标准差。反复这一迭代过程，直到找出匹配。在实际过程中，这种方法很快收敛，通常只要两到三次迭代。

最后，我们在不同的设置下再拍摄一张照片，然后渲染出对应的图像用作定性验证（见第 7 部分）。图 5 展示了两种不同物料的外观匹配结果。

7. 结果

我们的结果基于真丝缎、天鹅绒、毛毡和羊毛华达呢的样本。我们把它们送到了德克萨斯大学奥斯汀分校的高分辨率 X 射线计算机断层扫描设施进行处理。所有的织物均在 XRadia MicroXCT 扫描器中进行扫描，使用了 1024^3 个立体，体元大小为 $5\mu\text{m}$ ，观测了直径约为 5 mm 的圆形区域。我们的渲染实现基于开源渲染系统 Mitsuba，并对它进行了拓展以处理新的微片

表 1 我们的物料样本的纤维散射模型参数值： d ，密度乘数； γ ，薄片分布的标准差； α ，单次散射反照率

物料	数据大小	d	γ	α
华达呢	$992 \times 1012 \times 181$	5000	0.1	(0.892, 0.063, 0.048)
丝绸	$992 \times 1013 \times 46$	5000	0.01	(0.699, 0.030, 0.080)
天鹅绒	$992 \times 1012 \times 311$	500	0.1	(0.555, 0.040, 0.074)
毛毡	$992 \times 1012 \times 485$	500	0.125	(0.518, 0.915, 0.365)

分布（第 4 部分）。所有的渲染结果均使用蒙特卡罗体路径跟踪生成。

图 6 说明了得出的模型，这些模型使用壳状映射到悬垂布料的几何形状，并在环境光照下进行渲染。对应的高分辨率效果图可以从项目的网站 <http://www.cs.cornell.edu/projects/ctcloth-sg11> 处获取。人们通常使用基于表面的模型渲染织物，把它们当成了二维的薄片，使得剪裁边看起来不真实；另一方面，我们的立体模型清晰地捕捉了三维结构，妥帖地表现出织物的厚度和重量。不仅如此，基于表面的模型轮廓展现光滑的边缘，而我们的模型则可以生成细节丰富的绒毛状轮廓，让织物渲染的真实性达到了一个新的高度。

真丝缎（查米尤斯绸缎）的结构是其表面主要由多条平行的纤维组成，导致了相当强的各向异性的高光。在图 5（1）中，外观匹配对使用了呈圆柱弯曲的物料部分，而且选择的匹配区域还包含了高光区，以便在匹配过程中妥善调整 γ 值。虽然效果图中的理想圆柱体和较平坦的实际物料形状之间存在失配，我们仍然取得了良好的结果，这说明随意的设置也够用。使用从该视图中获取的参数后，验证对说明了旋转 90 度后在同一圆柱体上悬垂的布料的情况。在这种角度时，织物几乎没有呈现任何高光，而我们的模型正确地预测了这种各向异性的外观。

在图 6(a) 的悬垂配置中，我们展现了绸缎。在渲染这些效果图时，没有使用无反射率的模型（如 BTf）或其他的多视图图像数据——在渲染模型时，立体中的方向信息能让该织物的特有外观自动出现。

对于华达呢（一种羊毛斜纹织物），光照方向引起的纹理变化是一种重要的外观特性。在图 5(2) 中，外观匹配对通过低频环境光照亮。验证对准确地预测

图 5. 外观匹配结果：（顶部）丝绸，（底部）华达呢。（a）和（c）列展示了物料的照片，（b）和（d）列展示了渲染后的图像。左边两列构成了外观匹配对，其中蓝色方框指为执行我们的匹配算法而人工选择的区域。右边两列指验证对，在不同的配置下定性验证了我们的匹配情况。

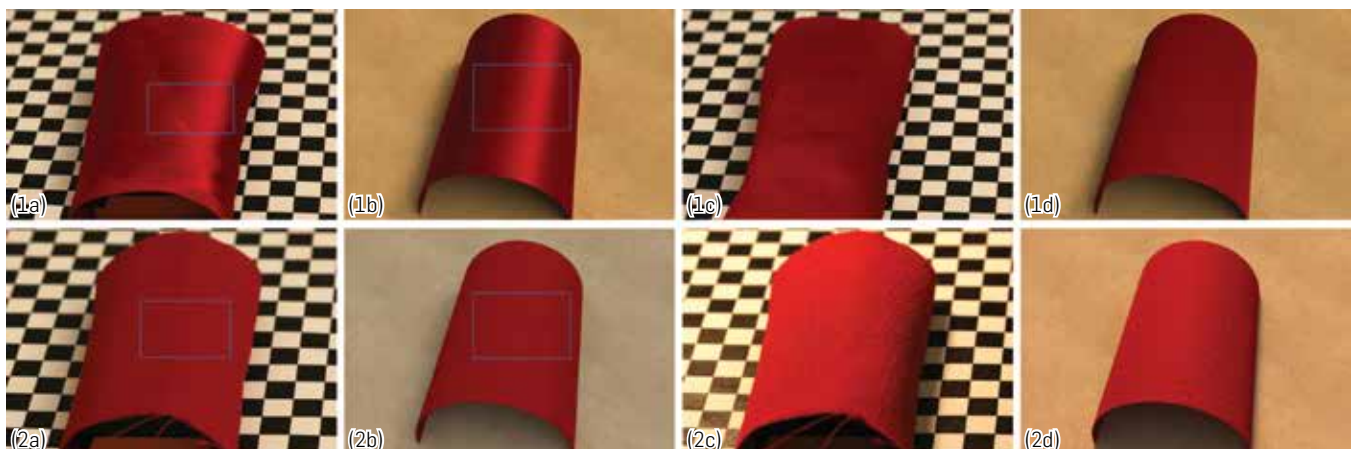
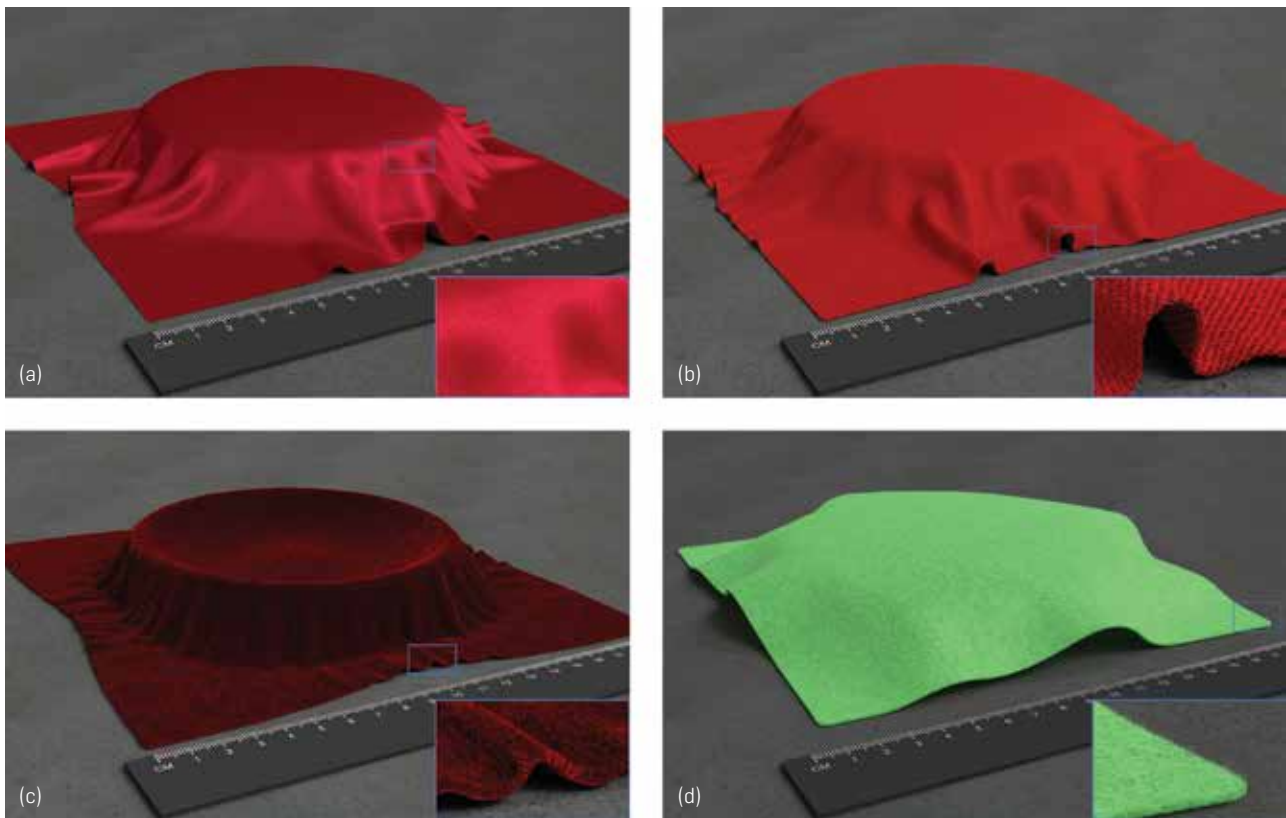


图 6. 悬垂配置下的织物以及我们的体外观模型：(a) 真丝缎，(b) 华达呢，(c) 天鹅绒和 (d) 毛毡。



了不同光照条件下的纹理情况，其中包含了上方有强光源的情况。在图 6 (b) 的悬垂配置中，立体模型捕捉到了微小的投影缩减效应，轮廓的外观，以及遍布表面的细微纹理变化。剪裁边的外观呈现出织物的厚度这种良好效果（与非常薄的缎料相比），而在表面模型中，这一直都是一个难题。

天鹅绒是一种带有割绒的物料（像地毯一样），其可见的表面由从基底材料突出的纤维组成。它的外观非常独特，带有一种特有的入射余角高光。天鹅绒的外观取决于纤维排列的方式，而我们的随机片旋转方法生成了纤维随机排列的天鹅绒。在图 6 (c) 中，我们说明了我们的模型重现天鹅绒高光的方式。另外，边缘和轮廓说明该物料厚度和重量都不小。

毛毡是一种非编织的纺织品，由杂乱缠结的纤维层组成。该物料的厚度和绒毛均是重要的外观属性，通常难以建模和渲染。因为毛毡总体上没有呈现出镜面高光，我们使用了平坦的小块布料进行外观匹配；由于场的深度有限，我们把匹配区域限制在相当薄的矩形范围内，此时照片的聚焦相当好。外观匹配和验证时使用的光照条件与华达呢时的一样。由于自身阴影属性而产生颜色和对比度匹配良好，并能很好地推广到第二种光照条件中。该物料的限制之一是，它的纹理中有相当多的低频内容，而在 CT 造影时我们的

小样本区域没有捕捉到这些信息，这导致我们平铺后的物料外观稍微有些齐整。图 6 (d) 说明了我们的体外观模型捕捉物料厚度和绒毛外观的能力。

与基于图像的方法相比，基于物理的三维模型还支持人们做出更有意义的编辑。在图 7 中，我们使用在空间上变化的反照率值拓展了华达呢的模型。反照率作为方向的函数计算得出，这样便可为经纬线中的纤维分配不同的颜色。使用蓝色的经纱和白色的纬纱，生成了类似斜纹布的织物，只不过用的是羊毛，而不是棉线。

8. 结语

本文阐明了一个新的多模式方法用于创造真实的布料立体模型，其不仅可以捕捉特写效果图中明显的三维结构，还能获取远景图中明显的反射情况。我们的方法与之前使用 BTF 获取布料外观的方法不同。我们的方法明确地为物料的三维结构建模。有趣的是，有了这种结构后，我们的方法能够自动获取物料的定向反射率。

我们的建模方法使用了在测量三维结构方面最强的 CT 成像，而且也使用了在测量颜色和纹理方面最强的照片。通过匹配纹理的统计数据，我们整合了两种信息源，得出了可以同时生成了带有绒毛和纤维的

图 7. 编辑立体表征数据后获取的效果图：华达呢样本使用了蓝色调进行渲染（左）；然后我们根据纤维的方向探测纬线，把纬线着色为白色，这样便生成了类似斜纹布的物料（右）。



丰富细节的特写视图和这些物料的独特高光。在渲染我们测量出的这种结构时，这些高光会自然产生。我们没有测量任何反射率的值，也只调测了光学模型中的几个参数。通过简单的各向异性相位函数模型以及从立体中抽取出的遮挡和方向信息，我们创造了布料的外观。本文证明，由于纹理的复杂外观是由几何结构创造的，一旦我们得到了这种结构，我们就大致接近了为外观建模的目标。

虽然在利用该结构为物料外观建模的过程中有很多复杂的因素，但是在为外观建模时，它仍是一种非常实用的方法。为物料建模时，只需要 CT 扫描的结果以及已知光照情况下的几张照片。CT 扫描可以从若干设施处获取，价格合理（或者，将来能通过快速改进的桌面 CT 扫描技术获取），而照片只需要使用照相机和球面镜，花上几分钟时间。另外，还可以用 CT 扫描一些基础的编织结构样本，然后把得出的立体组合起来形成很多花式各异的织物，¹⁹ 从而进一步减少每次设计的成本。得出的模型本质上是立体的，基于物理的，使得它们比基于图像的数据更容易编辑。通过缩放立体几何结构的参数，调整颜色、光泽度、不透明度和物料厚度相当容易；而且通过编辑立体数据，还可以让物料的结构产生很多更加根本性的变化。

本文证明了将 CT 建模方法用于织物建模时的有效性，但是这一方法也确实存在一些局限。特别是，这一方法要求光学属性的变化与密度的变化相关，而这一要求会限制这种成像模式可以捕捉的物料类型。另外，扫描仪只能用于为长宽高均小于一厘米的小样本成像，所需的分辨率需要达到能生成清晰的纤维取向图的程度。无法完全纳入立体的厚料（例如，带有非常长的飘动纤维的物料）不能被很好的处理。一些不常见的物料（比如金属纤维）在 CT 时可能会有问题，因为动态范围有限。而且，无法捕捉较大尺度的纹理内容。CT 扫描仪的分辨率和动态范围改进后，会减少这些问题。CT 非常适于织物，但尚不知道还有哪些其他物料适于用 CT 处理，也不知道在该技术中使用其他立体成像方法的情况。另外，我们的方法现在无法捕捉由不同颜色的纱线编织的物料。

未来的研究领域还很宽。本研究使用了极小的样本。在 CT 技术改进以及价格降低后，应该可以使用更大的样本，生成更好的纹理。为了拓宽可处理的物料的范围，需要使用能识别并把多个物料纳入单一立体的新的参数估算方法。为了提高准确性，可以在不同条件下拍摄的更多的照片，纳入更多的参数（比如，更复杂的相位函数）进行匹配。最后，对于那些很难用表面模型捕捉的多种类型的物料，可以拓展这种方法进行研究。

参考资料

1. Adabala, N., Magnenat-Thalmann, N., Fei, G. Visualization of woven cloth. In *14th Eurographics Workshop on Rendering* (2003), 180-185.
2. Ashikhmin, M., Premoze, S., Shirley, P. S. A microfacet-based BRDF generator. In *Proceedings of ACM SIGGRAPH 2000* (2000), 65-74.
3. Chen, Y., an Hua Zhong, S.L., Xu, Y.Q., Guo, B., Shum, H.Y. Realistic rendering and animation of knitwear. *IEEE Trans. Visual. Comput. Graph.* 9, 1 (2003), 43-55.
4. Dana, K.J., van Ginneken, B., Nayar, S. K., Koenderink, J.J. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.* 18, 1 (1999), 1-34.
5. Furukawa, R., Kawasaki, H., Ikeuchi, K., Sakauchi, M. Appearance based object modeling using texture database: Acquisition, compression and rendering. In *Eurographics Workshop on Rendering* (2002), 257-266.
6. Irawan, P., Marschner, S. Specular reflection from woven cloth. *ACM Trans. Graph.* 31, 1 (2012), 11:1-11:20.
7. Jakob, W., Arbre, A., Moon, J.T., Bala, K., Marschner, S. A radiative transfer framework for rendering materials with anisotropic structure. *ACM Trans. Graph.* 29, 4 (2010), 53:1-53:13.
8. Kajiya, J.T., Kay, T.L. Rendering fur with three dimensional textures. *SIGGRAPH Comput. Graph.* 23, 3 (1989), 271-280.
9. Kawabata, S., Niwa, M., Kawai, H. The finite deformation theory of plain weave fabrics. Part I: The biaxial deformation theory. *J. Textile Instit.* 64, 1 (1973), 21-46.
10. Lomov, S., Parnas, R., Ghosh, S.B., Verpoest, I., Nakai, A. Experimental and theoretical characterization of the geometry of two-dimensional braided fabrics. *Textile Res. J.* 72, 8 (2002), 706-712.
11. Magda, S., Kriegman, D. Reconstruction of volumetric surface textures for real-time rendering. In *Proceedings of the Eurographics Symposium on Rendering (EGSR)* (2006), 19-29.
12. Perlin, K., Hoeffert, E.M. Hypertexture. *SIGGRAPH Comput. Graph.* 23, 3 (1989), 253-262.
13. Pierce, F.T. The geometry of cloth structure. *J. Textile Instit.* 28, 3 (1937), 45-96.
14. Porumbescu, S., Budge, B., Feng, L., Joy, K. Shell maps. *ACM Trans. Graph.* 24, 3 (2005), 626-633.
15. Shinohara, T., Takayama, J., Ohyama, S., Kobayashi, A. Extraction of yarn positional information from a three-dimensional CT image of textile fabric using yarn tracing with a filament model for structure analysis. *Textile Res. J.* 80, 7 (2010), 623-630.
16. Thibault, X., Bloch, J. Structural analysis by X-ray microtomography of a strained nonwoven papermaker felt. *Textile Res. J.* 72, 6 (2002), 480-485.
17. Westin, S.H., Arvo, J.R., Torrance, K. E. Predicting reflectance functions from complex surfaces. *SIGGRAPH Comput. Graph.* 26, 2 (1992), 255-264.
18. Xu, Y.Q., Chen, Y., Lin, S., Zhong, H., Wu, E., Guo, B., Shum, H.Y. Photorealistic rendering of knitwear using the Lumislice. In *Proceedings of ACM SIGGRAPH 2001* (2001), 391-398.
19. Zhao, S., Jakob, W., Marschner, S., Bala, K. Structure-aware synthesis for predictive woven fabric appearance. *ACM Trans. Graph.* 31, 4 (2012), 75:1-75:10.

Shuang Zhao, Wenzel Jakob, Steve Marschner 和 Kavita Bala (szhao, wenzel, srm, kb)@cs.cornell.edu 来自纽约州伊萨卡康奈尔大学。

译文责任编辑：周昆

© 2014 ACM 0001-0782/14/11 \$15.00