

2014

ANNUAL REPORT



# 2014 科研成果年报



中国科学院计算技术研究所





中国科学院计算技术研究所  
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

2014  
科研成果年报

二〇一五年四月  
中国科学院计算技术研究所科研处编

编委主任：孙凝晖

编 委：李锦涛 隋雪青 陈熙霖 程学旗

主 编：罗瑞丽

责任编辑：吴金凤

编 辑：孙 愿 李 丹 梁 丽

通 讯 员：董 慧 何玉晓 李 楠 阮 麟 张 冬 李晓宇

宋艳梅 田卫平 何文静 任 菲

# 计算所简介

中国科学院计算技术研究所(简称计算所)创建于 1956 年，是中国第一个专门从事计算机科学技术综合性研究的学术机构。计算所研制成功了我国第一台通用数字电子计算机，并形成了我国高性能计算机的研发基地，我国首枚通用 CPU 芯片也诞生在这里。

在过去的五十多年里，计算所是我国计算机事业的摇篮。近五年来，计算所的定位是创建学术前瞻的引领产业型战略高技术研究所，保障国家信息空间安全，成为中国计算机产业人才与技术的源头。在计算技术学科的计算机系统、网络、智能技术三个主要研究领域，开展以体系结构与算法为特色的学术研究、技术创新、技术应用与技术转移，重点在龙芯处理器、曙光高性能计算机、天玑网络数据引擎上取得重大突破。截止 2014 年底，计算所获得国家、院、市、部级科技奖励 218 项，其中，国家级科技奖励 43 项（含非第一完成单位 7 项），院、市、部级科技奖励 175 项（含非第一完成单位 12 项）。

计算所坚持“基础性、战略性、前瞻性”的三性原则，坚持新时期办院方针，以“跻身国际前沿，关注国计民生，引领中国信息产业”为己任，落实“创新、求实”的理念，力争成为世界一流的科研学术机构。



# 目录

年度科研概况	1
科研成果与主要进展	4
重点 / 重大科研任务进展	4
面向 LTE-Advanced 的终端软基带技术	4
面向 C-RAN 的低功耗通用处理器平台研发	6
高通量计算系统的构建原理、支撑技术及云服务应用	8
面向服务的未来互联网体系结构与机制研究	11
网络大数据计算的基础理论及其应用研究	12
可重构大内存加速卡	14
新型多核 / 众核处理器编程与运行支撑环境	16
3D 内容视觉获取技术及设备	18
可编程虚拟化路由器关键技术研究及系统研制	20
异构媒体数据的关联与挖掘研究	22
高精度低成本低功耗的海水电导率传感器研究与设计	24
超并行高效能计算机体系结构与设计方法研究	26
云环境下的图像视频群体协同表达与处理	28
在线社会关系网络的挖掘与分析	30
后 IP 网络体系结构及其机理探索	32
数据并行与线程并行合一的可伸缩处理器体系结构	34
动态可配置的压缩感知成像系统	36
理论计算机科学	38
计算机体系结构	40
视觉模式分析与识别	42
多媒体内容分析	44
海云计算系统关键技术与系统研究	46
面向未来网络的可编程网络测试仪表	48
瓦楞纸生产线智能生产管理与工艺支撑系统	50
面向工业产品全生命周期的仿真公共服务平台	52
华为高通量服务器研制	54
研究方向科研进展	56
获奖成果	69
大规模网络视频处理与内容分析关键技术及应用	69
结题验收的代表性成果	70
面向 IMT-Advanced 增强多媒体多播技术	70
图像与视频处理	72
Web 搜索与挖掘的新理论和新方法	74
基于云计算的海量数据挖掘	76
绿色网络 - 降低网络能耗的理论与技术	78
未来异构无线网络信息理论与关键技术研究	80
云计算 I/O 服务器—VIO Server	82
高通量计算关键技术的研究	84
支持未来网络创新的可编程虚拟化路由器平台	86
机器博士 - 智能技术集成平台	88
天玑学术网	90
科研基地进展	94
计算机体系结构国家重点实验室	94
移动计算与新型终端北京市重点实验室	98
中科院智能信息处理重点实验室	101
中科院网络数据科学与技术重点实验室	103
学术活动	105
知识产权	120
学位论文	137

# 年度科研概况

## 一、承担课题概况

2014 年, 计算所共有在研项目 650 项 (包括新增项目 203 项), 其中, 主持或参与国家重大科技专项课题 32 项 (新增 3 项); 主持国家“973”项目 3 项 (新增 1 项), 参与课题 26 项 (新增 5 项); 主持或参与国家“863”课题 31 项 (新增 10 项); 主持或参与国家科技支撑课题 14 项 (新增 1 项); 主持或参与国家自然科学基金重点项目 20 项 (新增 4 项)、面上项目 64 项 (新增 17 项)、国家杰出青年科学基金项目 2 项 (新增 1 项)、优秀青年科学基金项目 4 项、青年基金 77 项 (新增 11 项); 主持或参与中国科学院战略性先导科技专项课题 2 项、院重点部署项目 5 项 (新增 2 项)、国家自然科学基金委和中科院重大科研仪器设备研制项目 5 项 (新增 2 项)、主持或参与国家自然科学基金委和科技部的国际合作项目 5 项 (新增 2 项)、院地合作项目 11 项; 承担横向项目 171 项 (新增 53 项)。

## 二、科教融合中心通过论证

按照中国科学院“四个率先”的要求, 计算所建设科教融合中心并承办中国科学院大学计算机与控制学院, 到 2030 年实现科学卓越和教育

卓越, 计算所的基础前沿研究水平达到国际一流, 计算机学院的教育水平达到国际一流。这项率先行动计划的重大改革举措使得计算所的一个定位有了“源头”, 能够按照“斯坦福模式”把计算所建成世界一流的科研机构。

## 三、基础前沿研究有了引领性工作

计算所的基础前沿研究出现了若干引领性的工作, 标志计算所在细分领域走在了世界最前列。 (1) IPU: 国际上首个深度学习处理器, 获 ASPLOS14 最佳论文 (大陆首次), MICRO14 最佳论文 (47 年来首次美国之外); (2) 消息式内存 (MIMs): 消息式内存专利群被华为公司评为“山顶专利”, 获华为“最佳研究创新成果团队”总裁奖; (3) PARD: 世界上首次提出面向云计算的动态可编程体系结构, 文章被 ASPLOS15 录用, 受邀参加由国际体系结构泰斗 David Patterson 组织的顶尖学者论坛 Dagstuhl Seminar; 发表在 HPCA2010 上的 I/O DMA Cache 技术思想, 在 Intel 最新发布的 Xeon E5 处理器中得到验证; (4) SimICT: 国际上第一个支持千核万线程规模的模拟平台, 速度比学术界主流 gem5 模拟器高 3 个数量级。



基础前沿研究的重要突破包括：（1）突破软件工程领域旗舰会议 ICSE，利用局部时钟重现并行错误的技术应用于两个知名系统 PRES 和 CLAP；（2）面向可穿戴计算的增量智能模型成果发表了 2 篇 UbiComp，基于“爱心小屋”的残疾人远程维权终端系统应用于北京市 13 个郊区县 41 个村；（3）SCN 路由器关键技术与未来网络体系结构的相关成果发表在 SigComm（大陆最近 10 年首篇）、CoNEXT 上，路由器在小规模未来网络试验网及国家相关部门部署应用；（4）5G 绿色通信关键技术成果发表在 IEEE Wireless Magazine（影响因子 6.5）上；（5）频繁场景挖掘算法被 ICDE2015 录用（计算所首篇）；结合深度学习和流形学习的成果，获 2014 年 ACM ICMI 视频表情识别第一名，FG2015 NISTPASC 视频人脸识别竞赛第一名。

#### 四、天玑有望“捅破天”

互联网信息监测系统在国家新闻办获得重点项目支持，建立起系统研发、运行维护、数据服务整体工作流程，成为特定通道信息监测的核心业务系统。大数据引擎系统完成了系统的产品研制，在某部重点项目中得到实际应用，进行生产性运行。

2014 年天玑团队夯实了以 BDC/BDE/BDA 为核心的网络大数据系统技术体系，网络数据感知获取系统实现百万量级信源频道的分布式采集，大数据引擎系统产品定型，高吞吐性能与企业同类系统相比优势明显，有望取得“捅破天”的应用效果。

#### 五、pFind3.0 迎接蛋白质组时代

pFind 团队历经 5 年半研制完成定量分析软

件 pQuant，相关论文发表在分析化学主流期刊 Analytical Chemistry，实现了从定性到定量的跨越。交联鉴定软件 pLink 继 2012 年后再次突破 Nature Methods，支持国际同行在 Nature 发表了文章；pLink 2 的速度提升了数十倍，利用 pQuant 技术完成的合作文章投稿到 PNAS；连续第三次主办了中国计算蛋白质组学研讨会。经过十年如一日的努力，pFind 团队有望在蛋白质组时代占据领先优势。

#### 六、超级基站工程样机研制成功

超级基站的工程样机研制成功，单基站支持 30 万用户并发通信，核心技术在我国第一代 WX 移动通信系统中得到应用。基于超级基站，正在构建仿真与原型验证相融合的 5G 移动通信科研与教学实验平台，能作为空、天、地一体移动通信系统的科研实验装备。DX001-0.5 LTE 基带芯片流片成功，形成了 LTE 小基站解决方案。

#### 七、新增人才与科研奖励

程学旗研究员获得国家基金委杰出青年基金资助，李晓维研究员入选科技北京百名领军人才培养工程，陈云霁研究员荣获 CCF 青年科学家奖，范东睿、孙毅入选院卓越青年科学家。在信息检索方向引入徐君研究员，在高性能计算方向引入尤海航研究员，尤海航研究员获得院百人择优支持，Luis 连续第三次荣获中科院外籍青年科学家项目支持。引入海外博士 6 人。

2014 年获得的科技奖励包括：（1）前瞻实验室“大规模网络视频处理与内容分析关键技术及应用”获北京市科学技术奖一等奖；（2）智能信息处理实验室“视觉模式的局部建模及非线性特征获取理论与方法研究”获北京市科学技术奖三

等奖；（3）网络数据实验室“社会化媒体数据的分析与检索”获钱伟长中文信息处理科技奖一等奖；（4）网络数据实验室“大规模社会化媒体数据的感知与分析处理系统”获中国电子学会科学技术奖二等奖；（5）前瞻实验室在计算复杂性、密码学领域的研究成果获中国密码学会密码创新二等奖；（6）东莞分部国云的“G-Cloud 云计算操作系统 V6.2”获广东省科技进步奖二等奖。此外，我所参与的项目“32 位星载容错控制计算机系统关键技术及应用”、“国家高性能计算应用服务环境关键技术及应用”均获北京市科学技术奖一等奖，“面向智能电网的大数据即时处理和分析技术研究与应用”获北京市科学技术奖二等奖，“普适计算关键技术及支撑平台”获教育部科学技术进步奖一等奖，“基于大数据的互联网机器翻译核心技术研究及产业化”获中国电子学会科学技术奖一等奖。

## 八、产业化工作亮点突出

历时四年，计算所控股的曙光信息产业股份有限公司于 2014 年 11 月 6 日在上海交易所上市，至 12 月 8 日连续 22 个涨停板，被誉为 HPC 第一股，市值超过 150 亿元（至 2015 年 4 月市值已近 300 亿元）。

2014 年 8 月北京市长王安顺到中科晶上公司调研，对团队成果充分肯定，在北京市科委的支持下，中科晶上获得北京市政府基金 4000 万元 A 轮投资。

新成立了北京中科睿芯科技有限公司，从事高密度视频处理加速系统业务；新成立了北京中科力量智能技术有限公司，专注于电信、银行和证券等领域的智能计算应用。此外，与华为公司的第二个 A 类项目“并行编程框架”得以签订。

## 九、分所工作进入新的十年

2014 年新成立了洛阳分所，是计算所在中部地区设立的第一个分所，定位于发展数字媒体技术等下一代信息产业；新成立了福州分所，开展数据相关应用技术研究和产业化，由此计算所基本完成了在沿海经济发达地区的院地合作布局。

苏州分所经过十年的发展，逐步由“平台型”发展到“引领产业型”，在已有的五大服务平台的基础上，为服务器处理器、视频处理芯片、基带芯片三大方向的产业化提供强力支持，带动本地企业的升级。

济宁分所主导建设的中科智造信息产业园，是计算所主动整合资源打造科技产业园区的第一次尝试，以推动济宁市产业结构的转型。

太仓分所利用自主研发的奥视佳摄像运动控制系统为江苏卫视《最强大脑 2》的节目摄制提供支持，实现了拍摄的全自动化、全资讯化、拍摄轨迹提前规划及精确的重复拍摄。

顺德分所为美的公司构建了智慧工厂的整体技术标准，将有效推动美的产业链的整体进步。

## 十、国际合作交流依然活跃

2014 年计算所共出访 224 人次。其中员工出访为 126 人次，占出访总人数的 56%；学生出访为 82 人次，占出访总人数的 37%；赴台 16 人次，占 7%。2014 年计算所引进了美国的 Sally A. McKee 副教授、Jason Oscar Mars 助理教授，他们被纳入中科院“国际人才项目访问学者计划”；西班牙的 Luis HerranzArribas 博士被纳入中科院“国际人才项目博士后计划”。

2014 年计算所主办了亚太计算机系统国际研讨会等 4 个重要国际学术会议，还邀请国内外知名学者到计算所做了 120 场学术报告。



# 科研成果与主要进展

## 重点 / 重大科研任务进展

### 面向 LTE-Advanced 的终端软基带技术

“面向 LTE-Advanced 的终端软基带技术”是计算所牵头的“新一代宽带无线移动通信网”国家科技重大专项课题（课题编号：2013ZX03003014），课题执行时间为2013年6月至2015年12月。

本课题的研究内容是以矢量处理器内核的研究和相关工具链开发为基础，对关键的基带处理算法进行并行化和矢量化的优化，同时研究矢量和标量处理的优化，设计并实现基于软基带思想的多模终端基带处理芯片架构，并通过原型验证平台和工程样片研制对课题研究成果进行验证和测试。

2014 年度，本课题在前续专项课题研发的基础上，取得如下研发成果：

1. 课题组进一步优化了自主矢量处理器的内核结构（图 1），完善了自动化辅助设计工具链：矢量处理器采用标 / 矢量紧耦合的多流水线结构，提高了处理器效率；在灵活性设计方面，

采用多层次化设计的指令集架构，在不同层面满足矢量处理器的高效性和灵活性，以满足 LTE - Advanced 所需的强大计算能力，同时也满足系统升级演进所需的可编程性。课题组设计和实现矢量处理器核及其多核架构的处理能力满足 LTE -

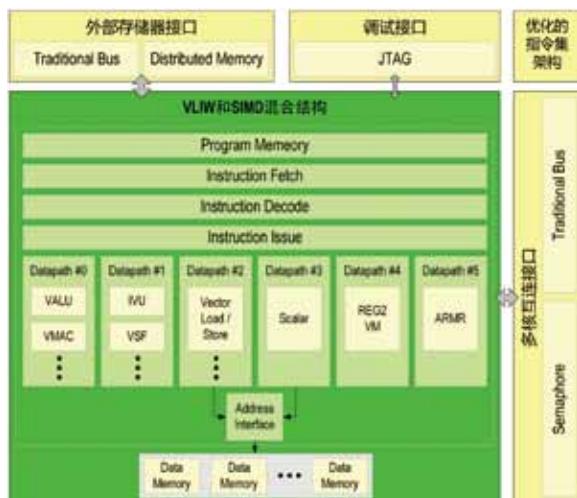


图 1 大规模并行处理的矢量 DSP 内核结构框图

Advanced 下行 600Mbps、上行 300Mbps 的峰值数据处理要求，并能兼顾 2G/3G/LTE 多模灵活性的要求。

2. 基于矢量处理器内核架构，从三个层面实现了 LTE-Advanced 算法参考模型的研发和核心算法的矢量化和并行化优化：① 任务级：多核的并行和流水处理；② 指令级：核内不同数据通路间的并行指令处理；③ 数据级：数据通路内相同运算的矢量化数据处理，使基带芯片对数据流的处理能够在最大程度上实现矢量化和并行化。

3. 设计并实现了以矢量处理器为核心的基带芯片的软硬件架构设计（图 2）。课题在对基带处理算法进行深入分析的基础上，结合矢量处理器内核的特点，并综合运用电路级、结构级和系统级的低功耗技术，设计了一个功耗可伸缩的终端基带处理芯片架构。

4. 完成了支持大规模基带芯片功能验证的 FPGA 原型验证平台的设计：课题组针对 LTE-Advanced 基带芯片设计规模大、验证复杂等问题开发了一套芯片功能验证平台，该平台包括支持 4000 万逻辑门功能验证的 FPGA 验证板及其配套的划片和互连工具、射频子板、协议栈处理板、各种芯片接口子板和测试仪表接口子板等，该平台可连接矢量信号发生器和分析仪，可对基带芯片进行快速的系统级功能验证。

尽管本课题指标要求高、设计难度大，但是课题组组织各单位联合攻关，多项工作齐头并进，在主要研究内容方面突破了处理能力、低功耗、算法软件优化等技术难题，取得了一系列创新性成果，软基带芯片各项指标能够达到课题要求。目前，LTE-Advanced 软基带芯片已进入后端设计阶段，并将于 2015 年 5 月进行 MPW 流片。

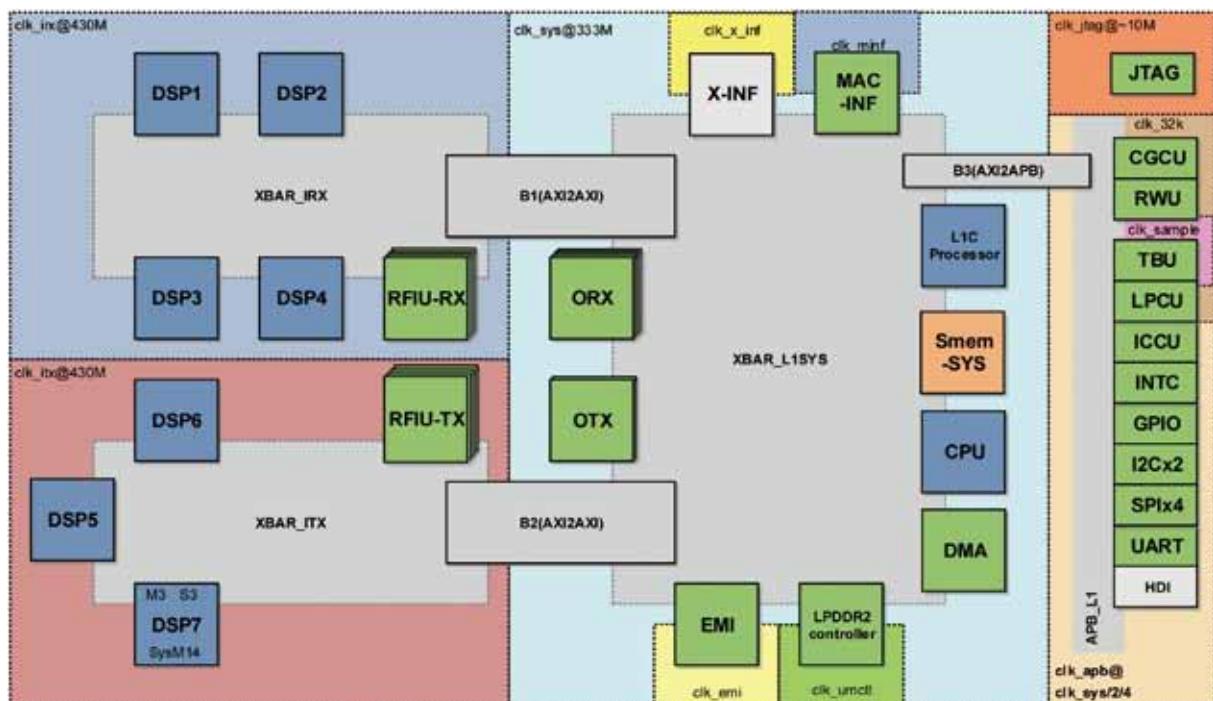


图 2 LTE-A 基带芯片架构



## 重点 / 重大科研任务进展

# 面向 C-RAN 的低功耗通用处理器 平台研发

“面向 C-RAN 的低功耗通用处理器平台研发”是计算所牵头承担的“新一代宽带无线移动通信网”国家科技重大专项课题（课题号：2014ZX03003004），起止时间为 2014 年 1 月至 2015 年 12 月。

本课题目标为面向新型集中式 RAN 架构，研究并实现基于通用处理器平台的 C-RAN 系统基站池关键技术，包括：软基带 DSP 的物理层硬件加速器研制，基于集中式基站池的实时虚拟化技术，载波的动态迁移及扩容，突破目前在通用处理器平台上实现 C-RAN 基站池面临的功耗高等技术瓶颈，提高系统单位功耗载波处理能力，降低系统功耗，从而满足未来异构融合、绿色、智能蜂窝通信网络建设和运营的需求。

2014 年度的主要工作进展如下：

## 1. 超级基站原型研制

本课题面向未来移动通信产业发展需求提出的超级基站设备，如图 1 所示，融合了超级计算技术与无线通信技术，属于移动通信网络架构创新。在超级基站原型机研制方面取得了实质突破，经过软硬件开发及联合测试，目前系统已实现 CS

通话、PS 业务下载功能，且话音质量清晰，业务下载流畅，初步通过了功能性验证，并在专网系统中得到了应用。



图 1 超级基站工程样机系统

## 2. 超级基站实训平台

本年度课题组对超级基站实训平台进行了需求、功能及软硬件架构设计升级。需求方面，确定了实训平台扩展需求；功能方面，扩展了多模通信共平台模式；软硬件架构方面，软件架构进行了信令、业务分离设计，硬件架构方面完善了系统时钟方案，进行了板卡升级提升性能及稳定性。

### 3. 超级基站资源智能管控

基于超级基站架构,课题组针对潮汐效应带来的负载差异性和负载动态变化,以及资源在虚拟映射时的资源利用率等问题,进行了如下研究:

(1) 提出一种和网络负载变化相适应的处理资源分配方法,研究了负载差异性条件下的集中式最优资源映射机制;

(2) 提出了多协议处理实体在同一个处理器上执行时的协议组织和运行机制,通过动态地对下行数据处理进行提前,在不降低处理资源利用率的前提下,保证协议处理的实时性;

(3) 研究了小区休眠技术在超级基站中的实现方法,如图 2 所示,超级基站可以通过集中式管控单元(OMU)对高虚拟化程度的处理资源和 RRH 进行动态映射和开关,同时,对比传统分

布式蜂窝网络架构下的 traffic-aware 休眠机制,提出了 centralized-best 休眠机制,不仅能够通过优化算法得到全局最优部署方案,更通过对业务的预测和休眠模型的建立,降低了休眠检测与决策的系统开销,减少了频繁切换带来的问题,提高了整体系统的稳定性。

此外,基于超级基站领域的研究,在北京市自然基金委的支持下,实验室主办了《通信与计算机科学技术融合》2014 年度国际学术研讨会。李国杰、邬贺铨、怀进鹏、陈俊亮、李幼平、戴浩等多位信息领域的院士参加了研讨会。与会专家认为,利用计算技术解决大数据时代的高通量、高性能宽带信息通信传输是一种必然的趋势,应加快推进通信科学与计算机科学融合的理论基础研究及实验平台建设。

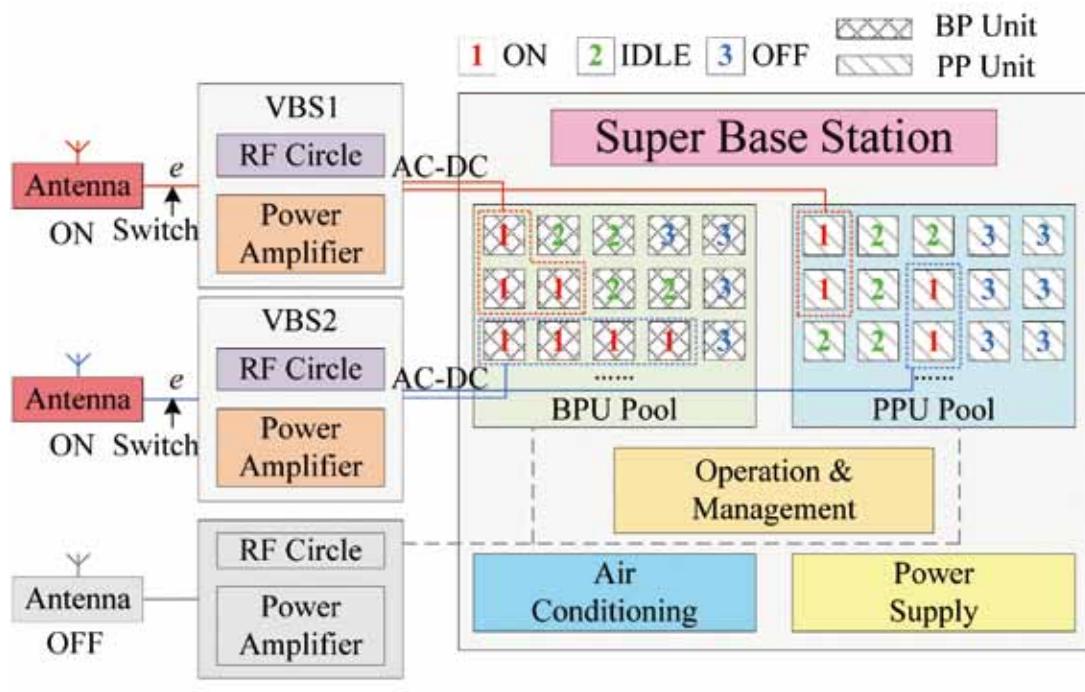


图 2 超级基站休眠机制示意图



## 重点 / 重大科研任务进展

# 高通量计算系统的构建原理、支撑技术及云服务应用

“高通量计算系统的构建原理、支撑技术及云服务应用”是计算所牵头的国家973计划项目（项目编号：2011CB302500），起止时间为2011年1月至2015年8月。2014年度，在高通量计算系统的微结构设计、系统结构设计、可靠设计、云计算服务与应用等方面进行了探索研究，取得了一系列成果，超额完成了预期目标。本年度取得了如下突出进展。

在2014年度，本项目在核心科学问题的研究方面完成了预期目标。共发表论文82篇，其中，SCI收录37篇，包括IEEE Trans. on VLSI Systems(TVLSI), IEEE Trans. on Architecture and Code Optimization(TACO), Communication of the ACM、IEEE Micro、IEEE Transaction on Computers, IEEE Transactions on Mobile Computing、Elsevier Journal of Network and Computer Application, Elsevier Computer Networks, Wiley International Journal of Communication Systems, Oxford Computer Journal等国际期刊和ISCA、ASE、HPCA、DASFFA、CGO、PACT、DAC、DSN、Hot

Chips、ITC、VTS、ICCAD、DATE、IISWC等国际会议。申请并受理发明专利67项，获得发明专利权2项。所取得的阶段性研究成果对本项目明年的研究奠定了坚实基础。各课题的主要成果统计如表所示。

## 产生的学术影响：

2014年课题1共发表了20余篇国际期刊和会议论文，包括TVLSI等领域内著名的SCI期刊，也包含在领域顶级会议PACT、DATE上的论文。PACT会议是计算机体系结构领域的顶级会议，具有很大影响力。课题组主要骨干也受邀在PACT、DATE会议上介绍了其在众核处理器微结构方面的研究成果，得到了国际同行的认可。

2014年课题2的BigDatBench基准测试程序发布以来，已有超过8000个独立IP访问，国际上大概有20个团队已经使用BigDataBench公开发表论文。包括：剑桥大学的Andy Hopper团队等国际知名学生团体。BigDataBench被TPC组织技术报告列为三大“大数据测试基准”之一，另两个为BigBench(sigmod'13)以及UCBerkeley的测试基准；BigDataBench被同时被欧盟组织、SPEC组织等在大数据相关报

告中引用；在国内正联合工信部电信院标准化所制定大数据测试标准，已经完成技术需求和测试规范初稿，并提交标准申请。并被国际顶级会议 MICRO 2014 作受邀 Tutorial。消息式内存专利群经转让给华为公司，被评为华为高价值专利，获得 2013 年华为公司总裁奖。一体化存储相关研究工作在第 20 届全国信息存储技术学术会议上进行一体化存储工作大会报告，引起国内各存储研究团队的广泛兴趣和积极反馈。

2014 年课题 3 共发表了 11 篇国际期刊和会议论文，包含在领域顶级会议 DAC、ICCAD 上论文，产生了重要学术影响。李晓维研究员和韩银和研究员分别受邀加入了国际测试会议和国际并行体系结构和编译技术会议程序委员会，在国际本领域学术组织中发挥了重要作用。李华伟研究员和韩银和研究员分别担任了 2014 年度亚洲测试会议的大会主席和程序委员会主席，亚洲测试会议是测试领域国际三大会议之一，是亚洲集成电路测试领域最高学术会议。

2014 年课题 5 本年度的学术成果如下：1) 提出了适用于分布式文件系统的 HRsync 算法。在分析了现有的单机文件差异同步算法后，修改了 Rsync 算法中的文件分块方式和文件差异比对方式，提高算法并行度，减少了文件同步所需的时间。同时，避免了当文件相似内容很少时，Rsync 算法在源文件系统和目标中源文件系统和目标中计算时间不均等的问题；2) 利用 KV 技术降低了 chunkserver 上不必要的元数据开销和数据读写过程中的系统调用开销，并在文件系统页缓存上提供了针对整个文件的缓存以加速读取过程。同时，我们采用了 LSM 技术在保证数据读效率的同时解决了随机写问题；3) 在 Paxos 算法的基础上加入了使用纠删码的形式对需要存入的数据进行编码，在多个节点中存储的是进行编码之后的数据，而不是存储多个副本的原始数据；4) 提出了一种新的并发控制协议 Rococo，Rococo 通过记录并发事务之间的依赖关系，能够在事务真正执行前进行重排，从而以一个可序

973 项目 2014 年度成果统计表

课题编号	课题名称	已发表论文				发明专利		出版专著
		国际	国内	其中 SCI 收录	其中 EI 收录	申请	授权	
2011CB302501	面向高通量计算的可扩展、高效能并行微结构	12	8	2	18	5	1	0
2011CB302502	面向 EB 级数据并发处理的新型系统结构设计及评价方法	17	0	3	4	33	1	0
2011CB302503	大规模高通量计算系统的可靠性设计	11	0	5	11	2	0	0
2011CB302504	面向亿级并发负载的编程模型与支撑环境	9	3	4	6	7	0	0
2011CB302505	高通量计算系统的云计算服务环境	17	1	15	17	12	0	0
2011CB302506	高效可扩展移动服务关键技术及应用	16	1	8	9	8	0	0
合计		82	13	37	65	67	2	0



列化的方式执行延迟的事务片段。

2014年课题6的三个相关研究成果产生以下影响：1) 第一个研究成果发表在国际知名期刊 Wiley International Journal of Communication Systems 2013 上。和已有算法相比，该成果显著提高了虚拟网络映射请求接受率和底层网络长期平均收益代价比，从而最终达到提高底层网络运营商长期平均运营收益的目的。2) 第二个研究成果发表在国际会议 In Proceedings of the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid) Workshop on Cloud for Business Industry and Enterprises 2012 和国际知名期刊 Oxford Computer Journal (在线发表) 上。和已有算法相比，该成果在保证并行任务的执行效率的同时降低任务执行的能耗开销。3) 第三个研究

成果，在国际重要学术期刊 Elsevier Computer Networks 在线发表。和已有算法相比，该成果费用节省率和用户请求满意度等指标上取得了更好的结果。审稿人对此成果评价为“文章对 CN 期刊做出了有意义和有用的贡献。”( “The revised paper makes an interesting and useful contribution to Computer Networks.” )。4) 第四个研究成果发表在国际知名期刊 IEEE Transactions on Mobile Computing 2014 上。已有并发多路径传输协议并未对移动视频分发提出有效的并发多路径数据传输，该成果提出一种视频失真感知的并发多路径传输协议，保障异构无线网环境的高质量移动视频分发。5) 第五个研究成果发表在国际知名期刊 Elsevier Journal of Network and computer Applications 2014 上。和已有算法相比，该成果能够保证高清视频在传输过程中有较低延迟。

创新人才

创新文化

创新环境

## 重点 / 重大科研任务进展

# 面向服务的未来互联网体系结构与机制研究

“面向服务的未来互联网体系结构与机制研究”是计算所牵头的国家973计划项目(项目编号:2012CB315800),起止时间为2012年1月至2016年8月。本项目针对互联网面临的可扩展性、动态性、安全可控性等问题,围绕着传输与服务的动态复杂耦合、海量差异化服务的透明映射以及可扩展服务路由与高效传输等三个关键科学问题,研究未来互联网体系结构,在网络设计与运行机理层面取得重大突破,形成一系列核心理论和机理方法。

2014年度的主要工作集中在体系结构模型、关键技术研发、研究平台建设等三个方面。在体系结构模型方面,围绕未来互联网体系结构模型SOFIA展开工作,实现了SOFIA原型系统和相关

演示应用。在关键技术研发方面,主要针对虚拟机迁移机制、拥塞和传输控制、数据中心网络的路由机制、软件定义网络的实现等重要问题开展研究并取得多项重要成果。在研究平台建设方面,针对未来互联网的性能评估和验证需求,基本建设完成了未来互联网研究平台,可为未来网络关键技术的研究、关键原型系统的测试评估、创新业务应用提供良好的示范平台。

本年度,相关成果发表国际学术论文129篇、国内学术论文20篇(其中SCI收录73篇、EI收录69篇),出版专著2部;在国内外重要学术会议做特邀报告15次;申请国家发明专利37项,获专利授权6项。在本项目资助下,培养博士研究生20名、硕士研究生89名。



## 重点 / 重大科研任务进展

# 网络大数据计算的基础理论 及其应用研究

“网络大数据计算的基础理论及其应用研究”是由计算所牵头的国家973计划项目（项目编号：2014CB340400），起止时间为2014年1月至2018年8月。本项目从网络大数据计算的面临的数据复杂性、计算复杂性以及系统复杂性三个挑战性问题出发，提出“网络大数据复杂性的内在机理”、“网络大数据可计算性及新型计算范式”、“网络大数据处理系统的效能评价与优化”三个关键科学问题，并设置了6个课题：课题一（网络大数据复杂性解析及计算模型）和课题二（全周期感知的网络大数据计算系统架构体系）进行理论基础研究，课题三（网络大数据感知与融合表示）、课题四（网络大数据内容建模与语义理解）和课题五（网络大数据模式发现与效应分析）进行技术基础研究，而课题六（网络大数据处理软硬一体化引擎系统及示范应用）是整个项目的出口，属于系统研究和技术应用示范。

2014年度课题完成的成果具体包括：

1. 在网络大数据分布规律发现、网络大数据复杂特征度量、大数据约简的基础理论与方法以及网络大数据的计算模型理论这四项研究中都有所收获。在网络大数据分布规律发现方面，重点研究了社交网络、合作网络中的大数据存在什

么样的内在规律，发现了其中的时序规律以及群体分配机制；在网络大数据复杂特征度量方面，研究了如何对数据的内在复杂特征有效地抽取和度量，提出了复杂特征抽取、特征选择和数据相似度度量方法；在大数据约简的基础理论与方法方面，主要研究了如何构建不同类型大数据的约简表达，并对大规模文本数据、复杂网络数据进行了约简建模；在网络大数据的计算模型理论方面，主要研究了面向机器学习任务的近似计算和优化算法，提出了面向排序学习、推荐系统以及在线学习的近似优化算法。该方向在国际、国内期刊会议上共发表论文37篇，其中SCI收录论文7篇，EI收录论文21篇，在国际著名期刊如TKDE、PNAS等上发表论文5篇，顶级国际会议论文8篇。

2. 在网络大数据计算系统架构体系与评价方法研究中，主要研究应用负载的特征表达及状态约简方法，解决基准测试程序的自动标识的问题；研究体现大数据应用共性访问特征的富语义编程模型，降低应用编程难度并方便针对多个应用进行统一优化设计；研究基于流水线的多源异构大数据计算框架，用户只需进行基于有向无环图的应用编程，而无需掌握分布式集群计算的细节；研究大数据应用的数据访问特征，透过数据

访问的弱局部性挖掘数据访问的关联规则。该方向共发表论文 11 篇, 其中 IEEE Transactions 论文 3 篇, 顶级国际会议论文 3 篇。

3. 在多源异构网络大数据感知融合与表示研究中, 提出多源异构网络大数据信息感知的理论和基本框架。在本领域顶级学术会议发表论文 2 篇, SCI/SSCI 检索论文 4 篇。

4. 在网络大数据内容建模与语义理解研究方面, 提出基于学习自动机的特征结构模型描述方法; 提出改进线性判别分析的大数据表示方法; 构建了一种多粒度、语义化的知识组织模式, 在服务搜索的时间和召回率方面具有明显优势; 设计了一种目标驱动、内容标签组织的知识库构建方法, 可有效提高服务发现和服务推荐的质量; 提出了一种两阶段的主题聚类方法, 在语义分类的基础上基于概率发现潜在的主题, 提高查询性能; 该方向新发表或接收论文 26 篇, 其中在 IEEE Trans 论文 2 篇和 SCI 期刊检索论文 12 篇, 国际会议论文 8 篇, 国内核心期刊 4 篇, 获得中国发明专利授权 1 项, 获得 CCF 科技进步一等奖一项并编制国家标准一项。

5. 在网络大数据模式发现与效应分析方面, 在网络大数据关联关系分析和挖掘、网络大数据演化、大数据的映射和约简机制这三项研究中都有所收获。在网络大数据关联关系分析和挖掘方面, 提出了一种基于属性和结构信息的关系推断方法, 以及融合多种信息的隐性关联发现算法; 在网络大数据演化方面, 提出了大规模网络中重叠社群的高效发现算法, 能对网络结构的演化特点进行追踪, 并发现演化中的关键节点; 而针对大数据的映射和约简机制, 建立了一个面向应用对知识图谱描述能力加强的框架, 能够提高数据映射的覆盖率和准确度。该方向共发表论文 19 篇, 其中 IEEE Transactions 论文 1 篇, 其他国际著名期刊论文 3 篇, 顶级国际会议论文 8 篇。

6. 针对软硬一体化的网络大数据处理系统研究所涉及的多层次研究内容, 多个研究小组对本课题的关键技术进行攻关, 本阶段重点大数据存储和处理引擎最小系统的部署和测试、以及大数据计算与分析算法研究工作, 完成了大数据处理系统硬件选项和最小一体机系统的研制, 已部署多套系统, 可稳定运行。针对示范应用需求, 课题组从资源平台建设、违法犯罪信息标注、违法网站行为分类、到违法网站行为结果的可视化 4 个方面, 开展了相应的工作, 按课题计划搭建了总存储容量 100TB 的试验环境, 可支持其它课题开展开展网络信息空间内大数据挖掘理论、基础技术研究及验证工作。初步建立了论坛(Discuz!、phpWind) 行为特征库、传销类及赌博类网站行为特征库, 为下步算法验证打下了基础。系统关键技术申请国家发明专利 1 项, 软著 2 项。

项目的学术交流 International Workshop on Data Science and Big Data Analytics 在深圳召开, 邀请怀进鹏院士、项目首席科学家华云生教授、及国际知名学者熊辉、杨强等人做了精彩的学术报告, 收到会议论文 6 篇。

项目年终总结会在香港中文大学深圳研究院召开, 科技部专家戴国忠研究员、彭群生教授、唐常杰教授听取了项目进展汇报并提出了下一年的建议。



项目年终总结会



## 重点 / 重大科研任务进展

# 可重构大内存加速卡

可重构大内存加速卡依托于973课题——“面向深度测序大数据量的计算模型和体系结构研究”，针对生物大数据及各种具有大数据特征的计算需求而开发。课题组利用可重构技术突破通用处理器指令集和系统结构的限制，以更加直接的运算方式实现对算法的有效加速：

◎ 大数据应用需消耗大量内存资源，加速卡最大板载96GB ECC DDR3内存，为大数据应用的内存开销提供可靠保证；

◎ 加速卡板载6组万兆网模块，实现对远

程大数据的采集与处理；加速卡支持OpenCL编程模型，体现良好的编程通用性和可移植性；

◎ 由算法到硬件逻辑的自动生成技术能够大幅提升设备的易用性和开发效率，并为算法提供后期功能升级的即时性保障。

由于算法规模的多样性，本加速卡支持一机多卡的扩展能力，满足不同规模算法的加速需求。

基于OpenCL的异构可重构加速体系统结构由Host端和Kernel端两条关键工具链支撑。Kernel端使用OpenCL SDK将算法的热点部分自动映射为FPGA片内核心逻辑，并与FPGA片



图1 可重构大内存加速卡

内预先定制的控制逻辑、外部接口逻辑及内部互连逻辑连接适配。

加速芯片内的核心加速模块以高并发深流水的定制计算资源堆组成。片内有两组互连网络，一组用于核心逻辑与 PCIE 和 DDR3 存储控制器之间通信，构成全局内存互连网络；另一组用于核心逻辑与高带宽低延迟的片内本地内存阵列之间通信，构成本地内存互联。片内本地内存采用

8 BANK 高并发阵列，用于 kernel 内局部数据的快速存取。

OpenCL API 调用设备 Driver 操作 PCIE 接口，以调度和使用可重构加速卡设备。Kernel 端工具链利用 LLVM 编译框架，在 IR 加入 RTL 生成器，分析 IR 的控制流和数据流，映射成高并发深流水的硬件逻辑表达，以实现 Verilog 代码的自动生成。

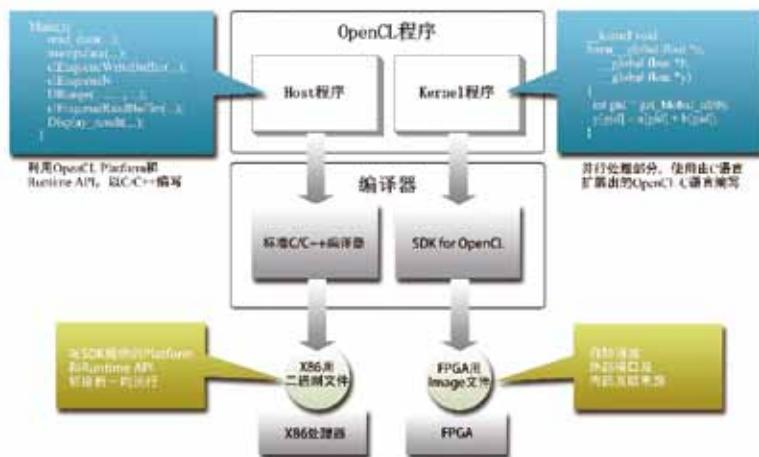


图 2 基于 OpenCL 的异构可重构加速体系结构

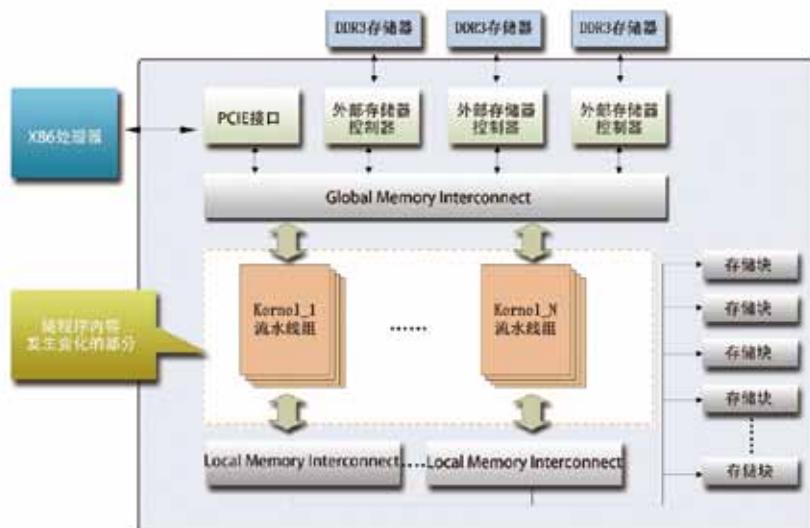


图 3 异构可重构大内存加速卡片内结构



## 重点 / 重大科研任务进展

# 新型多核 / 众核处理器编程与运行支撑环境

“新型多核 / 众核处理器编程与运行支撑环境”项目是国家“十二五”863计划信息技术领域重大项目（项目编号2012AA010900），下设5个课题：“面向多核 / 众核处理器的并行程序调试技术与工具”（课题号2012AA010901）、“面向多核 / 众核处理器的并行程序编程技术、框架和语言支持”、“面向国产处理器的并行程序综合优化技术与系统”、“面向领域的并行应用综合优化技术与系统”以及“面向多核 / 众核系统的运行时支持技术与系统”。本项目起止时间为2012年1月至2015年12月，于2014年11月初通过中期检查。本项目2014年的研究工作主要改进多核 / 众核处理器编程、运行时以及调试技术的总体设计，进行关键技术的验证，并初步探讨了针对国产处理器以及典型应用领域的综合优化技术。

计算所是该项目的项目首席专家单位，支持课题2的研究工作以及参与了课题1的研究工作。计算所相关团队在2014的主要进展如下：

## 1. 面向多核芯片的并行化编程技术研究。

在安全的并行化编程模型POMP方面，进一步完

善了并行优化框架，设计并实现了上下文敏感、流敏感的指针分析，循环分块等优化变换，提供动态的并行性分析，并开展了X86和ARM处理器的SIMD、GPGPU等多种平台的自适应优化。部分成果已交付华为公司，并促成与华为的新合作——多核编程框架研究项目。

**2. 面向领域应用的并行编程框架和多核优化研究。**完成AceMesh模式语言的编译器原型开发。在多核的并行优化方面，实现基于亲和性、以数据重用为中心的任务调度系统，调度策略可利用上层领域信息、并对MPI通信敏感。对Chombo示例程序集的测试表明，AceMesh编程系统可以把适应性网格的遗产应用自动转化为乱序任务图并行的程序，比MPI+OMP混合并行的加速平均可达1.8倍。

**3. 异构平台的编译支持。**针对众核处理器和异构平台结构各异、优化差别大的问题，我们设计了OpenCL的通用编译优化框架和层次化的编译支持系统，以灵活适配多种平台。我们在Tilera/Intel MIC/CPU等典型多核 / 众核平台上实现了OpenCL原型支持系统，验证该框架的有

效性。本工作获得了华为公司的高度认可，并将于 2015 年移植到其他国产异构加速平台。

**4. 新型编译框架的研究。**在可扩展多核编译框架中，我们提出语义模式构建接口来持续地将专家经验集成在通用编译器中。2014 年度，我们从循环结构、数组访问模式、标量访问模式、依赖关系几个维度来定义矩阵计算的语义模式。基于脚本语言 POET 实现了对上述语义模式的自动识别，并添加必要的标记信息，支持后续优化。对 BLAS 库的测试表明，本方法用 GCC 优化后在性能上接近专家手工调优库 MKL 和 ATLAS，对于某些核心性能优于这两个库。

**5. 并发错误的快速自动定位方法：**现有自动定位方法存在精度不高、开销大的问题。我们

针对触发并发错误的共享内存访问对的不同场景，设计不同的测试过程，进行错误定位。当程序执行中触发并发错误时，使用相应的测试过程来分析在成功执行和失败执行中收集到的共享内存访问对，直接找到导致并发错误的共享内存访问。试验表明，本方法的开销小且不需要大量的成功执行就能实现精确定位。

上述的相关成果发表在著名国际学术刊物 IEEE TPDS 以及 MICRO、CGO、ASE、DATE、CF 等著名国际学术会议上。在本项目的其它主要成果包括：改进基于数组类型的统一编程模型 PARRAY，并移植到异构的国产申威处理器。针对申威处理器的统一架构并行语言 (UFPL) 在主要关键技术研究上已经取得预期进展。

科研为国分忧  
创新与民造福



## 重点 / 重大科研任务进展

# 3D 内容视觉获取技术及设备

“3D 内容视觉获取技术及设备”是国家 863 计划课题（课题编号为：2012AA011501），起止时间为 2012 年 1 月至 2014 年 12 月。课题的主要目标是研发新型 3D 内容视觉获取设备和相应的处理技术与系统，方便、高效地生成保持细节信息的 3D 内容，带动相关产业的发展。

课题本年度申请国家标准“3D 获取设备数据交换技术规范”1 项，提交技术发明专利申请 1 项。课题主要进展有：

## 1. 基于多透镜阵列数据的三维几何重建

研究基于多透镜阵列的三维重建方法。根据多透镜阵列的光学特性，通过边缘检测、边缘深度计算和深度扩散，分析从多透镜阵列获取的光场图像，计算恢复出场景的三维信息（图 1）。

## 2. 基于红外散斑数据的三维几何重建

研究基于红外散斑的三维重建方法。首先将红外散斑信息转化为深度信息，然后将目标物体深度图像序列的表面重建问题转化为 3D 点云和表面的映射和匹配问题，最后利用随机采样一致性和迭代最近点方法确定连续 3D 点云的对应关

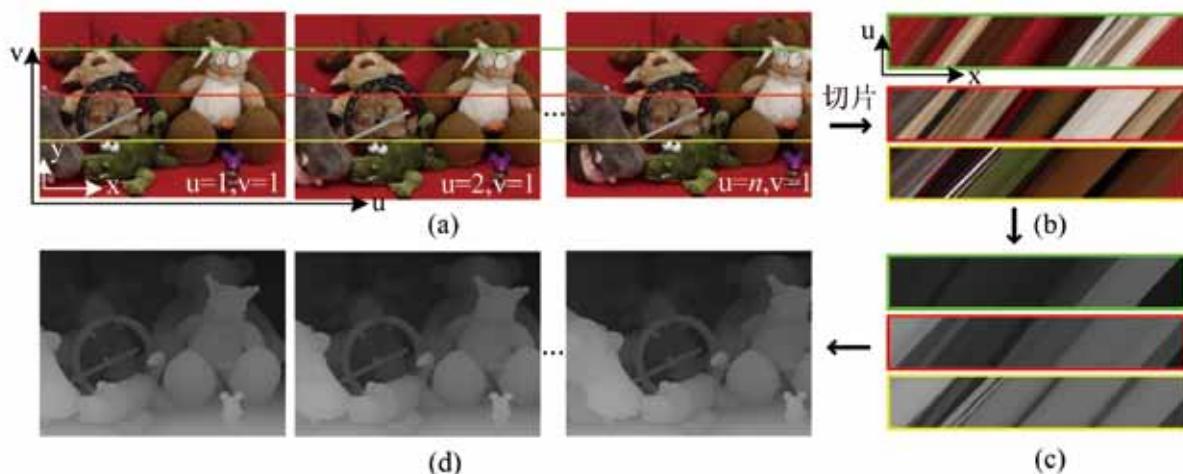


图 1 (a) 光场图像序列，(b) 对积图像，(c) 对积图像上的深度图，(d) 深度图像序列

系(图2)。物体表面顶点重建位置的平均相对误差为5%。

### 3. 三维动态火焰重建

研究基于物理的启发式火焰重建方法。提出了一种基于能量约束的火焰三维重建模型,结合热力学逆辐射过程,重建火焰的三维温度场数据。

首先,利用CCD相机同步采集火焰数据,实现不同角度的火焰数据获取。然后,基于拍摄的火焰图像,通过定义能量优化模型和基于光线投射的自适应温度调整算法,迭代求解,实现火焰温度场的重建,并将其应用于虚拟场景中(图4)。火焰场在R、G、B三个通道上的平均重建误差为13.1。



图2 转椅的三维重建结果

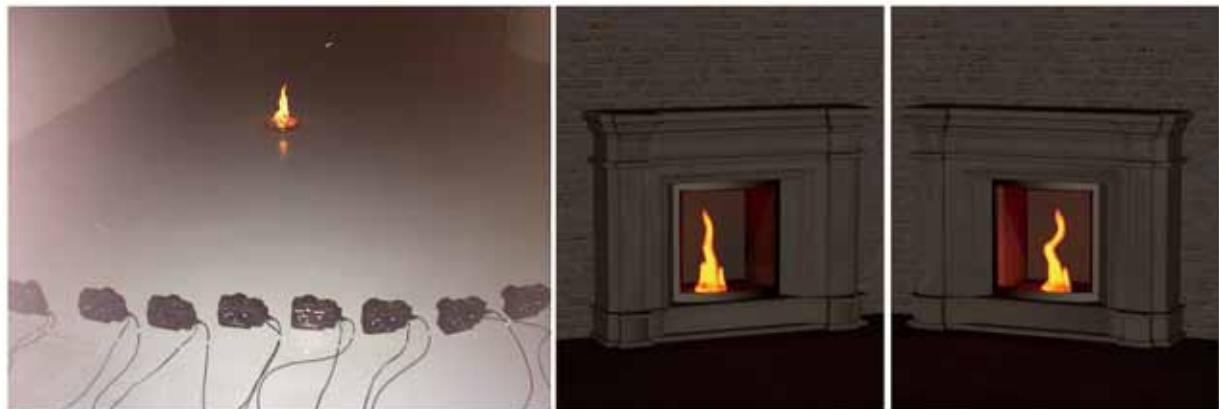


图3 三维火焰场的重建结果



## 重点 / 重大科研任务进展

# 可编程虚拟化路由器关键技术研究及系统研制

可编程虚拟路由器相关研究得到国家 863 计划项目（项目编号：2013AA013501）和中科院先导专项（项目编号：XDA06010303）的支持，旨在研究 SDN/NFV 关键技术及路由器系统的实现。本年度课题组重点研究了数据包处理关键技术，并在路由查找算法和多域数据包分类算法方面取得了研究成果：

路由查找作为一个成熟的领域，有大量的软件算法和硬件算法，但迄今没有一种算法可以达到理想目标： $O(1)$  的查找时间和  $O(1)$  的片内内存。课题组提出一种基于二维划分思想的路由表查找框架 SAIL，基本原理如图 1 所示。主要思想是将路由表的前缀长度信息和端口信息分开、将路由表前缀从 24 层分开，两维划分之后只将前缀长度小于等于 24 的前缀长度信息存放于片内内存，其他部分存放于片外内存。这样划分之后，片内的前缀长度信息的理论最大值只有 4MB，完全可以存放在现有的各种平台的片内内存之中。从而，在该框架下实现的算法都可以实现  $O(1)$  的查找时间和  $O(1)$  的片内内存。在此基础上，课题组提出了面向不同需求的三种基于 SAIL 框

架的查找算法：基本算法、面向更新算法和面向查找算法。这些算法实现了路由表查找速度、片内内存大小与路由表大小、特性无关，与虚拟路由表个数无关。其中面向查找的算法的速度是经典算法及已广泛部署算法的 7 – 50 倍，并且支持各种平台，包括 CPU、GPU、FPGA 以及 Many-core 平台。该部分研究成果发表在计算机网络及通信领域最高会议 ACM SIGCOMM 2014 上。之前中国单位在该会议上鲜有论文发表。

数据包分类是按照一定的规则集和数据包头部的域信息，找出与数据包匹配的规则条目的过程。数据包分类方法可以分为基于 RAM 的方法和基于 TCAM 的方法两大类。基于 RAM 的方法通过设计高效的决策树算法进行快速匹配，相比基于 TCAM 的方法，具有功耗与成本优势，但是存在查找性能不确定的问题。课题组针对不同决策树算法在不同规则集上所产生性能波动的现象开展研究，发现了规则集的覆盖均匀性对不同算法的访存次数具有直接影响，同时规则集内“正交结构”规则的数量将直接决定算法的内存占用。针对规则集的覆盖均匀性，设计了分析和

量化方法；根据正交结构规则的特性，提出了一种内存估算模型。结合所提出的量化方法和内存估算模型，设计了 SmartSplit 多决策树算法和 AutoPC 算法策略框架。对比已提出的最好算法，SmartSplit 算法的访存次数平均减少 1 倍，内存占用最多下降 10 倍。如图 2 所示，对于给定一

个规则集，AutoPC 框架能够根据规则特性自适应选择合适的算法，对比仅采用同一种算法，平均性能有 3.8 倍的提升。该部分研究成果发表在网络领域顶级会议 ICNP 2014 上。

上述算法在可编程虚拟路由器平台上试验取得良好的效果，均带来数倍的性能提升。

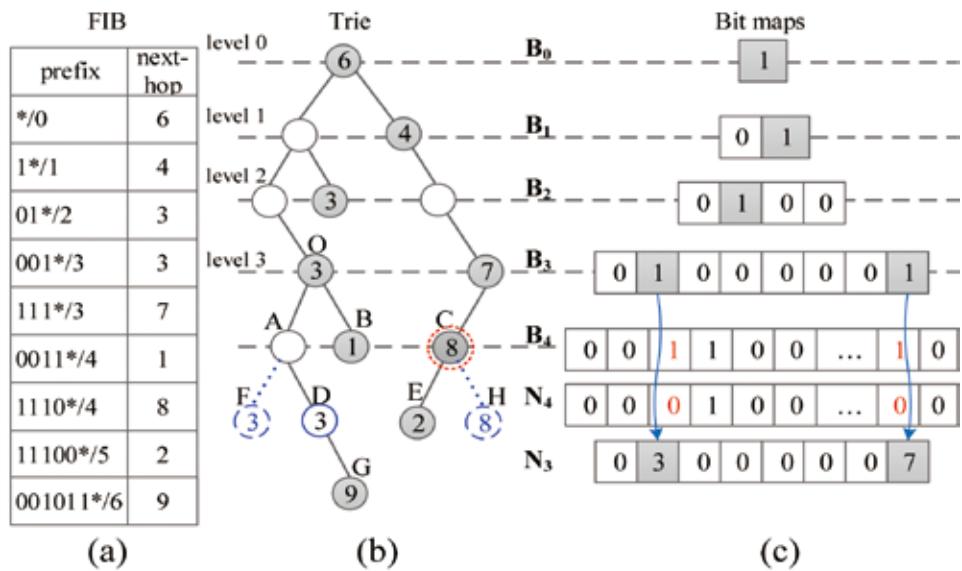


图 1 SAIL 框架

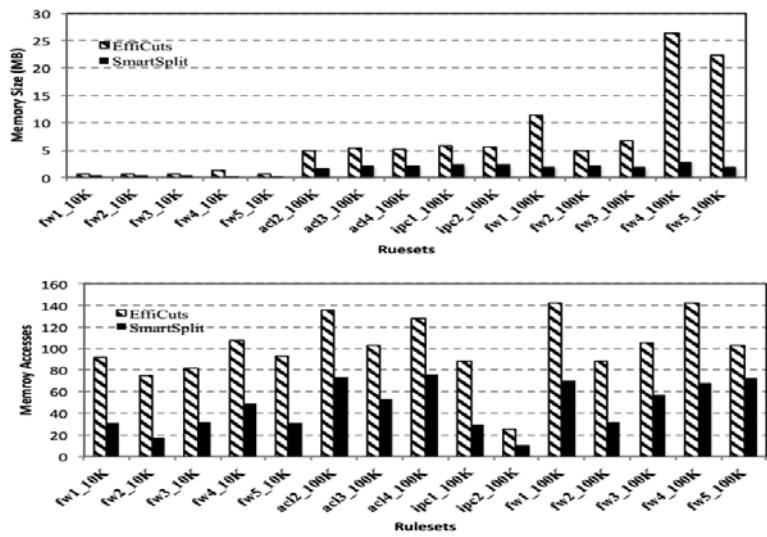


图 2 SmartSplit 和 EffiCuts 算法性能对比



## 重点 / 重大科研任务进展

# 异构媒体数据的关联与挖掘研究

“异构媒体数据的关联与挖掘研究”是计算所牵头承担的国家863课题（课题编号：2014AA015202），起止时间是2014年1月至2016年12月。本课题按照预定计划重点开展了异构媒体数据的多模态关联分析研究分析方法的研究，以挖掘不同模态数据间的横向语义关联关系，取得了一系列研究成果：（1）提出了一种深度哈希学习框架，并应用于跨模态最近邻搜索，通过深度学习、哈希学习技术有效解决了传统方法中的瓶颈问题，学习出高维数据跨模态的有效共享空间表达，实现了高精度的跨模态最近邻搜索。（2）在多源融合和跨模态层次化语义关联方面，提出了基于复杂语义表示的跨模态相关学习方法COLAR和TINA；（3）提出了一种异构媒体话题分析方法，通过提出一种多分辨率行为建模方法来度量不同用户两两之间多时间分辨率的行为相似性，从而识别出用户的社区结构。（4）提出了一种基于“多模态图”（Multi-modality Graph, MMG）的多模态异构媒体及时序关联建模方法，用于从跨模态、高噪声的互联网数据中检测出热点网络事件。（5）提出了一种层次

化语义的复杂多层特征学习方法，该方法采用一种共享模型学习方法，利用多层的特征表示提高层次化分类学习的效果。（6）提出了一种基于多特征融合的多语义视觉描述学习方法，通过三种层次的特征聚合产生多语义图像表示的框架。

（7）提出了一种基于用户点击信息的多模态图像检索重排序方法，通过结合用户点击信息（click-through data）和多模态特征（multi-modal）有效地解决了尾查询初始检索效果不佳的问题，实现了尾查询图像检索效果的大幅度提升。（8）基于特征学习、度量学习及哈希学习，提出了一套基于深度学习的图像检索框架。（9）提出了一种基于深度学习的视频事件检测方法，通过深度学习对特征进行融合，反映语义信息之间的联系，有效地进行复杂事件识别。（10）面向媒体大数据，我们提出了一个新的基准用于实例级视觉物体检索和识别（INSTRE），针对大规模物体检索问题使用多种评价标准全面地评价几种流行的算法。

（11）构建了新华社跨媒体知识库，对新华社大量原本无序的跨媒体数据资源进行整合。

到目前为止，本课题已发表（含已录用）学

术论文 25 篇, 申请专利 6 项。通过课题研究, 锻炼和培养了中青年学术骨干, 增强了研究队伍的整体实力。培养博士研究生 6 名、硕士研究生 4 名。通过组织全国特定音视频分析系统评测资格大赛, 扩大了课题组的学术影响。

在 2014 年度, 本课题遵循既定研究思路, 已圆满完成了计划的主要研究任务, 达到课题在 2014 年度的研究目标, 并在部分方面取得了突出的研究成果, 为研究工作的进一步深入开展打下了良好的基础。

创新  
求实  
团结  
高效



## 重点 / 重大科研任务进展

# 高精度低成本低功耗的海水电导率传感器研究与设计

“高精度低成本低功耗的海水电导率传感器研究与设计”是国家863计划课题（课题编号：2014AA093402），由计算所独立承担，起止时间为2014年1月至2016年12月。本年度课题重点开展了传感单元、测量电路、以及传感器样机的研究工作，并顺利通过项目组的年度考评。主要进展有：

## 1. 完成电导率传感单元的设计与加工

在电导率传感单元方面，我们首先通过有限元仿真测试，测试了单面矩形双电极、双面圆形双电极、单面对称矩形三电极、以及单面对称圆环双电极等多种方案的趋近效应，并选择了趋近效应最小的单面对称圆环双电极作为传感单元的设计方案。随后，我们采用陶瓷、聚四氟乙烯等基板材料，采用PCB加工工艺，进行电导率传感单元的加工试制，并采用镀铂、镀钯、镀金的方式在银导体上对制作了不同的电极。在三个月的浸泡老化试验之后，确定了陶瓷基板和镀铂的加工工艺，完成了传感单元的设计加工。

## 2. 完成测量电路的设计与测试

在测量电路设计中，我们基于三参数正弦波拟合的测量原理，实现了16位精度的全阻抗测

量电路。为测试电路的测量准确度与稳定性，我们采用铂电阻、温敏电阻作为温度传感器、以及一个100Ω的高精度电阻，对其进行持续18天的持续温度和阻抗测量。测试结果表明，在我们的测量电路支持下，经过粗略校准的两个温度传感器可以达到0.01°C的一致性与分辨率；阻抗测量的标准差为测量值的0.0018%，量程的0.0009%，说明该测量电路能够达到16位的测量精度。同时，我们还对测量电路的功耗指标进行了测量。其供电范围为6~16V，休眠电流6.5uA，工作电流为12mA，单次测量时间为40ms，在采用7.2V/2000mA双节5号锂电池供电时，能够满足课题任务中的对应功耗性能指标。



图1 用于单点测量的CTD传感器样机

### 3. 实现用于单点测量的 CTD 原理样机

在完成传感单元设计与测量电路设计后，我们对传感器进行了简易的组装和防水封装，形成了可用于单点测量的传感器样机（图 1），并对其在氯化钠溶液中进行了初步的电导率和盐度测量测试。如图 2 所示，在 4 个小时测试中，水温从  $26.158^{\circ}\text{C}$  上升到  $26.356^{\circ}\text{C}$ ；同时测量到的电导率从  $26.733\text{mS/cm}$  上升到  $26.851\text{mS/cm}$ ，

并且变化趋势与温度变化曲线吻合。将温度和电导率转化为盐度后，盐度数值显示出较好的稳定性，其抖动噪声在  $0.001\text{psu}$  以内。同时，在测试过程中，所测盐度逐步升高了  $0.01\text{psu}$ ，应是测试过程中水分蒸发引起的实际盐度升高，而非传感器的漂移造成。为实现对深海的测量，我们还设计了支持  $6000\text{m}$  水深的传感器封装，设计图如图 3 所示。

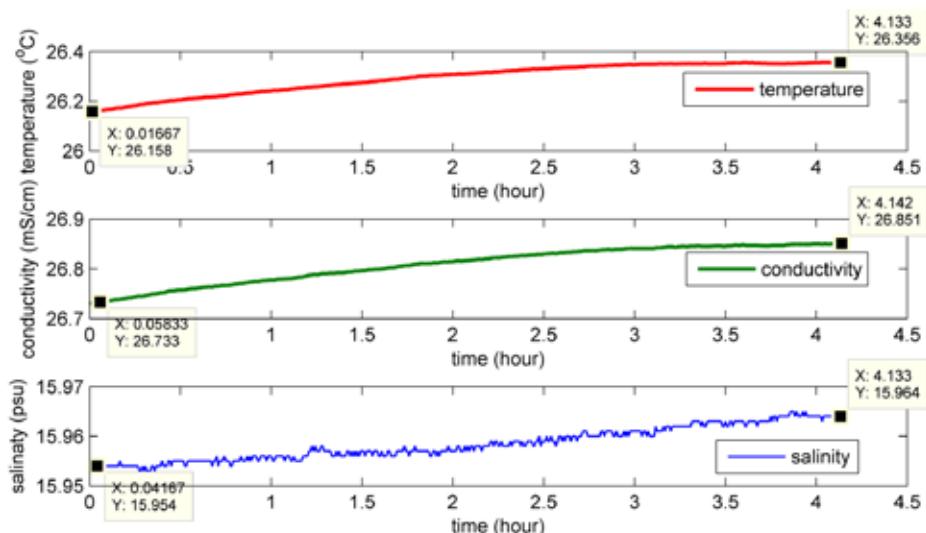


图 2 电导率传感器测试数据

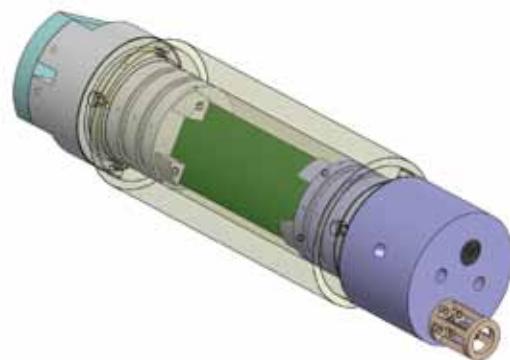


图 3 支持  $6000\text{m}$  水深的防水封装设计图



## 重点 / 重大科研任务进展

# 超并行高效能计算机体系结构与设计方法研究

“超并行高效能计算机体系结构与设计方法研究”是国家自然科学基金委员会创新研究群体科学基金项目，负责人是李国杰院士。本创新群体已经顺利通过了一期考核，顺延为二期。第二期起止时间是2013年1月至2015年12月。

研究群体在2014年度中，以高效能计算、低功耗计算、高可靠计算为研究方向，在高性能和高效能计算机的超并行体系结构，多核/众核处理器的新型体系结构，以及多核处理器的测试、验证与高可靠性设计方法等六个方面上开展研究，取得了一批国内外同行公认的重要成果。

在高性能计算机超并行体系结构方面：本年度研发的面向集群系统的图并行计算模型Graphine，采用了Scatter-Combine计算模型和Agent-Graph数据模型，有效解决了图计算并行分解问题。本年度研制了基于FPGA的40核高通量处理器、消息式内存扩展卡、一体化存储卡、DCN-FPGA、PARD等一系列硬件原型，验证了新的设计思想，为下一阶段的系统级原型奠定了基础。发布了大数据分析应用BigDataBench 3.0，并在多个国际顶级会议上进行了介绍，得到国际学术界的使用和认可。本年

度共申请了发明专利100多项，发表了ISCA、TPDS、ICS、IISWC、DAC、TAO等一系列高水平文章。

高效能计算机设计方法方面：为提高数据中心效能，我们在数据中心系统监控和低能耗负载调度方面展开研究，具体包括：（1）提出基于大规模分布式监测的实时压缩算法，包括一个线性压缩算法(LCA)将线性滤波器集成于实时监控，一个新的轻量级实时压缩的算法(ReCA)通过探索指标间的关联性来抑制远程信息的更新，一个探索指标间的关联性的压缩算法(MiCA)，与现有方法相比压缩率最高可提高44.5%，相关工作发表于CCF B类会议ICPP 2014；（2）设计和开发了多点温度及风速无线监控系统，相关工作受理专利1项；（3）提出计算与制冷能耗权衡的虚拟机迁移模型及算法，与负载均衡方法相比节能最高达54%，优于现有的VMAP方法3.9倍，相关工作受理专利4项。

在多（众）核处理器的新型体系结构和低功耗设计方法方面：（1）针对众核系统中的稀疏目录的存储扩展效率问题，研究稀疏目录替换策略、合作式目录替换策略，以及victim目录设

计方法和海绵目录的设计方法。并在国际顶级会议 PACT 上发表论文。(2) 研究面向能耗优化的多核处理器系统资源分配方法。通过分析处理器核数、频率、并行开销、访存开销、带宽划分等因素对能耗的影响, 针对同构处理器提出基于 DynamicVoltage/Frequency/Core Scaling 的能耗 - 性能模型; 针对异构处理器提出引入访存排队延迟的性能 - 能效模型。从而从模型和实际执行两个方面提出最优能耗所用核数和频率的设计空间探索方法。(3) 从高通量应用本身出发, 针对高通量应用的特点, 研究内存管理优化机制, 以提高高通量计算的性能。

在多核处理器的测试、验证与可靠设计方法方面: (1) 测试验证方面, 提出了同时考虑最长敏化路径和  $n$  倍检测覆盖率的时延测试向量选择方法、引导模拟验证激励生成的半形式化方法, 在 TVLSI 发表论文 2 篇。(2) 在片上互连和存储容错方面, 提出了跳变片上网络设计方法, 实现了带宽在时间维和空间维上的异构分布; 提出了 DRAM 的动态刷新机制, 实现了功耗和可靠性之间的折中。在包含 TVLSI、ICCAD 等期刊会议上共发表了 8 篇论文。(3) 在多核可靠性设计方面, 提出了基于分层隔离电压噪声与紧急线程优先线程调度的三维堆叠多核处理器电压紧急消除方法、基于热能功耗估算模型的多核处理器热能功耗管理方法、基于 Short-Set 的相变存储器高效能写方法, 在 DAC/DATE/ASPDAC 等领域旗舰会议上发表 4 篇论文。

在高性能计算机的并行编译方法方面: (1) 完善非精确平台上面向算法的编程支持, 被 TACO 录用; (2) 扩展网格应用的并行模式语言, 支持 MPI 应用; 提出对任务类型、位置敏感的时间扭曲调度, 性能提升可达 45%。论文投稿中。(3)

提出 GPU 平台上一种新制导和相应的编译优化方法, 性能可达 OpenACC 和 OpenHMPP 的 3.6 倍和 1.3 倍。论文投稿中。(4) 领域专家协同的编译方法方面, 提出语义模式的自动识别算法, 论文被 MICRO 2014 录用。(5) 针对大规模服务器, 提出差异性 I/O 调度方法, 显著地提高系统吞吐率, 发表于 JCST。(6) 建立异构集群上 MapReduce 应用的性能建模评估方法, 把典型 MapReduce 应用的性能提高 1.4 倍到 12.5 倍, 论文投稿中。(7) 提出并发错误的快速自动定位方法, 发表于 ASE14。

在本项目的研究过程中, 产生相关研究平台也是本群体的目标之一。在项目开展过程中, 高效能可重构大数据处理加速引擎研究平台、机器学习加速器的结构和编程研究平台、面向测量信息的日志记录分析平台、众核处理器并行编程方法研究平台等四大自主研究平台也取得了预期进展。

本项目本年度的主要研究工作取得了一系列成果, 并得到了国际同行的肯定。2014 年度总共发表学术论文 62 篇, 其中 EI 索引 44 篇, SCI 索引 20 篇, 期刊论文 26 篇, 会议论文 36 篇, 申请 / 授权专利 58 项。



创新群体年度总结会现场



## 重点 / 重大科研任务进展

# 云环境下的图像视频群体协同表达与处理

国家自然科学基金重大项目“云环境下的图像视频群体协同表达与处理”（项目编号：61390510）于2014年1月14日在中科院计算所举行了项目启动会。该项目的执行时间是2014年1月至2018年12月，下属5个课题的承担单位分别为中国科学院计算技术研究所、北京工业大学、同济大学、中国科学技术大学、北京大学。本项目旨在研究云环境下的图像视频群体协同表达与处理，发展以群体化、感知化为基础的信息理论，建立云环境下图像视频群体表达、编码、传输、评价方法，形成具有自主知识产权的AVS云媒体标准，促进我国信息产业由制造优势向技术优势、数据优势的转变。

经过执行期第一年的研究，各课题均按照预定计划取得了较好的研究进展，主要成果如下：

1. 在群体数据的协同结构化表达与处理方面，提出了跨欧氏-黎曼空间的异质度量学习方法，为解决图像与视频之间的异质数据匹配和识别问题提供了全新的理论框架；提出了混合多阶统计量的度量学习与多核分类器融合的判别学习方法，有效解决了视频识别中变化复杂、分布差异大、语义不对齐等难题；提出了一种高效稳定的

视频运动基元描述子，以有效刻画视频的中层语义信息，且对于时空对齐具有很好的鲁棒性，在视频动态表情识别任务中取得了成功应用。针对大规模图像快速检索问题，基于图像视频的结构化属性，提出一种时空高效的视频紧致二值编码表示，可有效应对大规模视频检索的挑战。初步建立了小规模的面向图像视频群体结构化的视觉概念库雏形，包括398个动物类别、20个属性的语义标注和关联。

2. 在群体化图像视频编码方面，提出了一种联合非局部相似性和稀疏性约束的高效光流正则方法，建立了相邻视频帧间的对应关系，取得了性能领先的光流场估计精度；提出了一种利用图像局部结构增强稀疏性的基元表达方法，改善了当前基于小波的压缩图像重建效果。提出了一种基于二维字典的编码方案，实现了跨越编码器的分块边界匹配。基于群体图像的层次化表示模型，提出将自适应滤波器用于图像分层次表示，在利用较少系数进行图像重构方面取得了优于小波变换的效果。

3. 在群体化图像视频传输方面，主要研究采用视觉感知失真度量代替传统的信号失真准则

研究率失真关系,提出了利用多用户的带宽异构特性进行伪模拟的失真可容忍传输方法,可有效优化每个用户分配的最优带宽与功率资源;提出了数模混合的失真可容忍图像视频传输方法,结合数字视频传输系统压缩率高和伪模拟系统的信道适应特性,实现了移动环境下的信道自适应传输;在平台建设方面,自主设计了数字中频信号处理电路板,支持2收2发的无线射频电路板,可以支持实现伪模拟图像视频传输的各项功能。

4. 在图像视频质量评价方面,为建立与感知一致的图像失真度量,提出了基于人脑自由能模型的无参考图像质量评价算法;进而提出了一种针对复合失真的图像质量评价通用的框架以及评价复合退化盲图像的方法,达到了与流行的全参考算法相当的性能。针对移动显示设备上3D视频的主观评价问题,提出了一种利用图像查询重建误差评价检索结果质量的方法。启动建设2D/3D图像视频质量评价数据库,旨在构建符合研究需求以及国际标准的评测平台。

5. 在面向群体数据的AVS云媒体标准及验证平台方面,提出了基于深度学习的视频动态建模方法,设计了深度动态编码器叠加和训练方法,以提高动态建模的准确性;针对视觉词典设计,提出了基于频域拉普拉斯高斯滤波的超低复杂度兴趣点检测子BFLoG,在精度与效率两方面均优于已有的SIFT和SURF等经典算子。提出了紧凑视觉模式包CBoP对视觉模式进行稀疏编码,能够以更小编码长度重建词袋模型的直方图,将检索速度提升20到25倍。提出了海量网络群体数据标识与压缩方法,在百度大规模图像检索中得到验证和应用;针对视频监控压缩,提出了低复杂度监控视频运动搜索算法、背景图像质量可伸缩编码方法、基于在线学习的多特征亲和度融合算法、基于变形敏感软级联模型的视频近似拷贝检测等方法,并初步建立了基于云计算的媒体编解码框架。

上述研究,2014年度共发表国际刊物论文33篇,国内刊物论文1篇,国际会议论文44篇,申请专利21项。



项目启动会会议现场



## 重点 / 重大科研任务进展

# 在线社会关系网络的挖掘与分析

本项目为国家自然科学基金重点项目（课题编号：61232010），起止时间是2013年1月至2017年12月。本年度在文本表达和排序学习、网络群体行为分析等方面展开了研究并形成了一系列有国际学术影响的研究成果。代表性研究成果概述如下：

## 1. 文本表达和排序学习

概率化主题模型是文本数据约简的一种有效方法，文档被建模成主题的混合分布，主题建模成词的多项式分布。针对短文本中主题模型会遇到严重的数据稀疏性问题，提出了双词概率主题模型 BTM 直接建模全局的词共现关系，利用全局词共现关系包含的上下文信息判断词的主题属性，克服数据稀疏性问题。进一步，针对短文本数据流中的突发话题，提出了一个突发话题模型 BBTM，利用双词的突发性指导突发话题的学习，无需任何后处理手段与启发式的技巧。在多样性排序模型方面，提出了一个新颖的关系排序学习的框架来求解，将多样性排序问题视作一个顺序选择过程，将排序函数定义为相关性得分与多样性得分的组合。在列表型算法 ListMLE 的基础上，提出了位置敏感的 ListMLE 算法，有

效建模了位置信息的影响。该算法可以有效改进传统 ListMLE 算法的性能。同时，理论分析也证明了该损失函数与 NDCG 型的评价准则是统计一致的。研究成果发表在 SIGIR 2014、AAAI 2015、WSDM 2015 和 TKDE 上，形成了很好的学术影响。

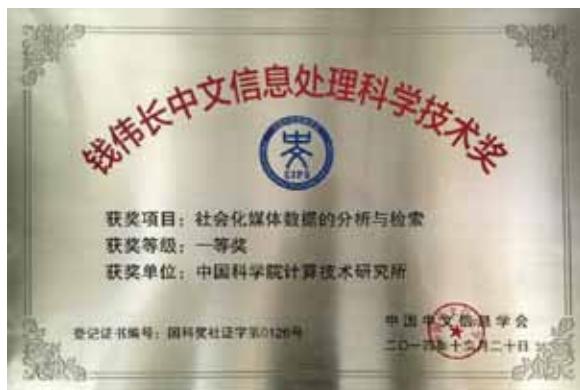
## 2. 网络群体行为分析

针对消息流行度预测问题，提出了一种基于自增强泊松过程的概率模型，建模流行度动态过程的微观机制。该模型直接建模各个消息获得关注的过程，可再现整个在线社会关系网络的流行度分布，显著超过了基于时间序列自回归的模型和对数回归模型。针对团队合作中功劳分配面临的学科间规则差异大、实际贡献难以量化等难题，提出了群体功劳分配方法，采用“被感知的贡献”作为功劳分配的依据，利用共引用关系作为度量“被感知贡献”的依据。该方法能够自适应地量化个体从团队合作中获得的功劳分配比例。针对影响最大化问题，提出直接通过寻找自洽排序来求解影响最大化问题，设计了一个迭代的框架 IMRank，在计算效率上显著超过了当前最好的贪心算法和启发式算法。研究成果发表在 SIGIR

2014 和 Sci. Rep. 上, 受到了国际研究同行的关注。研究成果发表在 AAAI 2014、Science 和《美国科学院院刊》(PNAS) 上, PNAS 论文为发表当期的六篇 highlighted article 之一, 并被 ScienceDaily 等多家学术媒体报道和转载。

研究成果“社会化媒体数据的分析与检索”, 获得了中国中文信息学会钱伟长中文信息处理科学技术一等奖; 课题负责人程学旗担任 Web

信息检索与数据挖掘领域国际学术会议 WSDM 2015 的大会共同主席; 课题负责人程学旗在组织了主题为“在线社会关系网络挖掘与分析”的专辑, 出版在《计算机学报》2014 年度第 4 期上; 项目研究成果在第二十届全国信息检索学会会议 (CCIR 2014)、第十届中国网络科学论坛上进行了介绍。2014 年底, 课题顺利通过了国家基金委组织的中期检查。



集 中 力 量 办 大 事



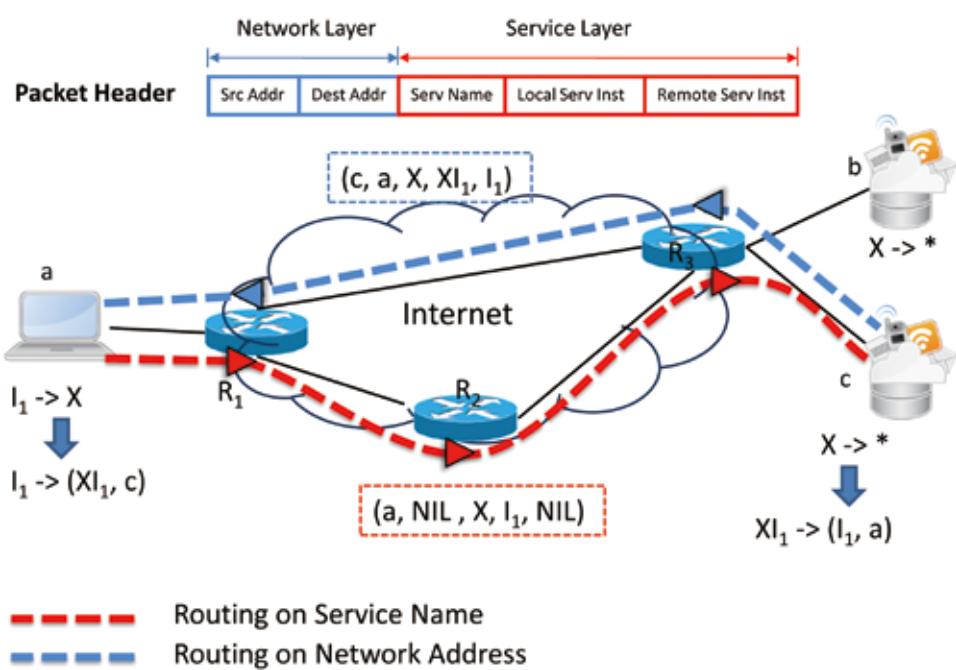
重点 / 重大科研任务进展

# 后 IP 网络体系结构及其机理探索

“后 IP 网络体系结构及其机理探索”是由计算所牵头、联合清华大学和北京邮电大学承担的国家自然科学基金重点项目（项目编号：61133015），起止时间是 2012 年 1 月至 2016 年 12 月。该项目旨在研究未来网络体系结构模型、理论基础和关键技术。项目组提出了面向服务的未来互联网体系结构 SOFIA，并在 2014 年度完整实现了 SOFIA 体系结构原型系统，研究了未来网络中的缓存 / 转发机制，提出了高效的互联网

路由查找和数据包分类算法，并刻画了互联网的资源访问模型。

在 SOFIA 中，服务获取由服务发现和数据传输两部分组成，其中服务发现使用基于服务名字的路由，数据传输使用基于网络地址的路由。服务发现的过程如下图所示。这种混合路由机制的优点在于一方面保证灵活的服务发现，充分利用临近的缓存节点，而基于网络地址的路由使用与网络拓扑一致的最短路径传输数据，提升传输效



SOFIA 服务连接基本元素及建立过程

率。项目组实现了 SOFIA 协议栈原型系统和典型应用系统（如互联网视频系统、语音通信系统等），并在未来网络试验床中部署。

考虑到未来网络中新的应用层出不穷，网络规模和链路带宽呈指数级增长。项目组深入研究了高效路由查找算法，并提出将整个路由查找按照两个维度（查找过程和前缀长度）进行分类处理，将整个路由查找按照处理过程和地址前缀长度分类。在四个平台（FPGA、CPU、GPU 和众核）上实现了这些框架算法或改进算法。实验结果表明，与已有算法相比，所设计算法可以同时实现查找时间较小（满足线速处理需求）和使用较小

的片上存储（片上存储大小较小，并且增加速度较小）。

此外，提出了路由经济学模型，利用该模型研究和分析了域间路由冲突的经济原因，路由的冲突根本原因是 ISP 之间的利益冲突。通过引入 Shapley 机制，在不同 ISP 之间建立协作解决经济利益冲突，在解决路由冲突的同时确保了 ISP 之间利益的公平性。

2014 年度，在该项目的支持下，在 ACM SIGCOMM、INFOCOM 等顶级国际会议，以及 IEEE JSAC、IEEE Network 期刊共发表（含录用）论文 20 余篇，申请国内专利 5 项。

跻身国际前沿  
关注国计民生  
引领中国信息产业



## 重点 / 重大科研任务进展

# 数据并行与线程并行合一的可伸缩处理器体系结构

本项目为国家基金重点项目（项目编号：61332009），起止时间是2014年1月至2018年12月。长期以来，指令级并行、数据级并行和线程级并行是现代处理器用来提升性能的三种并行机制，然而，为了追求高性能和通用性，处理器中指令级并行、数据级并行和线程级并行的结构通常是分别配备的，导致整个芯片占用较多资源，芯片的成本增大了。资源的过度配置和资源利用率的低下也是现代高性能通用处理器功耗过高的主要原因。根据对应用模式的简单分析可知，线程并行和数据并行等多种能力并不需要同时提供，经常可能是互相交错的。所以，为不同的并行能力提供不同的功能部件，并无必要。本项目重点研究在同一套控制和运算部件上，既能支持多线程并行执行，又能支持向量数据并行计算的新型处理器体系结构。核心是利用一组可重构的深度流水数据通路，同时支持线程并行代码和数据并行代码的执行，并可根据应用需求，在两种执行模式间动态切换。这种创新的体系结构可大大提高硬件资源的利用率，从而达到高性能、低成本、低能耗和通用性的平衡。2014年度的工作主要体现在以下四点。

## 1. 并行程序调试与错误定位

并行程序是发挥可伸缩处理器，以及传统多核处理器性能的关键途径。然而保障并行程序的正确性和健壮性，是一个难点，也是国内外研究的一个重点。我们从并行程序调试和错误定位两个角度展开了研究。

a) 并行程序中数据竞争导致的不确定性，使得并行程序调试极为困难。我们提出了一种新的错误重现技术，利用核间局部时钟的差值区间，降低了并行程序错误重现和离线分析的有效方法。同时还提出了核间局部时钟差值的度量方法，改变了人们普遍认为该差值无法度量的看法。b) 在错误定位方面，我们针对并行程序，提出了一种利用正确执行历史来定位程序中的错误根源，取得了较好的效果。

## 2. 线程并行和数据并行合一体系结构中缓存一致性

线程并行和数据并行合一的体系结构中，高效且低开销的维护缓存一致性是一个重要问题。本研究提出了三种方法解决稀疏目录的问题：a) Victim 目录（Victim Directory）。该方法是在主稀疏目录结构上附加一个小的全相联目录结构。实验证明该技术是一个减少稀疏目录替换的有效且低开销的方法。b) 合作式目录存储

(Cooperative Directory Storage, CDS)。选取合适的稀疏目录替换算法和采用 victim 目录均仅能减少稀疏目录的冲突替换，但无法减少稀疏目录的容量替换。在稀疏目录信息分布不均匀时，热点稀疏目录节点很容易发生容量替换。针对这个问题，我们提出了 CDS 技术，该技术可以动态细粒度地将末级缓存数据阵列的容量分配给末级缓存和稀疏目录，从而动态增加热点稀疏目录节点的容量。c) 海绵目录 (SpongeDirectory)。为了提高稀疏目录结构在未来超大规模众核系统中的可扩展性，我们提出了 SpongeDirectory，该技术是一种采用了多层次 memristor RAM 的稀疏目录。每个 SpongeDirectory 目录块均可以根据需求动态地调整其层数，从而改变其存储的容量。

### 3. 线程并行和数据并行合一体系结构的访存结构设计

在共享存储的线程并行和数据并行处理器上并行处理应用时，其主要的性能瓶颈来自于内存管理。而且，针对运行在共享存储平台下的并行应用，动态内存管理领域的研究主要集中于设计和实现高效的并行内存分配器，而并行内存分配器本身却受到并行度的限制，在并行度比较大的情况下无法实现高效的内存管理。本研究针对导致线程和数据并行应用内存管理性能下降的两个核心原因，分别从降低存储需求和减少存储频次两个方面，来研究内存管理的优化方法。首先，我们提出一种降低存储需求的功能流水机制，通过功能流水可以有效的避免中间数据结果的累积，可以减少绝大部分的中间数据的存储开销。另外，我们基于排队论和线性规划，提出一种减少存储频次的优化方法。基于排队论预测出系统的存储需求，然后根据存储需求和系统资源限制，利用线性规划的方法计算出最优的预分配存储空间大

小，并对这部分存储空间采用集中一次分配的策略，来代替大量离散小块内存的频繁申请，以此消去绝大部分的内存申请操作。

### 4. 线程并行和数据并行合一处理器的能耗优化

能耗优化的面积、功耗、带宽分配问题是多核处理器在未来发展中能否解决“暗硅”问题，延续摩尔定律的关键。本研究针对使用 DVFCs 技术的多核处理器提出一种能耗-性能建模方法，能够分别在能耗和性能中分离出与频率、核数相关和无关的组成部分，准确表达不同并行负载的能耗-性能随核数-频率的变化规律，相比已有的系统级建模方法能够更精确的反映核数、频率对能耗-性能权衡关系的影响。

基于能耗-性能在核数-频率空间的分布特点，研究提出了在 LM+PO+MO 模型所构成的核数-频率空间中搜索最优性能、能耗所用核数、频率的执行方法；同时，针对实测执行最优核数-频率级别搜索的巨大开销，提出一种基于可行方向法的最优核数-频率级别搜索方法，相比已有的二分爬山法和启发式爬山法在执行次数、执行时间和能耗开销方面都具有一定优势，并在核数和频率级别增加时具有更好的可扩展性。在充分研究了以上同构多核处理器能耗-性能制约关系的基础上，还研究了异构处理器上面向能效-性能优化的带宽划分方法，通过建立引入排队延迟的异构处理器能耗-性能模型，结合经验参数分析片外访存带宽资源分配对总吞吐率和能效的影响，阐明了异构多核性能-能效增益随主协处理器 cache 失配率之比率和可用峰值带宽的紧缺程度的变化规律，最终使用所提出的模型推导出最优性能和最优能效的带宽划分解析式，分析得出相比自然划分的性能-能效收益。



## 重点 / 重大科研任务进展

# 动态可配置的压缩感知成像系统

本项目是国家自然科学基金科学仪器基础研究专项（课题编号：61327013），起止时间是 2014 年 1 月至 2017 年 12 月，前期得到中科院装备研制项目的支持。本项目旨在面向压缩感知成像研究的实际需求，提供任意值测量矩阵的配置，用以验证不同的测量矩阵在光学成像中的实际效果；实现图像采集，恢复，呈现一体化，让研究人员高效快捷地验证自己的恢复算法和成像效果。2014 年研究进展如下：

## 1. 确定整个系统的设计方案

采用模块组件设计思想，以压缩感知成像理论为基础，以高分辨率，高速，可配置为目标，设计了无透镜 / 透镜压缩感知成像系统方案。整个系统由采集箱、重建服务器和控制客户端组成，模块之间使用网络连接交换数据。

## 2. 实现图像采集子系统观测矩阵设置与切换

提出了一种利用偏光片改造的液晶屏实现多级观测矩阵配置的方法，减少透过液晶光线的散射，实现高质量可分级的观测矩阵设置。通过显示控制接口使用的 VGA 信号转换模块，

将目前的当前刷新率提高到 60Hz。完成观测矩阵接口开发，提供二次开发接口和能力。

## 3. 实现单点压缩感知图像采集

采用可配置思想，支持普通光敏电阻、硅光电管、RGB 颜色传感器和 CMOS 摄像头采集。针对电源波形中的工频谐波，与观测矩阵负相关的超低频谐波，单次采集产生的白噪声以及大跳变脉冲信号干扰分别进行有针对性的滤波处理，提高图像采集质量。其中，普通光敏电阻、硅光电管、RGB 颜色传感器使用偏置电路接法，将所得信号通过 10000 倍的模拟放大电路放大后，由 ADS1256 24 位精度数模转换器采集。数模转换器最高支持每秒 3 万个采样，通过 SPI 接口将数据返回主控。CMOS 摄像头使用 Camera interface 500 万像素摄像头，可对白平衡等各种采集参数进行配置。

## 4. 完成部分恢复重建子系统设计实现

实现了一种基于 GPU 的 TVAL3 并行恢复算法，使得当前系统具备恢复高分辨率图像的能力，与传统算法相比，将 1920X1080 图像的恢复速度提高了 100 倍以上。利用纹理分布

的不均匀性和多变性，难以用一个稀疏基来完成整个图像的稀疏表示的特点，提出了一种基

于局部稀疏基的恢复算法，有效提高了图像重建质量。



图 1 系统控制流程图

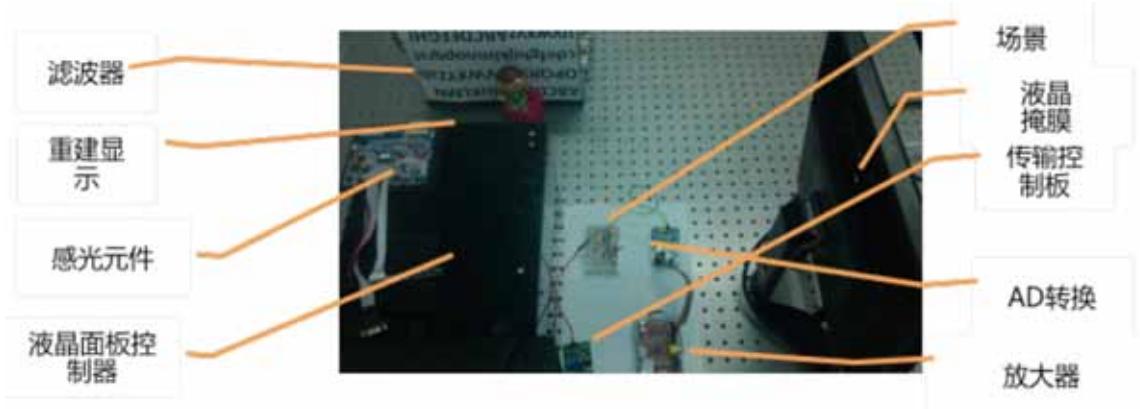


图 2 SOFIA 服务连接基本元素及建立过程



## 重点 / 重大科研任务进展

# 理论计算机科学

本项目为国家基金委优秀青年基金项目（课题编号：61222202），起止时间是2013年1月至2015年12月。2014年度研究工作围绕着问题的难解性，分别对判定树复杂度中的sensitivity猜想、解线性方程组等线性代数问题的通信复杂度等问题展开了深入的研究，20年来首次改进了关于sensitivity猜想的下界和指数上界，在被动通信模型下证明了求解线性方程组的最优的通信复杂度下界。取得的重要研究成果如下：

## 1. 判定树复杂度中sensitivity猜想的研究

Sensitivity复杂度最早由图灵奖得主Cook等提出，是判定树复杂度研究中的一个重要度量。ACM博士论文奖得主Nisan等在92年提出关于Sensitivity vs. Block Sensitivity的猜想，这一猜想断言：对于任意布尔函数  $f : \{0,1\}^n \rightarrow \{0,1\}$ ，其sensitivity复杂度和block sensitivity复杂度之间是多项式量级相关的，即  $bs(f) \leq s(f)^{O(1)}$ 。关于这一猜想的最好下界是Rubinstein在95年得到的结果  $bs(f) = 1/2 s(f)^2$ 。我们在研究工作中构造了新的复合布尔函数实例，20年来首次将猜想中下界的参数从  $1/2$  改进到  $2/3$ 。关于上界方面，

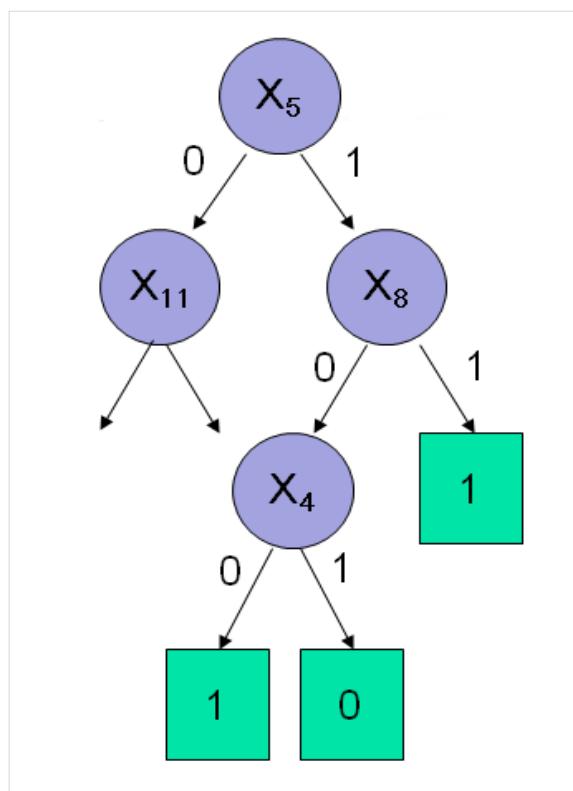


图1 判定树模型

我们引入等周不等式、傅立叶变换等数学工具，将猜想中的上界从  $e^{s(f)}$  改进到  $2^{s(f)}$ 。此外我们还研究了函数的多项式次数  $\deg(f)$  与 sensitivity 之

间的关系，证明了  $\deg(f)^{1-o(1)} \leq 2^{s(f)}$ 。这一工作最近发表在欧洲理论计算机领域著名会议 ICALP 2014 上。

## 2. 求解线性方程组等问题的通信复杂度研究

通信复杂度最早由图灵奖得主姚期智教授在 78 年提出，在过去的三十多年里，通信复杂度已经成为理论计算机科学中最重要的证明下界的工具之一，它也是研究流式计算空间复杂度的重要工具。我们在研究工作中讨论了在被动通信模型下求解线性方程组等线性代数问题的通信复杂度，我们提出了一种一般性的将多人通信问题归约到两方通信问题的方法，证明了  $s$  人的多方随机通信复杂度至少是  $s$  倍的两方通信复杂度。证明的关键是利用了线性代数问题的对称性，我们构造的一个对称的困难实例的分布；另外我们还给出了多个新的两方通信复杂度下界，结合上述两方面，使得我们能够证明求解线性方程组等问题最优的通信复杂度下界。我们的下界还适用于一些带有“threshold promise”的问题，例如判定矩阵的秩是  $a$  还是  $b$ 。这一工作已经发表在 DISC

2014 会议上。

## 3. 社会网络中近似算法的研究

我们围绕着 independent cascade 模型研究了社会网络中信息传播的效果。在 IC 模型下，信息被传播的效果是随机地，因此，人们希望信息能够被传播到一定比例的节点数的概率最大。我们在研究中首次讨论了如何计算最少需要的初始种子个数，使得传播到一定比例的节点数的概率超过给定的输入。传统的影响力最大化往往关注传播到的节点个数的期望，这里最大的困难在于概率保证下的影响力最大化不再是次模 (submodular) 函数，因此不能通过贪婪算法得到好的近似算法。我们在工作中首先证明了这个问题没有比  $\log n$  更好的近似算法（除非  $NP=P$ ）。我们设计了一个近似算法，该算法对有有限制的二部图的情况，能保证输出  $|S| \leq O(\log n \cdot |S_{opt}| + \sqrt{n})$ ，同时，这个算法在三个不同类型的数据集上 Wiki-Vote, arXiv, Flixster 均得到了很好的实验效果。这一工作已经发表在 ACM SIGKDD 2014 上。



图 2 社交网络



重点 / 重大科研任务进展

## 计算机体系结构

本项目为国家基金委优秀青年基金项目（课题编号：61222204），起止时间是2013年1月至2015年12月。面向神经网络的体系结构研究在技术上取得的系列突破得到了国际同行的高度认可，多次获得顶级会议最佳论文，在一些特定领域（尤其是深度学习）上相对传统芯片在性能功耗比上取得了显著的优势。

2014年在前两年的神经网络体系结构研究理论基础上，指导工程上进行了先导专项可重塑处理器测试流片设计、验证片系统的搭建以及软硬件解决方案的提出和验证，为后面进行正式流片和大规模系统搭建提供了宝贵和必要的基础条件。

目前工程方面的主要进展有：（1）完成可重塑处理器版图设计，采用65 nm工艺，包含不少于4个可塑加速核，芯片的峰值性能达到1000 GOPS/W；（2）完成可重塑处理器RTL和网表模拟仿真验证，可正确运行CNN、DNN和MLP神经网络应用；（3）在FPGA平台上完成可重塑处理器的原型验证，可正确运行CNN、DNN和

MLP神经网络应用；（4）完成面向领域的可重塑处理器编译器，可将CNN、DNN和MLP神经网络算法描述自动转化为可重塑处理器可执行的二进制码。

相关研究成果发表在2014年的各类顶级会议上。在体系结构顶级会议ASPLOS上获得2014年best paper。在体系结构顶级会议MICRO上获得2014年best paper。在体系结构顶级会议ISCA发表1篇论文。发表情况如下。

理论研究上通过硬件对神经网络结构的适配，提出了世界上学习速度最快的深度学习芯片。在32位运算下，达到1000 GOPS/W的性能功耗比。支持任意规模多种神经网络算法。以1/10的功耗达到同期的主流CPU(Xeon E5-4620)的100倍性能，以1/100的功耗达到同期的主流GPU(K20M)同样的性能。

相应理论成果被会议总结评价为“*This paper appears to be the culmination of a significant engineering effort in this area*”。

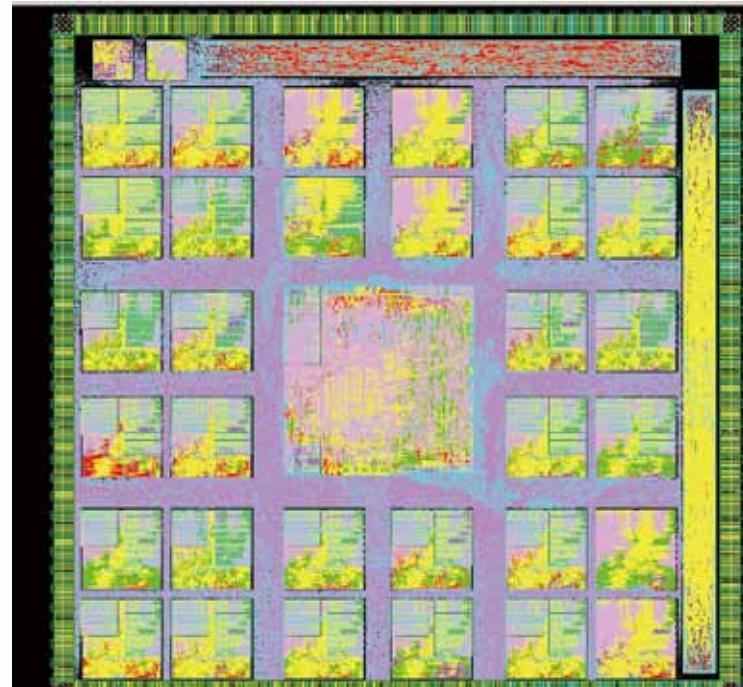


图 1 结构视角下的版图



图 2 ASPLOS 2014 最佳论文获奖证书



## 重点 / 重大科研任务进展

# 视觉模式分析与识别

本项目为国家基金委优秀青年基金项目（课题编号：61222211），起止时间是2013年1月至2015年12月。项目团队在2014年度继续围绕视觉模式分析与识别相关问题开展研究，重点在基于层级深度模型的人脸模式分析与识别研究、面向视频分析的视觉建模与识别研究以及跨视、跨域的视觉模式识别问题研究等方面展开工作，共发表了17篇标注学术论文，其中CCF A类论文6篇（包括国际刊物IJCV和IEEE TIP论文各1篇，国际会议CVPR论文4篇），领域顶级会议ECCV 2篇，领域顶级会议NIPS论文1篇。1篇论文获得ACCV workshop最佳论文。

本年度较为突出的成绩包括：1) 深入研究了深度层级模型，并将其引入人脸特征点定位、图像超分辨率、姿态鲁棒的人脸特征学习等方面，取得了较大进步，相关论文分别发表于领域顶级国际会议ECCV14和领域CCF A类国际会议CVPR14；2) 深入研究了基于流形的视频表示及分类问题，提出了通用流形模型、无监督通用流形对齐和异质空间流形距离测度学习等新方法，相关论文分别发表于领域CCF A类国际会议CVPR14和领域顶级国际会议NIPS14；3) 利用

以上理论和方法，研究团队参加ACM国际会议ICMI2014组织的表情识别竞赛(EmotiW2014)，取得了第一名的好成绩。典型的研究成果介绍如下：

## 1. 基于层级模型的人脸模式分析与识别研究

深度层级模型在语音识别和图像识别领域均取得了重要进展。在过去一年里，我们将深度模型引入人脸模式分析与识别，取得了很好的结果。主要包括四个方面的工作：1) 在面部特征点定位上引入深度编码，提出了一种由粗到细的多阶段深度非线性人脸形状提取方法CFAN。在该方法中，我们将面部特征定位建模成直接从人脸纹理特征预测人脸形状的“非线性回归”问题，通过级联多个由栈式自编码网络(SAN)实现。在LFW、HELEN和AFW三个公开的数据集上实验表明CFAN可以取得优于当前最好算法SDM的性能。该工作发表于领域顶级国际会议ECCV2014。2) 将深度模型引入姿态不变的人脸识别问题。姿态变化导致的人脸表观变化是一种复杂的非线性变化。为此，我们提出了一种栈式渐进自编码(SPAE)神经网络模型。该方法使用多个浅层的神经网络进行建模，逐步实现侧面图

像到正面图像的平滑变换。该方法在 CMU Multi-PIE 人脸库上的实验表明，该方法取得了 state of the art 的识别结果。该工作发表于领域顶级国际会议 CVPR2014(CCF A 类)。3) 将 CNN 特征用于人脸表情识别。在 ACM ICMI 2014 举办的表情识别竞赛 EmotiW' 14 中，我们将 CNN, SIFT 和 HOG 三种特征通过三种集合模型(子空间、协方差和高斯分布) 和三种分类器 (SVM、LR 和 PLS) 进行融合。并在主办方的盲测试集合上取得了 50.37% 的识别率，优于其他 8 个参赛算法。该工作应邀发表于 ICMI2014。

## 2. 面向视频分析的视觉建模与识别研究

视频中的目标识别或者动作、表情识别是计算机视觉中的关键问题之一。解决这类问题的关键在于视频的表示，项目团队以流形作为视频的表示模型，研究了流形的统计建模、流形对齐以及在流形空间的测度学习等，并取得了以下主要进展：1) 提出了通用流形模型的概念，并将其应用表情识别。我们将每个表情序列建模为一个时空流形，并对这些流形进行统计建模，形成适用于所有表情的通用流形模型 UMM，从而使得每个表情序列均为 UMM 的一个参数化的实例。这样就可以将表情序列建模为 UMM 的参数，并进而实现表情的分类和识别。该工作发表于领域顶级国际会议 CVPR2014 (CCF A 类)。2) 提出了一种通用的无监督流形对齐方法 GUMA。对于相关而结构相似的流形，我们将流形对齐问题形式化为一个显式的整数优化问题，其中联合考虑了待匹配流形的结构匹配、各自结构保持以及对应点在互空间中特征的相容性等约束条件，较好的实现了完全无监督的流形结构挖掘和对齐。该工作发表于领域顶级国际会议 NIPS2014。3) 针

对点到流形的距离计算问题，提出了一种欧式空间到黎曼流形的测度学习方法。为了实现欧式空间和黎曼流形这样的异质空间之间的测度学习，我们采用流形上定义的核方法，以再生希尔伯特空间为桥梁，将二者迁移到同质空间从而实现测度学习。在静态人脸到视频人脸的识别问题上的实验表明，该方法取得了显著优于已有方法的效果。该工作发表于领域顶级国际会议 CVPR2014 (CCF A 类)。

此外，我们还针对视频中的人脸识别和检索问题，提出了基于联合稀疏表示的人脸识别方法以及紧凑视频编码方法。

## 3. 跨视、跨域的视觉模式识别问题研究

跨视、跨域的视觉模式识别问题是一类典型的视觉计算问题。我们以不同人群的人脸识别、跨姿态的人脸识别等问题为实例对此开展研究，完成了以下工作：1) 提出了一种以公共子空间桥梁的源域目标化方法，用于不同人群之间的人脸识别模型适应问题。针对源域有监督、目标域无监督且与源域分布不同的情况，我们提出的方法试图将源域样本表示为目标域样本，同时保留源域的监督信息，从而实现目标域上的有监督学习。该工作发表于 CCF A 类刊物 IJCV。2) 提出了一种变形位移场概率建模方法，用以实现跨姿态的人脸识别。该方法的基本思想是要充分利用人脸姿态变化的 3D 几何基础，但又要避免复杂而高代价的 3D 重建，为此我们将 3D 几何关系简化为变形位移场，考虑其统计模型，即建立不同姿态人脸之间的几何形变关系的概率模型，以便隐式的实现非正面人脸到正面人脸的变形，进而实现跨姿态的人脸识别。该工作发表于 CCF A 类刊物 IEEE Trans. on Image Processing。



## 重点 / 重大科研任务进展

# 多媒体内容分析

本项目为国家基金委优秀青年基金项目（课题编号：61322212），起止时间是2014年1月至2016年12月。2014年度，在场景图像分类以及面向移动和智能终端的图像识别等方面开展了多项深入研究，取得了一系列成果，圆满完成了预期目标。下面介绍本项目本年度取得的突出进展。

提出了基于语义流形联合空间与多特征上下文建模的场景识别方法。语义流形是一种描述图像概念分布的概率空间。基于图像的视觉特征，利用提前训练好的概念分类器，可将图像表示成特定概念的概率分布，该概念概率分布空间即为语义流形。一幅图像可以提取多种视觉特征（可以为颜色，梯度，形状特征），可以得到多种概念概率分布，而这些概念概率都分布在相同的语义流形中。对于某一图像，由多种视觉特征得到的多概念分布之间的关系，即为语义流形中该图像的多特征上下文关系。同时，一幅图像可以划分为大量局部小块，相邻小块的概念分布间关系，即为语义流形中该图像的空间上下文关系。基于语义流形空间，我们提出了一种利用马尔可夫随机场对空间和多特征上下文关系建模的方法，以

增强图像语义描述，并将该语义描述用于场景图像分类，在大规模的场景图像的公开数据集上得到了较好的分类正确率。以该方法为基础的论文被国际会议 CVPR 录用。

提出了基于深度信息和多特征融合的手持物体识别方法。日常生活中，物体识别在人机交互和多媒体检索中有非常多的应用，而由于类内多样性和类间相似度，仅基于颜色信息 (RGB) 的物体识别仍然面对很大的挑战。同时由于在人与人，人与机器的交互中，使用手操作物体非常普遍。但是一般基于 RGB 的手持物体识别需要滤除背景噪音，很难准确定位物体。这里我们使用基于颜色和深度信息 (RGB-D) 的手持物体识别技术。首先利用 Kinect 获取人体骨骼图像，RGB 和深度图像信息。利用骨骼定位及深度区域增长的方式对手持物体进行分割，获取物体同时滤除背景噪声。然后对分割后的物体提取基于点云和深度学习特征并进行异质特征融合，以达到特征互补，提高识别准确率。该技术应用在智能技术集成平台机器博士中，手持物体识别技术不但可以增强机器博士与人的交互性，同时对进一步理解用户的状态、意图

及需求有重要意义。以该方法为基础的论文发表在期刊 JCST 上。

提出了基于地理上下文模型的手机图像识别方法。近年来,手机和移动互联网越来越普及,手机上传图像的数量急剧增加。对这些图像内容的准确识别,是各种网络应用以及许多潜在手机图像应用的必要技术要求。我们研究了基于地理上下文模型的手机图像识别方法,具体目标为手机拍摄的餐馆菜品图像识别。与大多传统的仅基于图像视觉内容的识别技术相比,我们的方法有效利用了地理上下文信息对图像内容的约束

性,提高了识别准确率。特别的,我们提出了地理约束的视觉分类模型——地理局部化模型,将餐馆的地理坐标及其包含的菜品类别等上下文先验信息,用于视觉分类器的训练,使这些分类器从根本上对地理信息更有针对性。在对手机拍摄的查询图像进行预测时,这些分类器再根据拍摄时所在的地理坐标进行自适应组合,协同进行分类判别。该方法在多个城市餐馆菜品图像的数据集上取得了较好的识别准确率,且可扩展性较高。以该方法为基础的论文被期刊 IEEE Tran. on Multimedia 录用。

## 坚持“基础性、战略性、前瞻性”的三性原则



## 重点 / 重大科研任务进展

# 海云计算系统关键技术与系统研究

“海云计算系统关键技术与系统研究”为中科院先导项目（课题编号：XDA06010400），起止时间是2012年1月至2016年12月。

以云计算模式为核心的现有信息服务系统在数据处理性能和能耗方面无法应对由终端规模急剧增长所带来的挑战。为适应“人-机-物”三元融合的趋势，以建立“感知中国”信息网络为背景，本任务提出海云计算系统的理念和关键技术，搭建海端和云端设备共同参与任务执行的计算系统，用以构筑一个海云协同计算和数据处理环境，从而为实现计算处理系统千倍效能比提升的目标探索一条有益途径。

海云计算系统在2014年完成了若干关键技术的突破与仿真验证，研制了两套概念原型系统，形成了一定规模的专利群，并在国际学术界获得了较高认可，部分成果已转移至华为等企业。从技术层面来看，本研究任务完成了海云服务器简约体系结构各功能模块的定义和设计，搭建了基于FPGA的16节点概念原型系统，完成了可扩展至千亿文件的海云存储系统16节点站点原型。另外，利用硬件对神经网络结构的动态适配机制，可重塑处理器的仿真性能功耗比达到1016

GOPS/W（每瓦特每秒十亿次操作数），并在FPGA平台上通过了功能验证。

海云服务器在2014年完成或优化了海云处理器核心模块自研芯片、第二代消息式内存、可扩展数据中心网络及资源管理等关键技术的软硬件实现。其中，基于专用化硬件及片上总线直接扩展等技术，首次实现了节点间应用级别远程主/从设备类型资源的直接访问，且性能优于主流商用系统，这将在较大程度上提升海云计算系统中数据处理资源的利用率和效能比。另外，海云服务器研究任务完成了第二代消息式内存和级联内存缓冲器的路由模块，并且实现了基于DcnFPGA的16节点数据中心网络原型。海云服务器概念原型系统运行内存数据库类应用的效能比达到x86商用服务器（基于Intel Xeon E5645处理器）的三倍。

可重塑处理器完成了包含4个可塑加速核的65纳米验证芯片设计和验证，生成了流片用版图。仿真结果表明，通过硬件对神经网络计算任务的动态适配，可重塑处理器运行神经网络应用的峰值运算效能比超过1000 GOPS/W，在世界上率先突破了处理器芯片级每焦耳万亿次运算的难题。

该项工作在 2014 年被计算机体系结构领域顶级会议 ASPLOS 和 MICRO (CCF A 类) 录用，并均获得最佳论文奖。此外，可重塑处理器任务搭建了 FPGA 原型平台，可正确运行 CNN、DNN 等神经网络算法。

海云存储系统完成了云端站点系统原型样机，突破了可扩展支持 ZB 级数据的关键技术。元数据集群基于大粒度子卷进行负载分布、定位、均衡，突破全异步分布式元数据一致性协议关键技术，

可扩展至数百台元数据服务器；针对元数据访问延迟，提出并完成了客户端缓存中全异步元数据存取机制；搭建了包含 16 台元数据服务器原型系统，实测支持 100 Billion 文件、元数据平均访问延迟低于  $50\mu\text{s}$  @36 万次，规模比肩谷歌文件系统，元数据访问延迟降低 1-2 个量级。广域存储实现了北京 - 天津站点间全局命名空间。网络盘阵可容忍节点级故障，经完善已达产品级成熟度，在中科院声学所、网络信息中心等单位进行部署。



图 1 海云服务器原型系统中的四类自研设备

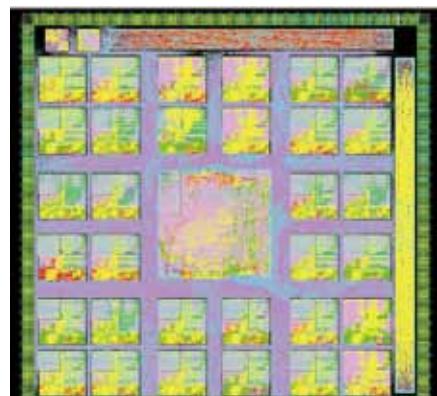


图 2 结构视角下的可重塑处理器 ICC 版图

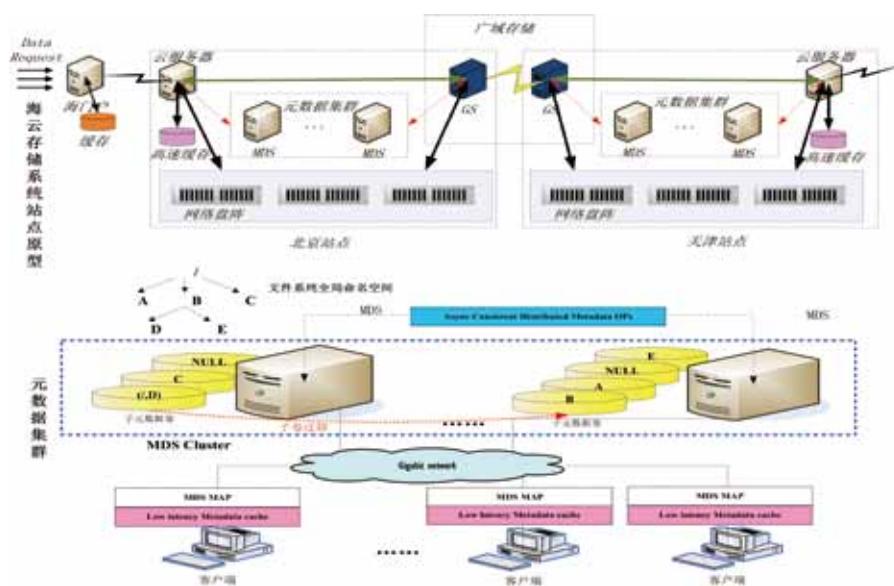


图 3 云端站点系统原型样机与元数据集群



## 重点 / 重大科研任务进展

# 面向未来网络的可编程网络测试仪表

“面向未来网络的可编程网络测试仪表”是计算所牵头、联合网络信息中心承担的中科院重大装备项目（课题编号：Y2201229），起止时间是2012年1月至2014年10月。旨在研究可编程高性能网络测试仪表体系架构，重点解决灵活可编程的数据包处理、可扩展协议配置等关键技术，研制面向未来网络的可编程网络测试仪表。2014年度，项目组完成了关键技术攻关和测试仪表的研制，并在实际环境中进行了测试。

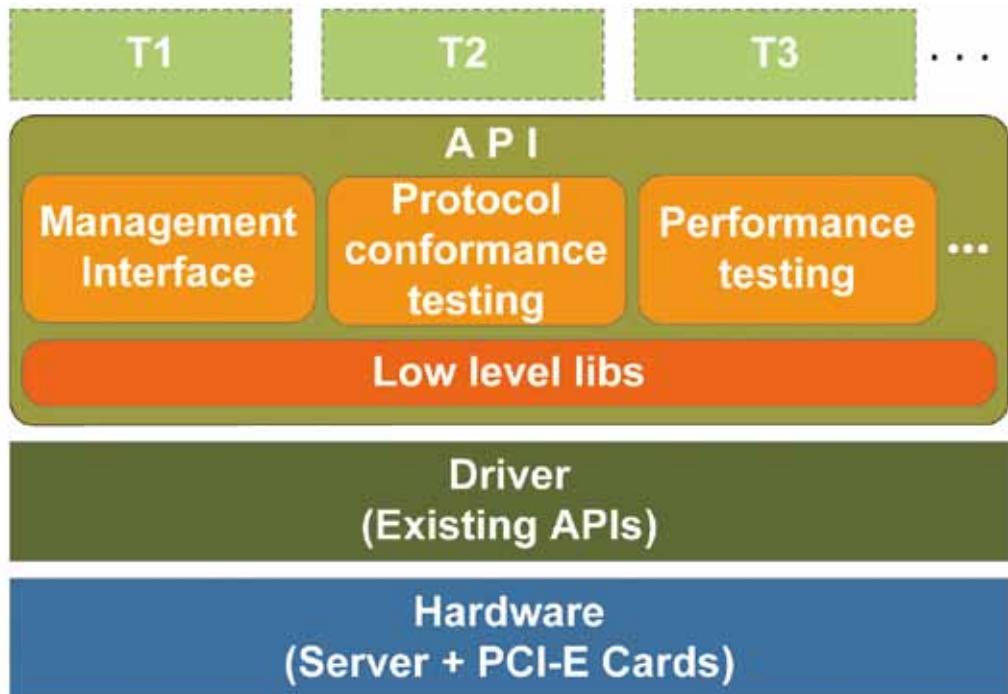
可编程测试仪表基于高密度可编程的网络测试硬件平台，通过软硬件可编程和优化技术全面支持未来网络体系结构、协议和关键设备的测试。该仪表充分考虑未来网络发展方向的不确定性，通过灵活高效的数据包处理方法，支持各种未来互联网体系结构和协议的功能、性能测试。可编程仪表的层次结构如下图所示，从下到上由四个层次组成：硬件层，驱动层，平台API层和上层

测试应用软件。

其中，硬件层由服务器平台+专用板卡组合实现，研制了支持千兆、万兆以太网和10G POS接口的高性能数据捕获卡。驱动层封装了直接对于底层硬件的操作，并且提供了丰富的API供上层软件使用。API层包括管理API、性能测试API、协议一致性测试API以及驱动层API扩展。最上层是不同的测试用例，即测试仪表应用软件及GUI操作界面等。

通过层次性结构，项目组构建一整套能够提供包括应用测试、协议测试、性能测试在内的面向未来网络的网络测试服务平台，为网络规划管理、故障分析、性能分析、安全分析、用户业务分析等提供服务。

下一步，项目组将着重把研制的可编程测试仪表在用户单位应用和测试，并在未来网络研究过程中逐步推广应用。在此基础上，准备项目验收测试，并完成项目验收。



可编程网络测试仪表研发平台架构图

从 优 秀 到 卓 越



## 重点 / 重大科研任务进展

# 瓦楞纸生产线智能生产管理与工艺支撑系统

“瓦楞纸生产线智能生产管理与工艺支撑系统”是由计算所牵头的中国科学院科技服务网络计划（项目编号：KFJ-EW-STS-039），起止时间为2014年1月至2015年12月。2014年通过了项目的中期检查，并得到了专家的一致认可。本项目利用物联网技术对分布在全国的瓦楞纸生产线的工作数据进行远程数据采集、传输和存贮，利用数据挖掘技术对采集的数据进行数据分析和挖掘形成成熟工艺库，并通过瓦楞纸生产线智能管理系统直接从工艺数据库中提取工艺参数发送到瓦楞纸机进行生产。本项目有效减少人工经验对产品品质的影响，提升成品率，降低能源消耗，同时带领传统制造业由过去单纯卖机器设备向卖服务的新型“云”制造的商业模式转变。

本课题本年度的主要进展如下：

## 1. 工艺支撑管理系统 CAPP 开发

工艺支撑管理系统是构建在瓦楞纸生产线现有基础架构之上，是实现设备控制的主要手段，也是设备与智能系统之间的信息节点，主要完成对工艺数据的采集、分析和挖掘，从而寻找并固化最优工艺。

### ◎ 数据片段和数据挖掘

工艺管理系统不仅要接收PLC发过来的生产数据，还要与客户原有的生产管理系统进行对接，获取到它的订单数据，如订单号、产量、能耗等等。这样，当服务器将数据存储到实时数据库的时候，订单数据就能够和生产的数据关联起来，从而为下面的庞大数据流分割成多个数据片段做好铺垫。

当庞大的数据经过系统的数据处理、挖掘之后，从开始连续的、无规则的数据变成了有序的数据片段。系统通过对数据片段的分析，从而更快地挑选出最优工艺参数的组合。

### ◎ 最优工艺的获取及确定

系统根据采集的工艺数据片段进行自动分析，针对不同的纸张材质、设备运行速度，提取出一组工艺并保存在本地，系统根据设备运行过程中不同速度段的平稳性，不同速度段按不同指标找出多个平稳速度，根据控糊量的大小、温度的高低等指标得出最优工艺数据包，将其固化到PLC中，减少对现场工艺师傅的依赖，提高工作效率。

### ◎ 最优工艺的自动发送

当客户根据订单开始生产前，系统自动寻找该订单对应的纸张材质对应最优工艺，将最优工艺包自动发送到设备的PLC缓冲区中，生产线上

的员工点击确认，即可快速生产，提高生产效率。

本项目申请发明专利一项：一种基于机械装备的远程售后故障诊断系统（申请号：201410098974.8），和一项计算机软件著作权：瓦楞纸生产线工艺支撑管理系统。

## 2. MESTALK1.0 的开发

MESTALK 是面向智慧工厂开发的离散制造车间管理和数据交互的技术标准，有效实现生产管理系统对智能设备的可靠指令下达，工艺对接、数据采集和传输以及信息安全等功能，包括：通讯格式，数据交互接口定义，与 PLC、工艺管理

系统的通讯交互，并能对 PLC 数据的读写，工艺管理系统分析最优工艺后，通过 MESTALK1.0，将工艺数据包下发到 PLC。

### 3. 示范推广

本项目在佛山高明客户的瓦楞纸示范生产基地实现了示范应用，目前已经试运行3个月，运行情况良好，得到客户的一致好评。同时本项目参加了全球顶尖的工业自动化展德国纽伦堡国际电气自动化系统及元件展览会（SPS IPC DRIVES 2014），向全球的观众展示来自中国的创新科技成果，彰显计算所的技术力量。



图 1 瓦楞纸工艺管理系统界面



图 2 瓦楞纸生产线应用示范现场



图 3 纽伦堡展会现场



## 重点 / 重大科研任务进展

# 面向工业产品全生命周期的 仿真公共服务平台

“面向工业产品全生命周期的仿真公共服务平台”是中国科学院科技服务网络计划（项目编号：KFJ-EW-STS-045），起止时间为2014年5月至2015年12月。课题的主要目标是：针对中小企业转型升级的现实问题，搭建面向工业产品的全生命周期仿真公共服务平台，对工业产品全生命周期（即从产品需求到原形设计、再到使用维修以及报废的全生命历程）进行仿真分析，解决企业在可持续发展的基础上创新、开发和生产新产品所面临的重要问题。

2014年，项目主要进展有：

## 1. 工业仿真云服务平台

通过构建工业设计的资源数据库、开发系列仿真分析工具软件接口、建设产品全生命周期仿真服务能力，从而构建一个面向工业产品全生命周期仿真的云服务平台，主要包括：资源建设、技术开发以及服务能力建设三个层面（如图1所示）。

面向工业产品的全生命周期仿真平台框架在资源建设（数据资源、软硬件资源、智力资源）的基础上（图3），发展云服务技术（结构优化技术、设计与仿真无缝集成技术、快速成型技术、

虚拟展示技术、虚拟体验技术、维修性分析技术），提供数据资源服务、软硬件资源服务、产品设计、仿真服务以及智力资源服务。

## 2. 面向工业设计的公共资源池

建设用于支撑工业产品全生命周期仿真的公共资源池，包括数据资源、软件资源、硬件资源以及人力资源四大部分。其中数据资源是开展相关工作的基础，包括标准规范库、零件3D模型库、领域知识库、经典设计案例等数据资源；软件资源包括：计算机辅助设计、计算机辅助工程、人机工程分析、虚拟体验等工具软件；硬件资源包括专用高性能计算机、3D扫描设备、快速成型设备等；智力资源主要是指整合国际、国内与工业设计相关的人力资源，为企业的产品设计提供涵盖多领域、全过程的服务。

## 3. 面向工业产品生产的快速成型平台

与传统制造相比，快速成型平台（图2）有如下优势：一是可以快速制造形状复杂的物体且不增加生产成本；二是具备一体成型的特征，无需组装，简化了产品的制造程序。结合产品逆向工程，还可实现对产品快速三维拷贝，用于产品的再设计、结构解剖、深化和再创造。



#### 4. 面向工业产品交易的三维拍摄云平台

“奥视景”三维拍摄云平台作为仿真云平台的重要组成部分，主要由图像采集设备、3D 自动成像系统、三维展示软件三个部分组成；通过 3D 自动成像系统控制图像采集设备采集物品图像信息从而生成三维展示图，支持 GIF、Flash 格式及 HTML5 等。三维展示图可存储运用于互联网、PC、iPad、LED 大屏幕等，便于物品的互联网展示及线下电子展示。





## 重点 / 重大科研任务进展

# 华为高通量服务器研制

本项目是由计算所和华为技术有限公司签订的“中科院计算所 - 华为联合实验室”A类项目，于2012年4月正式启动，并于2014年11月通过第二期审核。本项目的总体目标是基于新型的负载处理器、异步内存、一体化存储和高性能网络，配以异构操作系统和相对应的编译和开发环境，完成面向高通量负载的软硬件一体化系统。

截至目前，本项目共计发表 / 录用论文 70 余篇，关键技术成果主要发表在 ASPLOS、ISCA、HPCA、CGO、PACT、IISWC、ISPASS、Euro-Par、IPDPS、DATE 等国际会议及 ACM TACO、软件学报、计算机学报、小型微型计算机系统等国际和国内学术刊物上。累计申请并受理发明专利 248 项。

本项目在 2014 年度取得的主要进展如下：

**1. 基于 X86 的 16 节点原型系统及数据中心软件栈。**该原型系统用于对 DataCenter 3.0 体系架构资源池化的功能验证和性能评测。包括 4 个自研的定制电路板作为存储节点，以及 12 个计算节点，其中 4 个为 x86+GPU 异构节点。节点间使用定制高效通信协议 PRAP 互连成 2D Torus 拓扑结构，PRAP 可以提供或者借用远程内

存资源，或者直接访问存储节点资源。原型系统上部署数据中心软件栈，包括 DCOS 数据中心操作系统，Rainforest 节点操作系统，EKOS 存储系统以及 Hadoop+ 编程环境。

**2. 前瞻性资源管理可编程体系结构。**针对数据中心环境下多应用混合共享资源导致资源竞争不可控、应用服务质量无法保障的问题，国际上首先提出硬件上支持细粒度性能隔离和资源按需管理可编程的体系结构，并完成该结构的软件模拟器原型，对其功能和性能进行了全面的验证，目前已经对模拟器进行了开源（参见 <https://github.com/fsg-ict/PARD-gem5>）。

**3. 高通量数据中心的评测指标和基准测试程序。**2014 年 4 月发布高通量数据中心基准测试程序 BigDataBench3.0，包含 32 个高通量应用，12 月发布 BigDataBench3.1，扩展了应用的数据集规模并提供了多版本的实现。目前对该基准测试程序累计已有近 12000 个独立用户访问，包括剑桥、UIUC、Wisconsin 等。

**4. 集计算和通讯为一体的负载处理器 (WPU)。**WPU 能有效地将计算和通信融合在一起，提升数据中心内部处理数据的效率。已完成集成

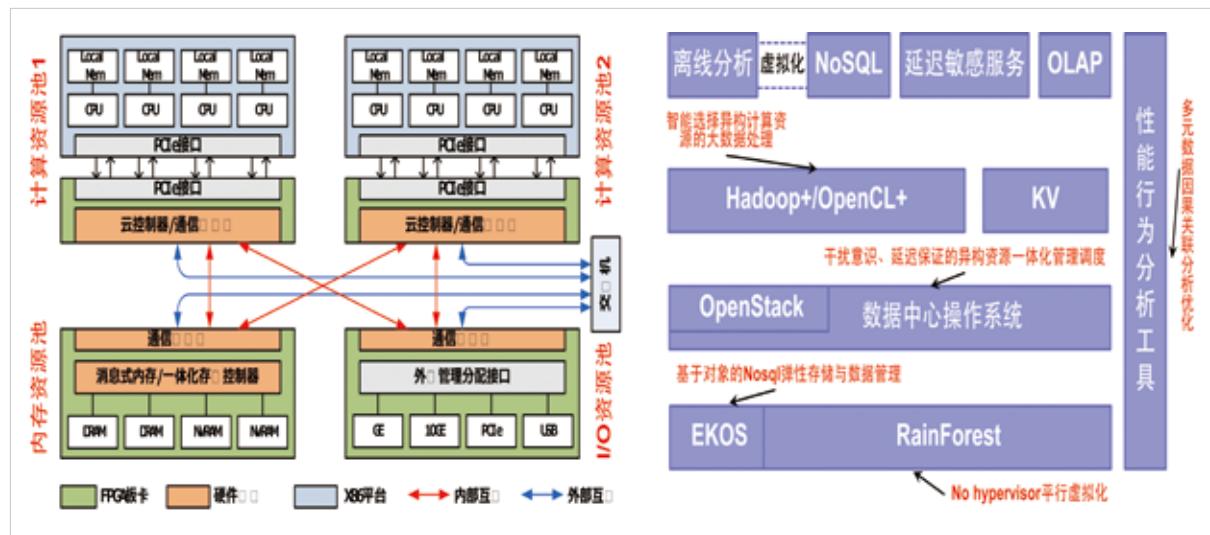
2路多线程的ARMv8单核芯片的流片，可成功运行Linux操作系统；完成40核ARMv8 SoC芯片的结构设计、逻辑设计和FPGA验证。

5. 新型内存、外存和网络子系统。设计实现了支持容量扩展和高效利用访存带宽的消息式内存系统，完成基于消息式内存的内存扩展板的母板和子卡FPGA原型；完成一体化存储UAS的FPGA原型，面向一体化存储的OS级资源管理内核模块、异构内存管理模块；完成数据中心网络DcnFPGA开发板的功能单元和分布式网络控制器，研究数据中心网络容错方案、基于快速交换设备的全光交换网络以及针对分层网络的负载均衡策略。

6. 支持异构系统管理的基本软件环境。以

高通量应用为驱动，研制基于高通量服务器的系统软件。完成X86平台上功能完善的Linux版Light OS，包含内存资源、CPU资源的动态调整功能，SMP支持功能，文件系统虚拟化功能，以及Light OS间通信优化功能；完成数据中心操作系统原型及数据中心性能分析优化工具。

7. 支撑高通量服务器的编程环境和高效编译工具。为了解决大规模数据处理在多个结点间并行处理所需要的编程框架和运行时支持，完成了异构MapReduce编程环境Hadoop+原型系统Beta版，并面向图计算领域和数据查询领域进行了编程扩展和优化；而在结点内部，完成了高性能的ARMv8编译器ARMv8cc-alpha。





## 研究方向科研进展

## 研究方向科研进展

### 编译与编程

研究聚焦在两个重点之上，一是降低多核处理器的编程开销，二是提升复杂软件系统的安全性和可靠性，这两点被誉为编译研究未来 50 年的两件大事。多核处理器飞速发展，已被广泛地运用于端、管、云环境中。实验室开展了面向海量并发系统的编程研究，以对云环境中的计算和资源进行有效地掌控；开展了面向应用的并行编程研究，以挖掘单个应用的并行度，利用多核对单个应用进行并行加速；开展了多核环境中的多级存储优化技术的研究，在片内高速缓存、片间带宽、片外存储等方面取得了一系列技术突破。在复杂系统的安全性方面，研发内构安全的软件开发和运行环境，通过系统性的缺陷挖掘和危害性消除、全生命周期的代码多样性、和多角度的完整性来提升软件的抗攻击能力，并探索进化机制，持续提升软件系统的安全性。上述研究得到了国家 973、863、自然科学基金的支持，取得了一系列研究成果，部分成果发表或录用于 ISCA、ICSE、MICRO、TPDS、CGO、ASE、DATE 和 TACO 等国际会议和学术期刊上。

### VLSI 测试验证、可靠设计与容错体系结构

围绕 VLSI 测试验证、可靠性设计、容错和可重塑体系结构等开展深入研究，取得的突出进展如下：1) 在 VLSI 测试验证方面，提出了基于马尔科夫链分析的功能测试生成方法，为时序电路状态机的激励生成提供了一个新思路；2) 在可靠性设计方面，针对多核处理器的热效应问题，提出一种功耗容量估算模型和功耗分配方法，在满足热效应约束的前提下显著优化了程序性能。3) 在容错体系结构方面，提出了在芯片低功耗态下的存储器容错和互连容错技术，显著降低了系统的功耗；4) 在可重塑体系结构方面，基于应用分类的异构体系结构设计方法，完成了一款基于 OpenCV 的视觉函数芯片 FPGA 原型。相关研究成果发表在 DAC、ICCAD、DATE 等国际知名会议，以及 IEEE Trans. on VLSI Systems 等国际知名期刊上。

### 类脑计算机

让计算机能以类似人脑神经网络的方式高效地进行学习等认知活动，是计算机科学的终极目标之一。但通常的计算机在模拟包含百亿个甚至更多的神经元及突触的深度神经网络进行学习时，处理速度和效率非常低下。未来计算团队在类脑

计算机方向取得重要的进展。他们研制了一种采用人工神经网络架构的机器学习运算装置——寒武纪 1 号，通过高效的分块处理和访存优化，能高效率处理任意规模、任意深度的神经网络，以不到传统处理器 1/10 的面积和功耗达到了 100 倍以上的神经网络处理速度，性能功耗比提升了 1000 倍。相关论文 (Tianshi Chen, Zidong Du, Ninghui Sun, Jia Wang, Chengyong Wu, Yunji Chen, and Olivier Temam, "DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning") 2014 年 3 月发表于本领域最好的国际会议之一 ASPLOS (ACM International Conference on Architectural Support for Programming Languages and Operating Systems)，并获得了最佳论文奖，这是亚洲科研机构首次获此奖。2014 年 12 月，后继工作 DaDianNao: A Machine-Learning Supercomputer 再次获得本领域顶级会议 MICRO 的最佳论文奖，这是 MICRO47 年来美国之外国家首获此奖。

## 并行软件研究

在数值计算核心算法优化方面，基于在 PPoPP 2013 上发表的 StreamScan 工作，我们针对 GPU 众核平台，提出一种新的稀疏矩阵存储格式 BCCOO 及其相应的 SpMV 算法，有效的减少了稀疏矩阵的存储开销。相关工作发表在 PPOPP 2014 上。继续开展 GPU 上国产自适应 FFT 软件包 MPFFT 的研制。针对目前主流的 GPU 提出了一个新的 FFT 自适应框架，且基于该框架实现了自主知识产权的 FFT 库 MPFFT，相关工作发表在 JCST 2014 和《计算机研究与发展》。面向多核 / 众核计算平台的并行编程方面，与编译

团队合作，参加“中科院计算所 - 华为联合实验室”A 类合作项目—多核编程框架研究。目前已经完成高层语言抽象，确定编译优化方案，编译框架的实现和优化正在进行中。深度学习方面，与旷视公司申请横向项目“基于 ARM 的高性能深度学习库”。相对于现有 Arm 平台上数学库、深度学习库以及 OpenCV 计算机视觉库，性能提升 2 到 4 倍，并已逐步将其应用到人脸识别相关产品中。目前，针对上述算法的 GPU 优化正在进行中。

## 操作系统研究

研究焦点在云计算数据中心面临的两大挑战上，一是数据中心服务器与制冷设备的节能降耗，二是数据中心面向大数据计算的新型计算架构与平台，目标是研发超越 Hadoop / Spark 的下一代高效能安全可靠的大数据计算平台，取得的研究进展：1) 为了解决大数据平台可控制 / 可观测性与海量观测数据实时采样、传输与处理上的矛盾，发掘了所采样数据在数值、物理意义等多维度的关联性，提出了设计高效组合压缩算法的新思路 2) 为了同时考虑数据中心中服务器计算能耗与为这些服务器制冷能耗的联合动态适配，提出了新的服务器负载与能耗模型，基于现有的数据中心制冷模型，提出了分析二者联动的动态实时优化分析模型，进而提出了面向联动能耗优化的 VM 迁移策略，为发展高效能数据中心提出了一个新思路，课题组的研究得到了科技部 863 计划和国家自然科学基金委员会的资助，部分成果发表在 ICPP 等国际学术会上。

## 数据管理研究

数据管理研究组围绕大数据对系统平台的三

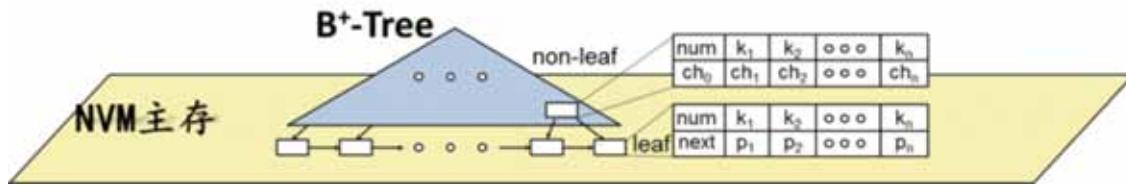


图 1 面向新兴的非易失主存优化广为使用的 B+-Tree 索引结构 (VLDB' 15 录用)

大主要挑战 (数据量大 Volume, 数据的产生处理速度快 Velocity, 数据的种类繁多 Variety) 展开深入研究。在 Volume 和 Velocity 方面, 主要采用硬件优化的思路, 充分发挥硬件设备的特征和优势, 高效地处理大数据。2014 年取得的突出进展包括: 在数据仓库系统中采用 SSD 支持在线的数据更新 (被 ACM TODS 录用), 面向新兴的非易失主存优化广为使用的 B+-Tree 索引结构 (被 VLDB' 15 录用, 图 1)。在 Variety 方面, 目标是研究和实现一个支持多种数据类型的大数据管理系统, 支持包括关系型数据、半结构化日志数据和图数据的统一存储与处理。2014 年实现了一个处理半结构化树状数据的数据分析系统 Steed, 基于 C/C++, 支持行式和列式的存储结构, 支持选择、投影、分组统计等通用的数据分析算法, 提供类似 SQL 的查询语言; 实现了一个图运算系统 GraphLite, 基于 C/C++, 支持 Pregel 方式的分布式同步图计算。此外, 研究了在 MapReduce 系统中如何灵活高效地利用索引, 成果发表在 EDBT' 14 会议上。

### 人脸识别

人脸识别研究组继续围绕人脸检测、面部特征点定位、人脸识别三个方向开展研究工作, 均取得了较大进展。在人脸检测方面, 提出了漏斗形的多姿态人脸检测框架, 有效提升了多姿态人

脸检测的性能; 在特征点定位方面, 提出了由粗到细的级联自编码网络方法, 并针对复杂的变化进行了话题自适应建模, 在公开测试平台上取得了最好的性能; 在人脸识别方面, 针对复杂真实场景下的视频识别问题, 提出了联合深度图像特征学习与混合统计量集合建模的方法。基于上述研究成果, 研究组本年度发表了 17 篇学术论文, 其中 CCF A 类论文 6 篇 (包括国际刊物 IJCV 和 IEEE TIP 论文各 1 篇, 国际会议 CVPR 论文 4 篇), 领域顶级会议 ECCV 2 篇, 领域顶级会议 NIPS 论文 1 篇, 1 篇论文获得 ACCV workshop 最佳论文。此外, 利用深度学习技术在姿态、表情分析方面取得了较好的效果, 提出的融合深度特征与多核学习方法取得了 ACM ICMI 2014 国际会议上组织的 EmotiW 表情识别竞赛的冠军。

### 机器翻译

自然语言处理研究组在机器翻译建模、翻译质量评估和自然语言生成等问题上持续开展深入研究, 取得了一系列成果。具体地, 提出了基于分析 - 生成模式的统计机器翻译模型, 借助句法信息的翻译质量评估方法, 基于句法信息的结构化语言模型, 以及文档级术语翻译建模等。机器翻译方面, 本年度发表领域顶级会议论文 6 篇 (COLING: 4, EMNLP: 1, AAAI: 1)。语言分析方面, 研究组在自然语言标注标准的自动适

应问题上展开系统性研究，提出了高效的异种知识库融合与转换方法，在保持分析速度不变的条件下可以显著提升分析精度。相关论文被领域顶级刊物 *Computational Linguistics* 录用，刘群研究员和姜文斌副研究员针对该工作在领域顶会 COLING 2014 做大会特邀报告。在科技奖项方面，本年度刘群研究员获得了中国电子学会科学技术奖一等奖。

### 智能视频分析

视频智能分析研究组继续坚持围绕对象化的监控视频内容分析展开研究，在单对象分析的基础上，又继续深入展开多对象跟踪、小群体行为分析和大规模人群行为分析方面的研究。针对多对象跟踪过程中的多特征融合、场景适应性等问题，提出了基于在线结构化的 SVM 的多对象跟踪、基于多对齐和重排序的低延时在线多对象跟踪算法；针对人群行为识别中深层语义缺失的问题，提出了基于社会属性力的群体行为异常检测、融合运动和表观信息的人群行为识别，课题组构造的多视 RGBD 日常行为数据库 ACT4<sup>2</sup> 已经被全球的研究者下载超过 300 次；在学习算法方面提出了深度度量学习方法，并应用于行人的再现检测任务上。本年度发表 / 接收学术论文 11 篇，其中包含 TCSVT 等国际国内期刊论文。

### 图像视频建模

图像视频建模研究组围绕视觉编码和表示学习开展工作，在深度编码器模型的启发下，提出了一种深度动态编码器 (deep dynencoders) 模型，用来建模视频序列，可应用在不同的视频任务上，包括动态纹理合成、视频分类和视频分

割等。还提出了一种新的图像超分辨率方法——深度网络级联 (deep network cascade, DNC)。DNC 方法可以看作是一种深层的模型，逐层的提高图像的分辨率。大量的实验证明，我们提出的方法的性能已经超越了目前比较流行的图像超分辨率方法。提出了一种基于学习的图像表示方法——光滑自动编码器。这种方法能够学习鲁棒而且判别力较强的特征表示。传统的自编码器以重构样本自身为训练目标，而光滑自编码器则是重构样本的近邻。如此，学习到的特征表示能够在局部邻域范围内保持一定的一致性，对较小的变化比较鲁棒。这种方法可以用于学习图像的特征，在图像分类和人脸识别等问题中展现出了其优越性。

### 智能人机交互

智能人机交互研究组在人机交互方向继续围绕手势、手语识别开展研究工作。课题组提出基于格拉斯曼协方差矩阵 (GCM) 的手语描述，在孤立词识别方面，1000 词汇规模的数据集上可以达到 92.4% 的首选正确识别率。在连续手语识别方面，课题组将 GCM 进行扩展，提出层级 GCM (HGCM) 的手语表示，结合多尺度时域置信传播策略，以解决连续句子识别问题。在手语数据库构建方面，课题组已经完成对标准中国手语词汇的采集，并对其中的部分数据进行了发布 (<http://vipl.ict.ac.cn/homepage/ksl/data.html>)。此外，为联合手语识别相关研究人员，推进手语识别向可实用的方向发展，课题组联合微软亚洲研究院、北京联合大学共同成立了基于 Kinect 的手语识别工作组，实现数据、代码等资源的共享，以开展更为广阔的国际、国内学术合作。



**数据挖掘与机器学习**

针对互联网上复杂事件关联，且事件在不断地更新和涌现，研究组设计了一种紧凑、高效的数据结构来存储事件场景，提出一种新的在线频繁场景挖掘算法。该成果被CCF A类会议ICDE 2015录用为长文，这也是首篇计算所为第一完成单位的ICDE研究类长文。结合主成分分析(PCA)和大间隔学习，提出了一种基于后验正则化的贝叶斯方法，被CCF A类会议AAAI 2015录用为长文。在机器学习算法以及大数据并行数据挖掘技术研究方面发表录用26篇论文，其中CCF A类2篇，CCF B 8篇，SCI论文7篇。

### 智能视频监控

本年度课题组继续坚持围绕对象化的监控视频内容分析展开研究，在单对象分析的基础上，又继续深入展开多对象跟踪、小群体行为分析和大规模人群行为分析方面的研究。针对多对象跟踪过程中的多特征融合、场景适应性等问题，提出了基于在线结构化的SVM的多对象跟踪、基于多对齐和重排序的低延时在线多对象跟踪算法；针对人群行为识别中深层语义缺失的问题，提出了基于社会属性力的群体行为异常检测、融合运动和表观信息的人群行为识别，课题组构造的多视RGBD日常行为数据库ACT4^2已经被全球的研究者下载超过300次；在学习算法方面提出了深度度量学习方法，并应用于行人的再现检测任务上。本年度发表/接收学术论文11篇，其中包括TCSVT等国际国内期刊论文。

### 社交媒体计算

在社交媒体计算方向，2014年度重点研究了网络群体行为分析、消息流行度预测、影响最

大化算法、网络信息传播等基础问题，取得的主要研究进展如下：

#### (1) 团队合作中的群体分配机制研究

团队合作逐渐成为科学研究的主要模式，打破了个人科学研究所贡献和功劳的对称，给功劳分配提出了严峻挑战。为了区分团队成员在团队合作的贡献大小，不同的学科所采取的策略差异很大，而现有方法大多基于作者顺序或作者声明的贡献来确定每个人的功劳分配比例，无法应对学科间规则差异大、实际贡献难以量化等难题。针对该科学难题，我们提出了群体功劳分配方法(Collective Credit Allocation)，该方法不再试图量化个体的真实贡献，转而采用“被感知的贡献”(Perceived Contribution)作为功劳分配的依据，利用共引用关系作为度量“被感知贡献”的依据。该方法独立于学科分配规则和作者顺序等信息，能够自适应地量化个体从团队合作中获得的功劳分配比例。预测诺贝尔奖得主的实验表明，所提出的群体分配方法在预测准确率上达到了86%，远高于基于作者顺序和作者角色的方法。同时，所提出的群体分配方法还可以帮助我们研究获奖带来的马太效应，并可用以比较不同学者在其共同领域中的影响力。该研究成果发表在综合性学术刊物《美国科学院院刊》(PNAS)上，论文为发表当期的六篇highlighted article之一，并被ScienceDaily等多家学术媒体报道和转载。

#### (2) 信息传播中的时间标度研究

信息传播中用户间交互行为的时间间隔对于研究信息传播具有重要的启示作用。现有方法大多假设用户间交互行为的时间间隔服从指数分布或泊松分布，基于该假设的模型和方法将数据按照时间进行切片来研究，缺少大规模数据上的实证分析支持，且预测精度受限。我们以WISE

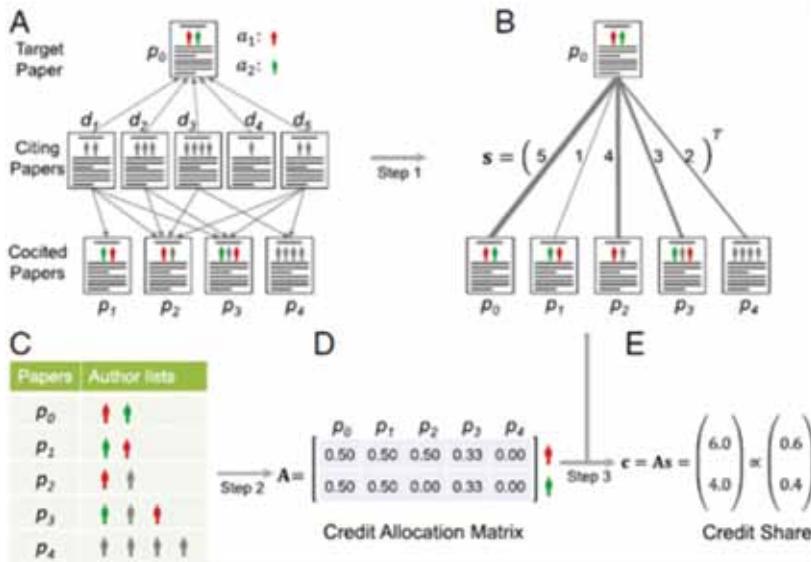


图 2 群体功劳分配方法

2012 国际会议提供的新浪微博数据为基础, 分析了用户间交互行为的时间间隔分布。分析发现, 对于任意一个用户和其任一关注者, 他们之间的大部分转发行为密集发生在一些较短的时间窗中, 这些频繁转发的时间窗之间有大段的空白期, 表明其间关注者很少转发被关注者的微博。这表明用户交互行为表现出了个体行为中时常观测到的阵发现象: 个体间短期的频繁交互被长期的沉默所间隔, 时间间隔服从幂率分布, 而不是现有模型所假设的指数或泊松分布。基于该发现, 我们建立了一个含时模型 (Decay Model), 来估计和预测未来社会推荐的成功概率, 并在新浪微博数据上进行实证检验。在任意比例的训练样本下, 所提出的含时模型都一致优于现有不含时的基准方法。当训练样本较少时优势更加明显, 表明含时模型只需少量样本即可准确估计影响力。该研究成果发表在 Nature 旗下开放访问的学术刊物 Scientific Reports 上。

### (3) 社交网络上的高效影响最大化算法

影响最大化 (Influence Maximization) 旨在解决如何在网络上选择一部分初始种子用户, 由他们口口相传将企业的产品或信息尽可能地推广开来。该问题是在社交媒体上进行口碑营销或病毒式营销的核心问题。现有方法中, 贪心算法精度高但速度低, 启发式算法速度快但精度无保障。为此, 我们拟设计高效的影响最大化算法, 从启发式的角度切入研究影响最大化问题, 分析发现贪心算法

所得的解集是一种自洽排序——节点的边际影响力和它们的排序是自洽的。据此, 我们提出直接通过寻找自洽排序来求解影响最大化问题。我们提出一个迭代的框架 IMRank, 将任意给定的初始排序通过迭代调整的方式得到自洽排序。迭代框架包括两个步骤: (1) 根据当前排序计算各个节点的边际影响力; (2) 按照边际影响力对节点进行重新排序。我们在理论上证明了算法的收敛性, 并在 PHY、DBLP、EPINIONS、DOUBAN 和 LIVEJOURNAL 等 5 个数据集上测试了 IMRank 的性能。实验结果表明, IMRank 在计算效率上显著超过了当前最好的贪心算法和启发式算法, 计算精度和当前最好的贪心算法相当。该研究成果长文发表在信息检索领域 CCF-A 类会议 SIGIR 2014 上。

### (4) 流行度动态过程的建模与预测

在线社会关系网络中, 消息之间相互竞争用户的关注度。根据平均场理论, 对于单个消息而言, 其流行度可以视为该消息获得关注的一个动态过程。实证分析发现, 消息获得关注的过程受

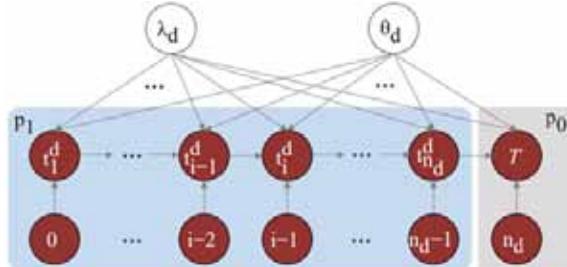


图 3 流行度动态过程的模型

到三个微观机制的影响：（1）适者生存，即消息自身的吸引力对其最终的流行度起决定作用；（2）富者愈富，即消息的流行度越高越容易获得新的关注；（3）时间效应，即随着时间推移消息获得关注的可能性会下降。针对该问题，我们提出了一种基于自增强泊松过程 (Reinforced Poisson Process) 的概率模型，建模流行度动态过程的上述三个微观机制。该模型是一种产生式模型，直接建模各个消息获得关注的过程，可再现整个在线社会关系网络的流行度分布。进一步，通过将消息自身的吸引力视为隐变量，并为其引入一个共轭先验，采用贝叶斯框架对消息流行度进行预测，进一步提高了预测的准确度。以美国物理协会旗下 11 个期刊从 1893 年到 2009 年间的引文网络为例，将论文的引用次数视为流行度，通过预测论文的引用次数来验证我们所提出模型的有效性。实验表明，所提出的基于自增强泊松过程

的模型，在预测流行度方面显著超过了基于时间序列自回归的模型和对数回归模型。该研究成果发表在人工智能领域 CCF-A 类会议 AAAI 2014 上，获得国际同行的广泛关注和跟进。

### 大数据机器学习

在大数据机器学习方向，2014 年度重点研究了短文本话题建模、多样性排序学习、基于深度学习的文本表达、分布式机器学习框架等问题，取得的主要研究进展如下：

#### （1）短文本话题建模

概率化主题模型是文本数据约简的一种有效方法。在这类模型中，一个文档被建模成一个主题的混合分布，主题建模成一个词的多项式 (multinomial) 分布。然而，在短文本中这种建模方式会遇到严重的数据稀疏性问题：（1）短文本中大部分词都只出现一次，因此词频信息对词相关性和重要性判断没有区分度；（2）由于短文本文档过短，上下文信息的缺乏会给一些二义性的词的主题判别带来困难。为了解决在文本约简表达中面临的大量短文本数据显著的特征稀疏问题，提出了双词概率主题模型 (Biterm Topic Model) 来进行有效的文本主题表达学习。其核心思想是直接建模全局的词共现关系，这样做的

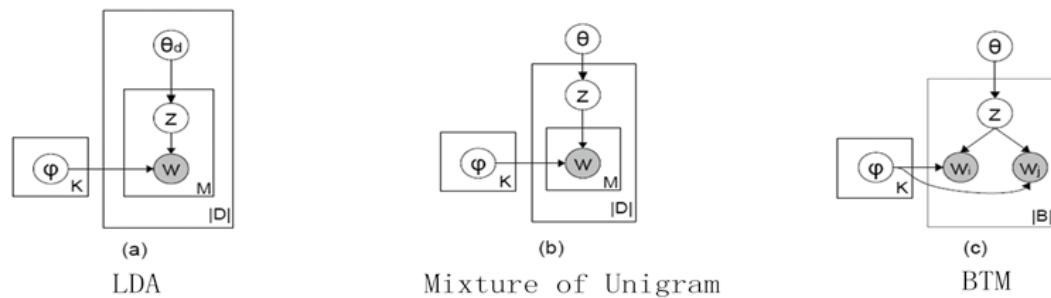


图 4 短文本概率主题模型

好处是 1) 词共现关系包含了上下文信息, 比单个词更容易判断其中词的主题属性; 2) 词共现关系与文档长短无关。虽然单个文档内部的词共现关系比较稀疏, 但只要数据足够多, 全局的词共现关系仍然很充分。实验表明, 基于 BTM 学得的文本约简表达, 在分类性能上, 准确性比经典的 LDA 算法提高 13%, 在聚类性能的评价上, 性能比经典的 LDA 算法提高 17.7%。相关论文发表在数据挖掘领域顶级期刊 TKDE 上。

## (2) 多样性排序学习

检索结果多样性是信息检索中的一个重要问题, 现存方法主要使用基于经验定义的函数来实现多样性的效果, 难以加入有效的特征, 也很难在实际中进行调节。我们提出将多样性问题看做一个学习的问题, 并提出一个新颖的关系排序学习的框架去解决该问题。具体的, 我们首先阐述多样性排序问题是一个顺序选择过程。在此基础上, 我们将排序函数定义为相关性得分与多样性得分的组合, 其中, 相关性得分与传统排序学习中的定义类似, 而多样性得分定义为当前文档与排序在前面的文档的差异化程度, 具体数学表达:

$$f_S(x_i, R_i) = \omega_r^T x_i + \omega_d^T h_S(R_i), \forall x_i \in X \setminus S,$$

其中  $x_i$  代表当前文档,  $S$  代表已选择文档, 即排序在前面的文档,  $X$  代表该查询下的待选文档集,  $R_i$  代表其他文档与  $x_i$  的关系张量,  $h_S$  代表多样性特征向量, 具体的, 可用如下三种形式

$$h_S(R_i) = (\min_{x_j \in S} R_{ij1}, \dots, \min_{x_j \in S} R_{ijl}).$$

$$h_S(R_i) = (\frac{1}{|S|} \sum_{x_j \in S} R_{ij1}, \dots, \frac{1}{|S|} \sum_{x_j \in S} R_{ijl}).$$

$$h_S(R_i) = (\max_{x_j \in S} R_{ij1}, \dots, \max_{x_j \in S} R_{ijl}).$$

在论文中, 我们使用了包括文档内容多样性特征, 子话题多样性特征, 链接相关的多样性特征等六种不同特征。需要注意的是, 我们提出的框架是一般的, 可以处理其他特征的情况。在排序函数的基础上, 我们提出使用 Plackett-Luce 模型的次序化损失为学习的目标, 最终使用极大似然为损失, 使用梯度下降方法来进行求解。实验效果标明我们方法的有效性, 如下图所示。相关论文发表在信息检索顶级国际会议 SIGIR2014.

## 跨媒体计算

2014 年度积极推进科研仪器研制项目和院地合作进展, 坚固在国家互联网音视频监管任务中的地位, 在学术研究、关键技术应用和项目申请三个方面取得了显著的进展。学术研究方面, 在相似视频检测技术、语义概念检测技术和高维特征索引技术的研究上取得了突破: 提出了提出基于分段稀疏表示的语义基元视觉关键词方法, 实现了高达上百万维的视觉关键词快速提取, 码表内存占用仅为传统方法的 1/1000; 提出了稀疏集成学习的视觉语义识别方法, 解决了大规模训练集的快速训练和识别问题, 识别各类别仅需 11 毫秒。发表多篇高水平论文: 包括国际顶级期刊 ACM TIST 1 篇 (SCI IF=9.39), IEEE T-IP 2 篇 (CCF A 类期刊), IEEE T-MM 4 篇; 国际顶级会议 ACM Multimedia 长文 2 篇。关键技术应用方面, 相关核心技术在互联网音视频分析和大规模社会媒体数据分析中得到了良好应用, 国际计算机法证领域最著名公司冰岛 Videntifier 公司引入我们的语义概念检测技术, 为其承担建设的国际刑警组织反儿童性剥削数据库 ICSE 提供关键技术, 荣获 2014 年度北京市科学技术一等奖。



## 生物信息学

2014 年度继续在计算蛋白质组学及其相关领域开展深入研究。蛋白质鉴定方面, pFind 团队完成定量分析软件 pQuant, 突破了分析化学领域主流国际期刊 *Analytical Chemistry*; 完成蛋白质鉴定新一代引擎 pFind 3, 实现定性定量综合、速度精度领先; 连续第三次主办中国计算蛋白质组学研讨会。在糖的质谱鉴定方面, 提出了多级质谱鉴定的层次贝叶斯模型, 有效融合了上一级质谱鉴定信息到下一级中, 依据最大熵的原则指导质谱鉴定中的谱峰选择; 在蛋白质结构预测方面, 参加 CASP-11 比赛, 取得第九名和第 10 名。中国科学院知识创新工程重要方向项目非

编码 RNA 和系统生物学网络分析算法研究、非编码 RNA 调控网络及功能研究, 通过结题验收; NONCODE 数据库第四次升级, 升级文章发表在 *Nucleic Acids Research* 杂志。围绕生物显微图像处理和三维重构技术, 提出一种大分子复合体多重构象的三维结构局部优化算法, 提高精度 10%-20%; 提出一种先验约束的重构算法和图像增强算法, 提高重构结果的精度; 提出一种新的超分辨成像算法, 提高了原有算法的空间和时间分辨率。

## 虚拟现实

2014 年度在人体运动仿真、群体运动仿真、

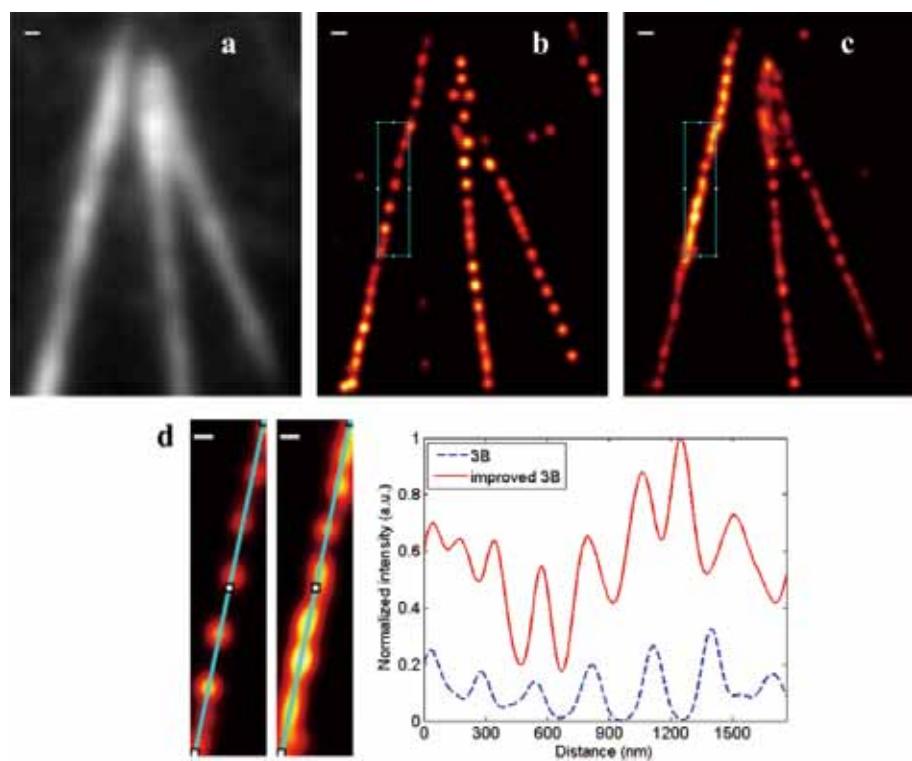


图 5 超分辨光学成像实验结果

(a) 200 帧衍射图像的叠加图 (b) 使用原有成像算法的重构结果 (c) 我们方法的重构结果 (d) 左图为 b、c 区域的放大比较, 右图为线条上的密度曲线。其中 a、b、c 图比例尺为 200nm, d 图为 100nm

火焰等自然现象建模仿真、虚拟体验技术开展了研究工作。本年度代表性工作包括：在群体运动仿真方面，重点研究并提出了一种高效车道网络语义表示方法，能够支持车辆运动仿真中里程信息的快速三维转换和周边路况查询，有效提高车辆运动仿真计算的效率；研究并提出一种群集行为建模方法，从心理和生理机制出发描述多维情绪状态对人群聚集行为的影响，实现逼真群集行为的建模与仿真。基于相关技术开发的核电周边人员撤离模拟系统，完成泰山核电厂周边人口撤离模拟的示范应用，通过仿真不同辐射源项、不同天气、不同交通情况下的人口的撤离，为影响撤离关键因素的研究和分析提供科学依据。

## 分布式计算

2014 年度对科研目标进行凝练，对相关科研方向进行调整，聚焦于云计算平台及其应用开发两个方面，取得了如下进展：在系统开发及应用方面，项目组针对教育行业对虚拟化教学应用的需求，在自有的云计算平台上设计开发了面向高等院校、中等 / 高等职业院校的虚拟化教学应用系统。与新疆大学，杭州电子科技大学等单位合作筹建云计算实验室，虚拟化教育应用得到具体应用。在项目争取方面，争取到国家基金委项目 30 万，电科院横向项目 50 万。在关键技术研究方面，在虚拟化技术、云计算方面取得多项关键技术，相关论文已发表在 IPDPS workshop 和 IEEE TPDS 等国际会议和重要期刊上。

## 计算机基础算法

2014 年度主要围绕着社会网络与博弈论相关算法、传统计算复杂性、无差错

量子算法等三方面进行了深入的研究：首次提出了社会网络中有概率保证的影响力最大化问题并给出了  $O(\log(n))$  的近似算法，十年来首次改进了判定树研究中 sensitivity-block sensitivity 猜想中的下界和指数上界，在消息传递模型下给出了包括解线性方程组等在内的多个问题的最优通信复杂度上下界，对无差错量子算法展开了初步研究。14 年共发表会议论文 5 篇，SCI 期刊论文 1 篇，包括 ICALP, DISC, KDD, ISAAC, WINE 等会议和 Theoretical Computer Science 等期刊。

## 数据存储技术

本年度数据存储技术在海云存储多元数据服务器、高可靠存储服务器方面有阶段进展，数据集群基于大粒度子卷进行负载分布、定位、均衡，突破全异步分布式元数据一致性协议关键技术，可扩展至数百台元数据服务器；搭建了包含 16 台元数据服务器原型系统，实测支持 100 Billion 文件，规模比肩谷歌文件系统；支持高可靠的网络盘阵可容忍节点级故障，经完善已达产品级成熟度，在中科院声学所、网络信息中心等单位进

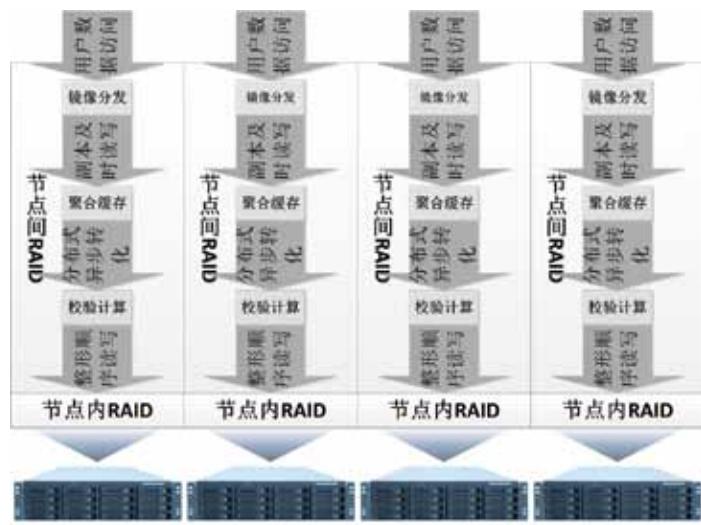


图 6 BWRAID 高可靠阵列系统示意图



图 7 高通量众核处理器 DPU-m 芯片



图 8 高通量视频数据 40 路实时转码

行了部署。围绕以上研究内容顺利完成了 863、973、院先导、院重点部署项目的相关科研任务和年度考核。数据存储技术研究中心在学术培养方面持续努力，在 2014 年计算机学会全国存储大会上获得了 5 篇优秀论文中的 3 篇，并获得其中的第一、二名。

### 高通量媒体大数据众核处理器 DPU-m

面向媒体大数据应用的高通量众核处理器 DPU-m 成功流片。该芯片采用 TSMC 40nm 工艺，芯片面积 54 mm<sup>2</sup>，处理器主频 800MHz，含有 220 个处理部件，单芯片功耗不超过 3W。基于 DPU-m 众核处理器研制的高通量媒体大数据处理加速卡，其正常工作时的单卡功耗不超过 19W，在视频数据编解码处理器方面，单卡性能与现今视频数据处理领域的主流服务器 Intel 的 Sandy Bridge 4 核相比，性能功耗比可达其的数十倍以上。

### 面向数据中心的新型计算机体系结构： PARD

传统的指令集 (ISA) 抽象体系结构接口无法传递更多应用信息到底层硬件（如安全级别、

QoS 需求等），导致硬件无法区分不同安全级别或不同 QoS 需求的应用，从而在硬件层次出现应用间相互干扰。传统抽象接口的不足已对云计算、互联网等新兴应用场景带来了负面影响。例如，2013 年 Google 最新数据表明包含在线应用的数据中心 CPU 利用率依然只有 30% 左右，但不考虑 QoS、只运行离线批处理作业的数据中心则能达到 75% 的 CPU 利用率。

针对数据中心环境下资源利用率与应用服务质量之间的矛盾及其原因，先进计算机系统研究中心前沿系统小组提出一种新计算机体系结构 PARD (Programmable Architecture for Resource-on-Demand)。PARD 核心思想是将计算机看作是小型网络，将软件定义网络 SDN 思想应用到计算机体系结构，在计算机硬件资源管理机制中引入网络 QoS 技术，提供新的接口运行应用将高层语义信息传递到底层硬件，从而在硬件上支持资源容量隔离与性能隔离，使数据中心能在保障关键应用服务质量的前提下提高资源利用率。PARD 已与华为联合申请了包含多项高价值专利的专利群以及国际专利，并且在学术上已被计算机系统领域国际顶级会议 ASPLOS 2015 录用。

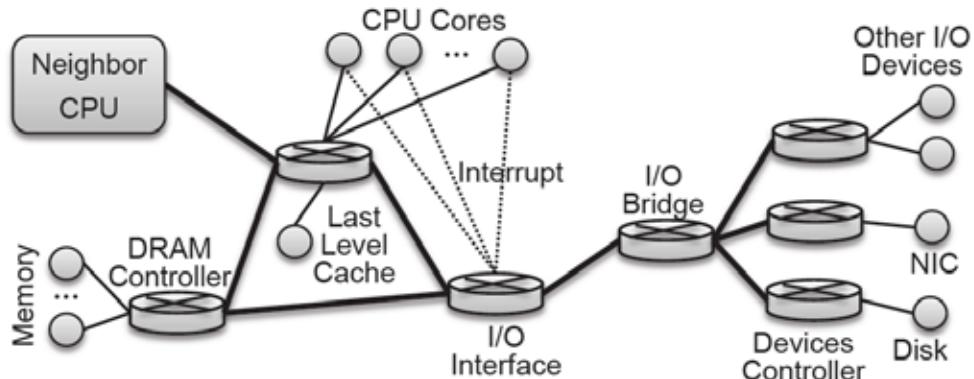


图 9 计算机可以看做是网络

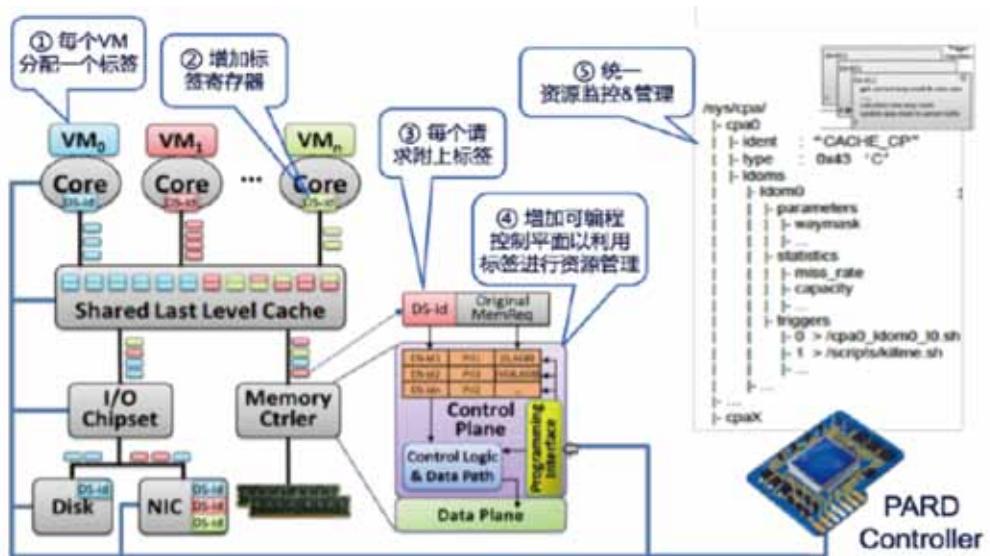


图 10 PARD 总体系统结构

### 大数据基准测试程序集: BigDataBench

BigDataBench 是一个开源的大数据基准测试集。BigDataBench 包括 14 个数据集, 可以生成 PB 规模的文本、图、表数据。同时 BigDataBench 包括搜索引擎、电子商务、社交网络、多媒体、生物信息学等 5 个应用领域的 33 个典型负载。

BigDataBench 发布后, 独立 IP 访问累计近

12,000 次, 被剑桥、三星、UT Austin、OSH 等多个国际知名学术团体和工业界使用并公开发表论文, 其中剑桥大学的团队为 ARM 处理器的创始团队; BigDataBench 在欧盟组织、SPEC 组织等在大数据相关报告中引用; BigDataBench 已经联合工信部电信院标准化所制定大数据测试标准并完成技术需求和测试规范撰写。

BigDataBench 的相关论文发表于体系结构

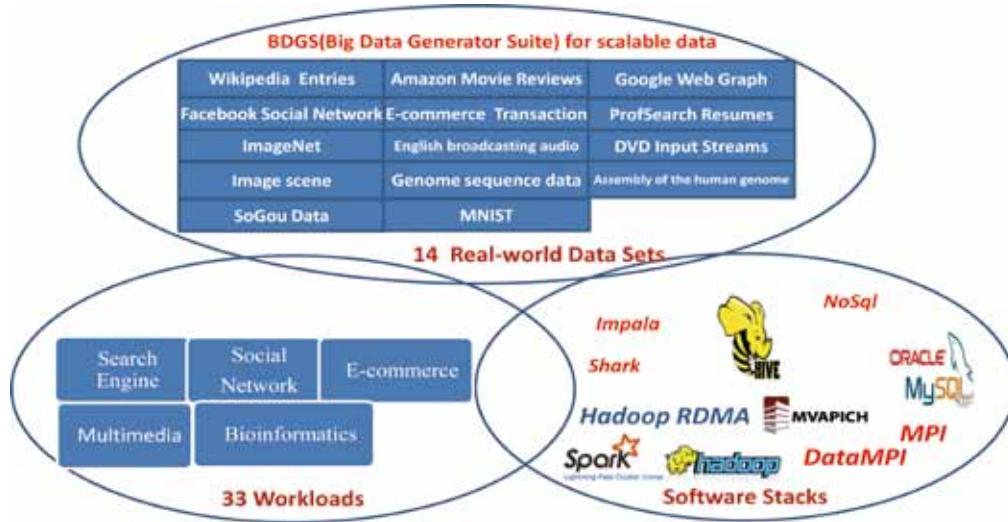


图 11 BigDataBench

领域著名会议 HPCA-2014、并获得负载分析领域著名会议 IISWC-2013 的 Best Paper Award。同时 BigDataBench 在 HPCA 2013 大会上进行了 Tutorial，并受邀参加 MICRO 2014 的 Tutorial。

### 残疾人信息无障碍建设

普适计算研究中心在残疾人信息无障碍建设方面继续发力。中心以残疾人的基本需求为导向，从残疾人参与远程信访维权活动的实际需求出发研制了残疾人远程信息无障碍交流系统。该系统能够实现简便易操作、自然状态的多方远距离可视交流与任务协同，从而为残疾人提供不出社区、村情况下的无障碍信访维权服务。同时为残疾人服务机构实现高效的视频会议和监控管理，提升残疾人基层服务的信息化水平。目前，系统已经成功部署在北京市 16 个区县的残疾人温馨家园，受到中国残疾人联合会理

事长及其他政府人员以及残疾人的广泛好评。另外，面向智力残疾人和老人 / 儿童研发的“防走失定位胸卡”，获北京市残疾人辅助器具“金点子”创意发明活动三等奖。面向视力障碍人士研发的可穿戴导盲避障系统，能实时鲁棒检测和反馈前方路况及障碍物信息。相关系统在 2014 国际福祉博览会上进行了展示，获得了广泛关注。



图 12 北京市残疾人远程无障碍交流系统部署示意图

## 获奖成果

# 大规模网络视频处理与内容分析 关键技术及应用

为了促进网络视频信息消费的发展，同时加强网络视频内容监管，保障网络视频产业健康有序成长，项目在国家863计划、北京市科技计划、国家自然科学基金的支持下，开展大规模网络视频处理与内容分析关键技术研究及应用。

项目的主要技术创新点有：1) 发明了一套视频编解码优化及适配处理方法，有效提升编解码处理速度和解码重建的视觉质量，满足实际应用对带宽、实时性及功耗的要求，同时能够高效适配不同终端特性。2) 发明了一套适用于大规模相似视频检测的特征提取、匹配与高维索引方法，实现了高速、高精度的大规模相似视频检测。3) 发明了一种适用于大规模网络数据的视觉语义概念训练和检测框架，能够有效应对复杂变化的网络视频数据，具有较高的训练效率和检测精度。4) 发明了一种网络视频话题检测与跟踪方法，解决了对稀疏、噪声大的网络数据进行话题检测和热点话题发现的难题。5) 发明了一套分布式采集框架与策略，实现了大规模网络视频数据的高效采集和相关信息精确抽取。



项目成果获国家授权发明专利 19 项，在 IEEE T-MM、T-CSVT 等著名国际学术期刊及顶级国际会议上发表论文 48 篇，获 3 次著名国际会议最佳论文奖，在视频检索国际评测 TRECVID 中取得第一名的成绩。

项目成果在网络视频服务和国家互联网视音频监管领域取得大规模应用，近三年直接经济效益 12040 万元，间接经济效益 1150 万元。由于技术性能先进，项目所研发的手机视频服务平台被选为中国移动和央视网合作开展的手机视频业务的运营平台，取得了显著的经济效益。项目成果在国家互联网视音频监管主管部门—国家计算机网络与信息安全管理中心和国家广电总局信息网络视听节目传播监管中心取得成功应用，在一系列国家网络安全重大专项任务中发挥了重要作用，为全面提升国家互联网视音频监管能力提供了关键技术与系统，社会效益显著。

该项目由中国科学院计算技术研究所、北京北纬通信科技股份有限公司、北京神州天脉网络计算机有限公司共同完成，荣获 2014 年度北京市科学技术一等奖。



### 结题验收的代表性成果

# 面向 IMT-Advanced 增强多媒体多播技术

“面向 IMT-Advanced 增强多媒体多播技术（课题编号：2010ZX03003-004-03）”是由计算所牵头的“新一代宽带无线移动通信网”国家科技重大专项课题，执行期为 2010 年 1 月至 2013 年 6 月，2014 年 5 月 23 日顺利通过工信部验收。

本课题的主要研究内容及成果包括：

#### 1. 增强多媒体多播传输技术

提出了基于叠加码的高能效单播与多播混合业务的机会多播方法、针对多播传输的 MIMO 技术优化、单频网多基站协同分集传输机制等，从而提高了 IMT-Advanced 系统中的多播传输容量和质量。

#### 2. 增强多媒体多播无线资源管理

提出了包括单小区多播资源分配、单频网资源管理以及单频网与单小区多播业务的资源动态复用等方法，充分考虑不同业务的特性，在不影响单播业务性能的前提下，提高系统对多播业务的提供能力。

#### 3. 保障多播可靠分发的网络层分发技术

提出了基于数字喷泉编码的多播数据传输方式、协作多播机制、多播业务同步方法、移动性支持方法等，为此提高了多播传输的可靠性。

#### 4. 增强多媒体多播网络部署优化技术

提出了包括单频网区域选择、拆分、动态调整等部署优化方案，提高单频网传输的频谱效率；同时，对广域多播传输、中继协助传输等机制进行了研究，提出了相关的融合组网方案。

#### 5. 增强多媒体多播协议

面向 IMT-Advanced 系统设计增强多媒体多播核心协议内容，包括单频网管理、多播与单播业务信道复用、MBMS 服务连续性保障、MBMS 计数等，参与完成了“宽带无线多媒体系

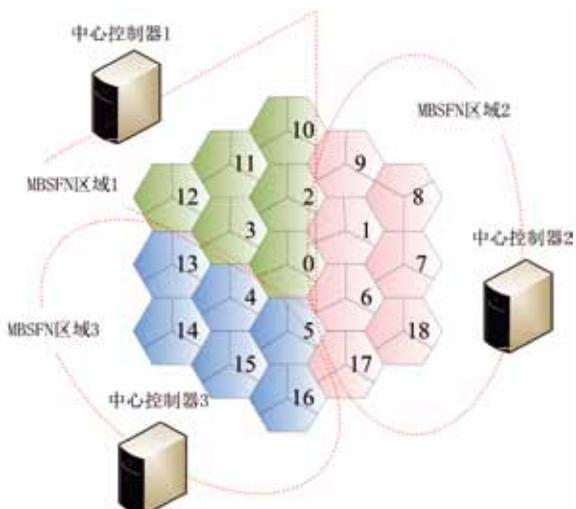


图 1 单频网多播仿真环

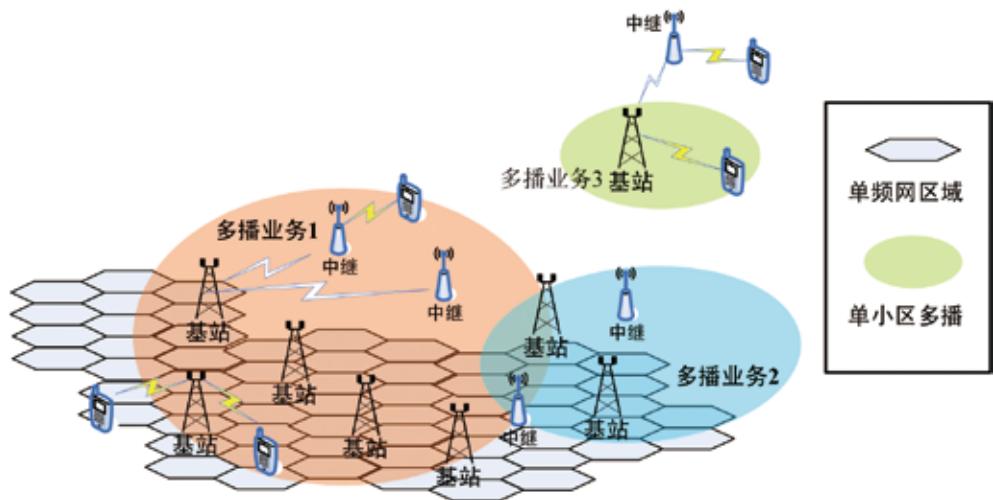


图 2 支持广域覆盖的多媒体多播混合组网

统的空中接口”国家标准（20067544-T-339）的制定。

#### 6. 验证平台研制

实现了支持 IMT-Advanced 增强多媒体多播的仿真平台以及原型验证平台，为增强多媒体多播技术验证提供试验环境。其中，仿真验证平台支持面向 IMT-Advanced 的单小区和单频网多播功能，能够支持从链路级到系统级的多播新技术验证；原型验证系统物理层基于 OFDM 技术，能同时支持数据、语音和多媒体业务，支持单播和多播传输。

本课题在增强多媒体多播协议框架的基础上，研究了包括物理传输、无线资源管理、多小区数据分发和网络部署等几个方面的关键技术，利用分层和协同的方法极大的优化了系统性能。同时，课题的成果输出通过专利和 IMT-A 标准的

形式进行落实。具体的，申请了 25 项专利，并通过积极参加国内、国际标准化工作，提交了 28 项项标准提案，部分提案已经被标准组织接纳，提高了我国在 IMT-Advanced 标准中自主知识产权的比例。从而促进我国蜂窝网络多媒体多播业务的发展和推动 IMT-Advanced 系统的产业化步伐。



图 3 原型验证系统



## 结题验收的代表性成果

# 图像与视频处理

国家自然科学基金杰出青年基金项目“图像与视频处理”于2014年12月25日在北京通过了验收。该项目的执行时间是2011年1月至2014年12月。本项目旨在研究图像、视频处理中的表示和识别等基础性问题及其在检测、识别和编码等方面的应用。探索建立高效的层次化图像、视频表示模型，实现局部特征表示与全局特征表示的过渡和统一，同时兼顾表示的判别和重构能力。

经过四年研究，项目所取得的主要成果如下：

1. 深入研究了图像集合的流形表示方法，提出了一种具有重构能力的流形学习方法——最大线性嵌入 (Maximal Linear Embedding, MLE) 以及利用流形间距离的通用识别框架；提出了一种通过流形对齐进行先验知识迁移的方法，并提出了利用流形相似性从局部稠密数据集构建全局稠密数据集的方法。

2. 在稀疏和紧致表示方面，提出了一种稀疏编码的局部描述方法和一种结构化稀疏的线性判别分析方法，并将稀疏约束用于单样本识别的知识迁移，以及图像重构的基底选择。利用稀疏和紧致表示，提出了一种同时将图像质量分析与视频帧选择、待识别对像对齐和多帧图像同时表达

与识别的框架，用于视频对象的识别与分析，取得了明显的效果。

3. 在图像 / 视频的层次化、多视表示方面，提出了一种跨视的图嵌入方法以及多视判别分析方法，通过将所有视图的类内变化和类间变化（同时考虑视图间与视图内）结合到一起，最终形式化成 Rayleigh 熵的形式，从而得到一个判别式的公共子空间，提供了同时度量多个视图的方法；提出了一种异质空间距离度量的方法，为跨空间表达提供了一种新的度量手段。跨空间的表示及其距离度量为解决复杂变化下的识别问题提供了有效的方法。

4. 在纹理表示方面，提出了一种考虑结构相似性的纹理描述子，使得描述子的数值和纹理的结构相关联，从而可以和其他描述方式相结合，用于纹理的分类；利用多层谱聚类进行视频的分割与表示，并利用多层网络结构建模动态视频，可以实现高效的动态纹理合成，这一方法还可用于视频分类和视频分割等多个不同的任务。

5. 在特征定位方面，提出了利用多维度上下文信息的特征定位方法，综合利用包括图像内、图像序列间、同类别图像间等不同关联信息，提高了大尺度复杂变化和在缺乏明显纹理结构区域

的定位精度，改善了复杂成像条件下的人脸识别的性能。

在本项目支持下，项目组在本领域主流刊物和会议上发表论文 50 篇 ( 国际刊物论文 15 篇，国际会议论文 33 篇，国内刊物论文 2 篇 )，其中含 CCF A 类期刊 (TPAMI, TIP, IJCV) 论文 7 篇，CCF A 类会议 (ICCV, CVPR) 论文 7 篇，上述论文据 Google Scholar 统计累计引用 400 余次，其中单篇引用超过 20 次的有 8 篇；获授权国家发明专利 11 项；获 Emotion Recognition in the Wide 2014 竞赛和 FG 2015 Video Person Recognition Evaluation 的第一名；团队成员一人获基金委优青资助。培养博士 8 人，获中科院优博论文和 CCF 优博论文奖励各一项。在国际交流方面，作为大会主席等主要成员参与承办了 IEEE International Conference on Automatic Face and Gesture Recognition

2013 (FG2013)、IEEE International Symposium on Circuits and Systems 2013 (ISCAS 2013)、IEEE International Conference on Multimedia and Expo 2014 (ICME 2014) 等大会，项目负责人应邀多次赴欧洲讲授短期课程和交流，并在 DICTA 2014 上做特邀大会报告。参与撰写完成国家标准一项，IEEE 标准两项，获得国家科技进步二等奖一项。



IEEE FG2013 组委会合影

跨 越 发 展  
持 续 发 展



## 结题验收的代表性成果

# Web 搜索与挖掘的新理论和新方法

国家自然科学基金重点项目“Web 搜索与挖掘的新理论和新方法—支持舆情监控的 Web 搜索与挖掘的理论与方法研究”2013 年 12 月在南京通过了会议验收，2014 年 2 月提交了正式的结题报告。该项目起止时间是 2010 年 1 月至 2013 年 12 月，项目立足于 Web 信息处理的多维度、不确定性、演化性等挑战性科学问题，从社会舆情计算和新型 Web 搜索与挖掘的实际需求出发，提出 Web 信息处理的新理论与新方法，为网络舆情监测预警等新型应用奠定理论与技术基础。

在基于维度挖掘的多维计算理论，大规模信息处理的不确定性计算理论，基于复杂系统网络模型的聚集计算理论等三个科学问题的研究上取得了突出进展，分别重点研究了网络内容计算的不确定性理论、多维度的话题模型理论和复杂网络模型理论。在关键技术层面解决了国家舆情分析面临的 Web 搜索和挖掘的关键问题。在社交网络的个性化搜索方面，提出了基于用户搜索行为分析的高效查询推荐、基于 MapReduce 模型的范围分析优化，及基于指纹的文档集快速搜索等技术；在网络话题发现与演化分析方面，提出了基于群体智慧的微博

流新事件检测算法和基于情感突发的在线突发事件检测模型；在社区发现与演化分析方面，提出了不完全信息社交网络的社区发现方法及结合社区属性熵的社区发现方法；在网络舆情态势分析方面，提出了微博事件中的公众关注点演化分析算法、基于多关系网络的话题层次影响力个体挖掘、基于多元情感向量的情感演化分析技术，以及基于爆炸性渗流的信息传播分析方法。基于上述关键技术的突破，开发了“中科天玑网络舆情系统”等产品和系统，在国家舆情分析领域中得到应用验证。

项目在国内外本领域重要的学术刊物和会议上发表论文 196 篇，发表论文数是计划的 2.45 倍，其中在 IEEE TKDE、Physical Review E、IEEE Translation、PLoS ONE 等国际一流刊物发表 18 篇，在 SIGIR、CIKM、WWW、NIPS、WSDM 等 Web 搜索与挖掘国际顶级学术会议上发表 21 篇，其中，SCI 检索 33 篇、EI 检索 108 篇。另外，有 4 篇会议文章获得最佳论文 / 学生论文奖，包括《查询意图感知的相似度度量方法》获 CIKM 2011 最佳论文奖，《Top-k 排序学习》获 ACM SIGIR 2012 最佳学生论文奖，《一种基于表情符号的在线突发事件检测模型》获 NDBC2013 最佳

论文奖,《增强型协同知识获取及共享系统研究》获第 27 次全国计算机安全大会最佳论文奖。项目申请专利 33 项、获得授权专利 8 项,是计划的 4 倍以上;获得授权软著 17 项、获批国家通信行业标准 3 项,亦超过计划的 4 倍;获国家科技进步二等奖 1 项、电子学会电子信息科学技术一等奖 1 项;培养博士硕士研究生 81 人,博士后 6 人,为本研究领域培养青年学术带头人 11 人,培养研究生数是计划的 2.7 倍,培养青年带头人是计划的 3.6 倍。此外,项目执行期间,项目组

共主办 / 承办国际、国内学术 / 技术会议 13 次,共 7000 余人次。具体组织了第四 - 六届 (2010-2012) 中国 Hadoop 与大数据技术大会 (HBTC) ,第六 - 九届 (2010-2013) 全国信息检索学术会议 (CCIR) ,第一届 (2013) CCF 大数据学术会议,第一、二届 (2012-2013) 未来数据论坛 (FFD) ,第一届 (2013) 大数据挖掘与理解国际 Workshop (BigMUD) ,以及以“社会网络”为主题的计算机学会第 25 期学科前沿讲习班等学术 / 技术会议。



2013 年第一届 CCF 大数据学术会议现场



## 结题验收的代表性成果

# 基于云计算的海量数据挖掘

本项目为国家基金重点项目（课题编号：61035003），起止时间是2011年1月至2014年12月。经过四年的研究，项目组全面完成了研究计划，并顺利通过验收。取得了如下成果：

## 1. 基于云计算的海量数据挖掘方法和算法

提出了利用数据内在结构的特征挖掘框架和结构正则化的SVM算法。在海量数据挖掘中，充分利用各种数据所蕴含的内在结构信息能有效提升数据挖掘性能。据此提出了一个统一的能充分利用数据内在结构的、大规模稀疏学习特征挖掘框架。该框架将各种结构先验，如光滑结构先验、继承结构先验、群组结构先验等作为惩罚统一集成到目标函数之中，进而发展出了相应的有效优化技术，实现了快速优化，最终达成了针对小样本大维数问题的高性能结果。

提出了多模态特征表示和多视图特征选择的海量视频图像数据挖掘方法。针对计算机视觉应用中，视频与图像数据的特征复杂、异构、动态等特点，我们着重研究了多表示、多层次的特征学习与融合方法，提出了基于空间约束、兼顾局部与全局结合的特征选择方法，研究了不同模态、不同类型特征间的相互关系及其融合方法，同时分别从时间与空间两方面分析了

视频行为规律。由此使所提方法能够更好地挖掘视频图像数据中最具判别性的特征，有效支持了医学诊断与视频异常检测等应用。

提出了混合生成式和判别式模型的图像语义标注方法。为克服训练图像关键词较少问题，提出了混合生成式和判别式模型的图像语义标注方法 HGDM。该模型采用连续PLSA 建模图像特征，并利用图像的主题分布作为中间表示向量，由此将图像标注问题转化为一个多标记学习 (multi-label learning, MLL) 问题。HGDM 使用集群分类器链方法进行多标记学习，在构造一个集群分类器链过程中能同时学习标记间的关联信息，从而不仅使 HGDM 能获取各个语义关键词在每幅未知图像中的置信度，同时还具有生成式模型和判别式模型的综合优势，标注和检索精度高且易扩展。

提出了跨领域迁移学习算法和典型相关性分析方法。基于特征映射迁移学习思想，提出了一种跨领域典型相关性分析 CD-CCA 算法。该算法在保持各领域自身特有特征与领域共享特征相关性基础上，通过选择合适的基向量组合训练分类器，使降维后的相关特征在领域间具有相似的判别性。

提出了一种并行数据挖掘方法，利用数据库来模拟链表结构，管理挖掘出来的知识，提供了树形结构、图模型的分布式计算方法，提出一种在 Hadoop 上高效数据挖掘框架。

## 2. 海量数据预处理

提出了将粗糙集中的属性约简问题转换成优化问题；利用数据的内在结构实现高维数据的维度约简和稀疏化；提出了具有最小堆结构约束的优化技术和快速优化算法。

## 3. 基于云计算的海量数据挖掘按需服务

基于强化学习和多智能体技术，提出了面向多层次用户云服务的系统资源分配和任务调度的方法和策略；提出了一种全新的 web 服务分解算法，将 QoS 优化和冗余服务去除两个功能整合到一起；针对大规模离散（或连续）状态空间强化学习问题，提出了基于分段线性基的时间差分学习方法 PLVF-TD。

## 4. 面向海量数据挖掘的云计算模式

研究了存储虚拟化，提出了一种分布式存储技术：GCFS 云存储系统。对其系统架构采用分布式架构进行设计，对系统可用性、数据安全、数据去重等关键技术进行了研究。

## 5. 研制了大数据挖掘云引擎 CBDME

大数据挖掘云引擎系统（cloud-based big data mining engine, CBDME）综合利用云计算技术、大数据挖掘技术、并行分布处理技术、自治计算技术、服务计算技术等方面

的最新研究成果，基于云计算的大数据挖掘理论和方法，在云计算环境下实现跨域、异构、动态大数据的集成、分析、挖掘，以满足行业实际应用的需要。CBDME 系统总体架构分为四个层次，即云平台、数据工厂、智能大脑和应用服务。该系统已取得软件著作权，登记号 2014SR163925。该系统及相关技术已应用于开发社会舆情分析、视频点播推荐系统、跨媒体检索等，也可应用于智慧城市大数据知识中心建设。

课题组共发表学术论文 203 篇，其中 SCI 收录 70 篇；获国家自然科学二等奖 1 项；获发明专利 1 项，申请专利 23 项，获软件著作权 3 项；培养博士后 2 人，博士 21 人，硕士 23 人；项目组成员在国际学术会议作特邀报告 6 次。课题负责人史忠植获得 2013 年中国人工智能学会吴文俊人工智能科学技术成就奖（见下图）。





## 结题验收的代表性成果

# 绿色网络 – 降低网络能耗的理论与技术

本项目是国家基金委重点国际合作项目（课题编号：61020106002），起止时间是2011年1月至2014年12月。2015年1月通过了基金委组织的会议验收。

数量庞大的网络设备对能源的消耗，已成为一个急需解决的对于环境和经济发展具有重要影响的重大社会问题。降低网络能耗，构建绿色网络已经成为当前计算机领域一个意义重大、迫切需要解决的关键问题。传统网络系统设计的两个原则，超额资源供给和冗余设计，有悖于节能的目标。网络服务节点规划与部署的不合理和网络的低效使用是影响网络能耗的两个重要因素，其中如何部署网络节点的位置及规模，如何从网络全局的角度设计路由和调度算法以尽可能使网络设备能耗与其负载流量成正比，是当前的研究趋势。

本项目的主要成果如下：

**1. 网络服务节点的规划与部署：**针对虚拟化数据中心实时迁移所面临的挑战，提出了虚拟化数据中心的能源高效动态资源供给机制，实现

服务器整合和资源供给，达到节约能量、维持较好的负载均衡的目的；针对现有部署方案的不足，利用部署服务器所需要的全局网络信息、综合考虑用户的体验和提供服务的成本，提出了新型服务器部署方案，能提高内容分发网络的使用效率，减少所需服务器数量，降低资源消耗。

**2. 网络能耗模型与能效算法：**基于数据传输模式和能耗机制，构建了网络能耗系统优化模型，对5种关键系统模型给出了详细描述；提出了具有完成时间限制的网络节能调度/路由算法、基于延迟与能耗双目标优化的分布式路由算法，特别是针对网络最坏情况下，提出一种普适稳定且能耗优化的调度算法；结合虚拟机布置和流量

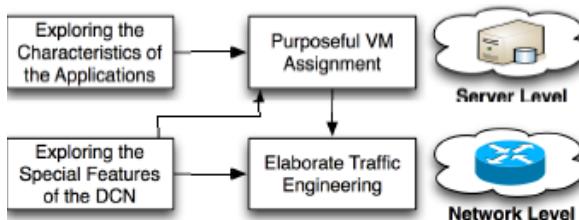


图1 GreenDCN：数据中心网络的普适优化框架

工程方法, 提出一种数据中心网络的普适优化框架 GreenDCN (图 1), 可应用于数据中心网络中的多种优化问题 (能耗或性能等)。

**3. 绿色网络的评价指标与度量:** 首次提出绿色网络的评价体系框架, 通过模型-评价-优化-实施这一循环过程, 逐步优化网络的服务质量和能效之间的权衡; 提出一种基于网络能耗模型的节能机制评价理论框架 (图 2)。将耗能资源划分成不同状态, 不同状态的能耗和服务能力不同, 通过低能耗状态实现节能, 并通过不同状态之间

的相互转移来实现 QoS 要求。

项目组在 IEEE JSAC、Trans 等国际期刊和本领域重要国际会议 (INFOCOM、SIGMETRICS 等) 上发表 (含已录用) 论文 91 篇, 其中在 SCI 国际期刊 (如 JSAC, TCS, TON, TPDS 等) 发表论文 32 篇, 在国际主流会议上发表论文 51 篇; 申请国家专利 7 项。培养研究生 38 名。项目组通过积极参与相关领域主流国际会议、邀请国外专家访问等方式, 与国外同行进行了深入的学术交流与合作。

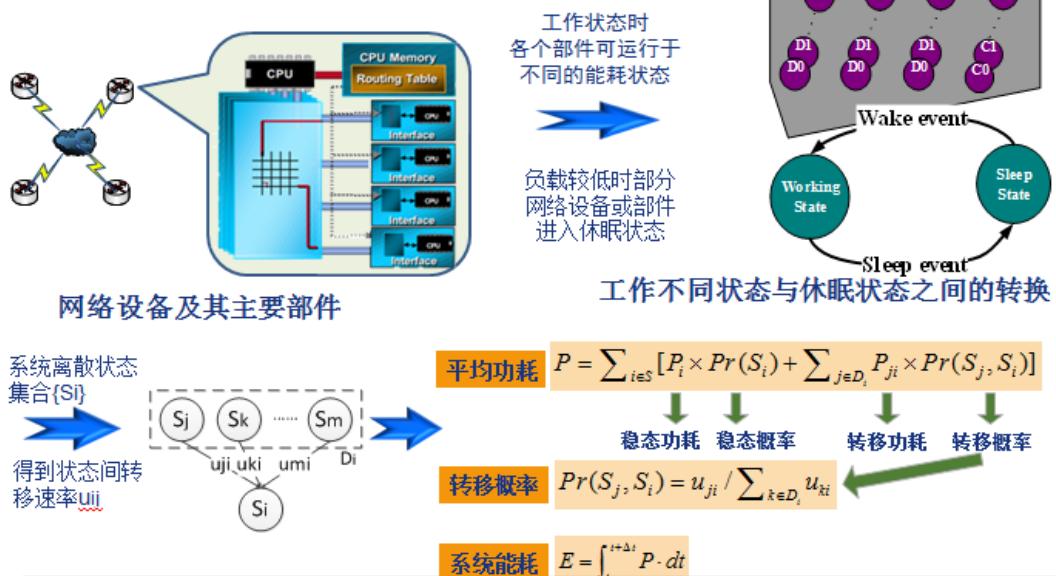


图 2 基于网络能耗模型的节能机制评价理论框架



## 结题验收的代表性成果

# 未来异构无线网络信息理论与关键技术研究

“未来异构无线网络信息理论与关键技术研究”是由中国科学院计算技术所牵头承担的北京市自然科学基金重大项目（项目编号：4110001），起止时间为2011年1月至2014年12月，于2015年1月20日顺利通过北京市验收。

围绕项目任务书的研究规划，本项目就异构无线网络信息容量理论分析和资源管理、自适应协作通信、智能控制等关键技术展开了研究，并搭建了相应的仿真平台进行技术验证，主要研究内容及具体成果：

## 1. 异构网络无线信息理论

利用随机网络理论，采用易处理的理论分析方法，给出了分层异构无线网络的成功覆盖概率，分析了分层异构无线网络的干扰分布特征，提出了遍历容量闭式解公式，为异构无线网络容量性能优化提供了理论指导；并面向分层异构无线网络的网络编码方案，对中继网络编码的多维资源最优化理论进行了研究，为异构网络中进行网络编码优化提供了理论依据。

## 2. 高效异构无线资源管理机制

面向宏蜂窝和D2D的双层异构网络就功率分配进行了研究，融合广播网与通信网，研究D2D高能效的多播传输机制，并从系统能耗、信号合并方式、用户密度因素的角度对提出的方法进一步分析优化，有效地提高了宏蜂窝和D2D双层异构网络的能耗性能；此外，针对双层异构高铁无线通信网络的切换机制研究，面向高铁双层异构网络，提出一种双层双链路的无缝切换机制，有效降低切换失败率，提升用户体验。

## 3. 异构网络自适应协作通信理论

首先，分析了以用户为中心的异构网络自适应CoMP性能，利用随机网络理论对整个系统的干扰性能进行了建模与分析，推导得到了采用CoMP协同通信的系统覆盖率、最佳协同基站数和系统容量等性能指标。然后，研究了面向异构网络层间干扰的协作传输机制，提出了一种能够实现在资源块级别的协作传输技术来解决异构网络中存在的严重层间干扰问题。

## 4. 智能异构无线网络协同控制方法

针对LTE和WiFi的异构网络，提出基于FFT

的智能业务分流，实现不同系统间的资源动态按需分配，在最大化系统业务提供能力的同时提高频谱资源利用率；同时研究了基于用户的异构无线网络智能接入选择机制，针对用户在不同通信场景下多种业务的实际 QoS 需求，给出了异构无线网络智能接入选择方法。

### 5. 面向未来高速宽带异构通信网络的仿真验证平台

为了验证上述理论和机制的研究，本项目搭建了面向未来高速宽带异构通信网络的仿真验证平台，支持多种异构无线网络的业务、资源管理算法以及协作处理等功能。

项目组在任务执行期间，共发表论文 40 篇

（两篇获得最佳论文奖），申请发明专利 9 项，申请相关软件著作权 2 项，完成技术报告 12 份，并完成了面向未来高速宽带异构通信网络的仿真验证平台 1 套；此外，本项目在异构无线网络研究领域共培养博士生 8 名，培养硕士生 56 名，晋升高级职称研究人员 5 名，全面完成了项目目标。

基于本项目的研究成果成功申请到国家自然科学基金项目，并研制了支持异构网路协同的超级基站原型设备，在专网领域开展了应用，为北京市未来无线网络产业发展的部署提供了科学指导和技术支撑，为推动从北京制造向北京创造转型贡献力量。

自 主 创 新  
科 技 强 国



## 结题验收的代表性成果

# 云计算 I/O 服务器—VIO Server

“云计算 I/O 服务器—VIO Server”是计算所 2011 年 5 月部署的所创新课题，并于 2014 年 8 月通过结题验收。云计算技术通过将大量的计算、存储、网络资源整合并虚拟池化，向用户提供按需的动态服务，进而有效地降低 IT 支出、减少能源消耗、提升运营效率，因此，云计算技术已经在各种数据中心、超算中心的环境中取得了广泛的应用。

但是当前云计算基础设施架构还存在着许多不足之处，制约了云计算系统的服务能力和效能的进一步提升：

- ◎ 传统服务器（单操作系统域）中各种 I/O 资源（包括加速计算、存储、网络等）与服务器紧密绑定，一方面导致服务器间 I/O 资源冗余，另一方面限制了跨服务器 I/O 虚拟资源的分配；

- ◎ 整机柜交付的高密度计算部件部署方式极大地限制了 I/O 资源的扩展。因此，亟需探索一种解耦合 I/O 系统结构，通过将 I/O 资源网络化，从硬件层次解除 I/O 资源与服务器间的紧密耦合，实现 I/O 资源的灵活扩展、聚集和共享。

在计算所创新课题的支持下，研究工作突破了构建解耦和体系结构的核心使能技术—多根 I/O 虚拟化和基于 I/O 总线扩展的高效互连。其中，全硬件多根 I/O 虚拟化机制，实现了透明的设备虚拟和多 OS 域 I/O 空间管理映射，无需对软件进行任何修改，即可将单个物理 I/O 设备虚拟为若干虚拟 I/O 设备，同时为多台服务器（或虚拟机）提供近似设备原生性能的 I/O 服务。为实现虚拟资源在服务器间的流动，本课题还提出了虚拟热插拔机制，实现在线虚拟资源分配。但服务器最终获得的 I/O 服务性能，取决于它与 VIO Server 间的网络通信效率，因此本课题定制了基于 PCIe 总线扩展的网络协议，消除了 I/O 总线与网络间的协议转换开销。此外，该网络的远程 Load/Store 和 RDMA 通信机制，使其应用于服务器间、服务器与 VIO Server 间、VIO Server 内的高速互连。

上述虚拟化和互连机制均由 VIO Server 的核心控制芯片——cHPP 控制器实现，具体负责 I/O 设备的多根虚拟呈现、互连网络协议处理和

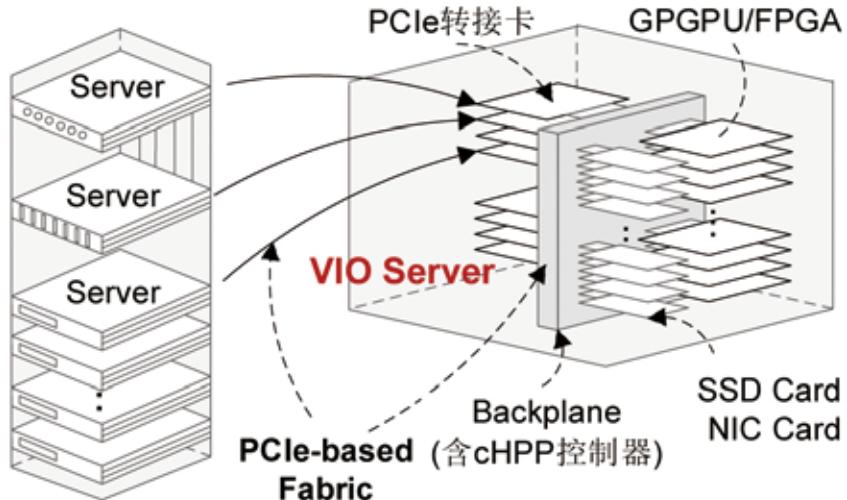


图 1 VIO Server 架构

内部数据交换。基于 cHPP 控制器的 VIO Server 架构如图 1 所示：VIO Server 对外、对内互连均采用基于 PCIe 扩展的互连网络（PCIe-based Fabric），cHPP 控制器（Backplane）位于背板之上连接所有 I/O 设备，并通过 PCIe 转接卡与服务器互连。

为了验证 VIO Server 的功能和性能，本课题实现了 10 端口 VIO Server 原型系统，如图 2 所示。其中 10 端口 cHPP 控制器芯片基于 Xilinx XC7VX690T FPGA 实现，每个端口采用峰值带宽 40Gb/s 的 PCIe Gen2 x8 链路，既可用于设备扩展，也可用于与服务器互连。服务器的上层软件实现融合通信中间件，全面兼容 MPI 和 TCP/IP 协议栈，同时实现基于虚拟热插拔技术的在线虚拟资源分配功能，用于 I/O 服务能力的弹性部署。

基于 PCIe 存储（Intel SSD 卡）和网络设备（Intel 82599EB 万兆网卡），在原型系统对多根 I/O 虚拟化功能进行了验证。测试结果表明，VIO Server 可以向多台服务器提供公平的 I/O 服

务。以共享 Intel 82599EB 为例，当仅有一台服务器使用时，该服务器可获得 9.41Gb/s 的以太网传输带宽，与该网卡的实测原生带宽相同，当 4 台服务器共享使用时，每个服务器获得近似平分的传输带宽，均在 2.35Gb/s 左右。此外，在线虚拟资源分配测试表明，虚拟热插拔机制可以实现快速的设备接入和移除，其中动态接入延迟为 127ms，移除延迟为 30ms。

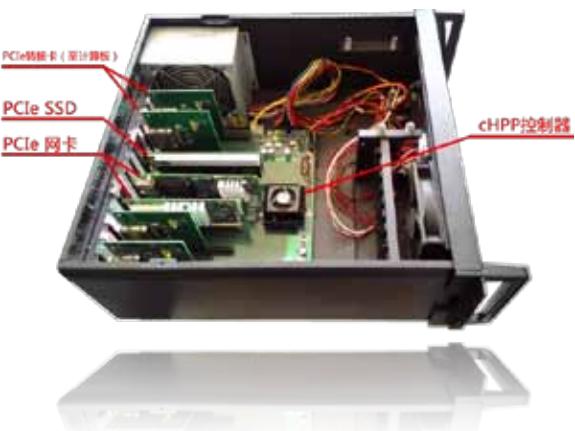


图 2 VIO Server 10 端口原型系统图



## 结题验收的代表性成果

# 高通量计算关键技术的研究

“高通量计算关键技术的研究”是计算所 2011 年 5 月部署的所创新课题，并于 2014 年 8 月通过结题验收。本项目实施高通量 (HTC, High Throughput Computing) 方向十二五规划中的第一阶段计划，其目标是在两年之内掌握设计高效能的高通量服务器的一些关键技术，实现高效能、高可靠的高通量计算需要在软硬件的各个方面进行创新，为了降低项目的风险，考虑现有资源的限制。

本项目的主要技术成果体现在：完成高通量服务器 (DC3.0) 详细系统设计，包括系统架构、内外存储以及数据中心网络；搭建多套硬件平台，包括 48 节点 ATOM Cluster, 200 节点 Xeon Cluster, PCIe Switch 混合系统两套，桌面云测试环境一套；累计发表论文 30 余篇，关键技术成果主要发表在 HPCA、CGO、PACT、IISWC、ISPASS、Euro-Par、IPDPS 等国际会议。累计申请并受理发明专利 100 余件。

本项目主要关键技术完成情况如下：

1. 高通量计算测试床系统。开发成功高通

量计算测试床系统，该测试床系统底层基于科研人搜索系统，包含数据处理和数据服务两类典型高通量计算负载；测试床系统已经收集了 10TB 数据，包含 12 万研究人员的信息以及 4 千万个主页。科研人搜索系统已经有超过 4 百万的访问量，能支持每分钟十万次请求处理、每天 TB 级别的数据处理。在科研人搜索平台的基础上，该测试床系统支持不同类型的测试，包括 Xeon, Atom, Tilera 等丰富的硬件平台、后台管理系统调度监测分析等、统一资源管理测试。

2. 基于 PCIe Switch 的 X86 原型系统。设计与实现基于 PCIe Switch 的混合系统，通过支持底层通信库、虚拟网卡实现数据中心内跨节点间资源共享。基于该系统可实现借用远端节点的内存（直接通过 load/store 访问或通过 DMA 方式访问虚拟磁盘两种方式）、借用远端节点的 GPU、借用远端节点的网卡。基于 PCIe 的虚拟磁盘比基于 InfiniBand 的虚拟磁盘带宽提高 3 倍、IOPS 提高 4 倍、延迟缩短为 1/3；远程共享 GPGPU 可达到本地 GPGPU 的 4%~95% 的性能；

通过 PCIe 互连可以比 10GbE 以太网互连在带宽上提升 2.6 倍、延迟上降低 22.5 倍。

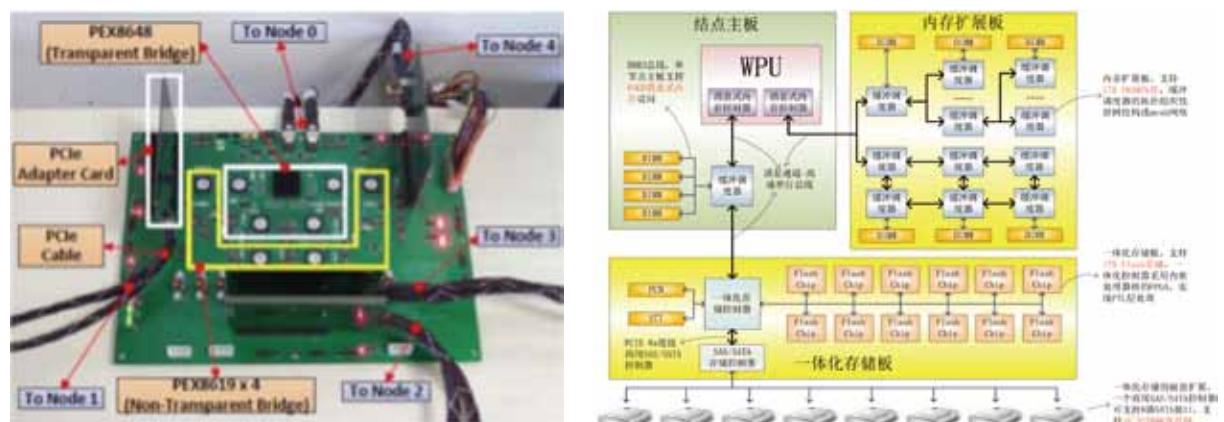
3. 标准测试集。从四大典型领域（搜索、社交网络、OLAP、电子商务）入手，一方面分析提取数据特征，开发了数据生成工具 BDGS，另一方面对计算单位进行抽象，形成了 77 个负载扩展。进一步为了便于体系结构模拟器级别研究，本项目对 77 个负载进行分析，最终约减为 17 个具有代表性的负载。基于上述研究成果，本项目发布 BigDataBench，有超过 6500 次访问，包括英国剑桥大学等国际著名研究机构均开始使用该 Benchmark。

4. **高通量芯片**。首先对网络服务类应用和网络安全类应用的特性进行分析抽象，形成微测试程序，并用此来指导完成 DPU 架构设计。DPU 架构特色之一为由双环网络加锥形网络形成的低延迟通路。DPU 采用多线程并发处理器核架构来容忍延迟。通过硬件机制保证实时性需求，实现多级任务分发机制，从而缩短任务实际执行时间，从而有效避免了任务的失效。首创性地提出访存请求收集机制，可以高效利用访存带宽，尽管访

存平均延迟有所增加，但应用整体执行时间提高 1.9 倍。

5. 消息式内存。完成消息式内存总体架构的设计，以及支持级联的缓冲调度器设计。消息式内存原型系统所能达到的性能指标包括：LRDIMM 控制器支持单条 16GB，总计可插 16 条，共 256GB；逻辑工作频率 200Mhz；母板上缓冲调度器与设备之间的总带宽可达 51.2GB/s(SERDES) + 25.6GB/s(LRDIMM)，端口路由延迟 5 cycles（仿真结果）；子卡上缓冲调度器与设备之间的总带宽可达 25.6GB/s(SERDES) + 19.2GB/s(LRDIMM)。

6. 数据中心网络。提出了一个两层的数据中心网络设计方案，第一层是 Hoffman-Singleton 网络，最多 50 个 WPU 构成一个大点，该网络构成度为 7，直径为 2 的 Moore 图，具有编码和路由简单的特点，可划分为 10 个子 Moore 图，可通过逻辑计算获得每一跳出口端口号，省去了路由表的开销。第二层是以优化网络直径为目标的自由扩展，采用全局负载均衡的路由策略，使用“零延迟”路径池技术来减小流调度开销。



基于 PCIe Switch 的 X86 原型系统及消息式内存系统



## 结题验收的代表性成果

# 支持未来网络创新的可编程 虚拟化路由器平台

“支持未来网络创新的可编程虚拟化路由器平台”是计算所 2011 年 5 月部署的所创新课题，2014 年度课题完成了软硬件平台、部分功能组件以及设备管理系统的研发与测试，并于 2014 年 8 月通过结题验收。

## 1. 硬件系统

1) 研发了全新的 ATCA 众核数据包处理板（如图 1 所示），取得了近十项功能升级。该板卡换用了 40G 端口交换芯片 BCM56842 取代 BCM56803，实现了板卡与背板的 40G 交换；换用了新型接口 MAC 芯片 BCM8726，实现了前置端口的千兆万兆兼容；重新设计了后置 RTM 端口与前板的连接方式，实现了端口接入模式的统一；设计了全新的处理器硬存储系统，实现了存储容量的扩展；优化了处理器内存走线，实现了对大容量内存更好的兼容；优化了主从处理器之间的连接方式，实现了高速主从通信。2) 研发了单芯片处理板系统，作为接入设备的硬件平台。该系统提供 12 个千兆端口和 1 个万兆端口。支持 36 核、16 核、9 核处理器的兼容。

## 2. 平台软件研发

1) 优化了平台的虚拟化与可编程功能，实现了更高效的虚拟路由器实例创建与启停管理，提供了更方便的编程接口 API。2) 提升了路由协议一致性，通过大量测试与编码优化，将 BGP、ISIS、OSPF 等主要路由协议的一致性大幅提升，通过 TestCenter 测试可以达到 90% 以上。3) 通过优化算法大幅提升了路由平台的处理能力，单业务处理板能够达到处理 64 字节小包 20Gbps 线速路由查找转发的能力。4) 完善了路由器设备管理接口，提供更丰富的北向接口，支持更高效的设备管理。

## 3. 功能组件

1) 研发了网络入侵检测功能组件，该功能组件能够在线对数据包进行深度解析，实时监测网络上发生的各种入侵攻击行为，并进行阻断与记录。该功能组件能够根据用户需要动态开启关闭，并允许针对不同用户加载不同的安全过滤规则，实现用户定制的安全服务。2) 在路由器平台上部署了新型网络体系结构 SOFIA 协议栈，验证

了路由器平台对新协议的支持。

#### 4. 设备管理系统

1) 设计实现了路由器资源管理系统 PRMS (如图 2 所示)。该系统服务于物理网络管理员, 能实现对网络物理设备资源的集中分配与调度, 并对资源申请进行审核。2) 设计实现了路由器用户管理系统 PUMS。该系统服务于虚拟网络用户。它提供用户对虚网的管理接口, 包括申请创建虚拟平面, 申请加载不同的功能组件与应用等。用户通过 PUMS 提交的申请会在后台直接送达 PRMS, 由管理员进行申请的审核与批复。

工业和信息化部电信传输所的第三方测试结果表明, 可编程虚拟路由器功能与性能均达到预定指标。当前设备已经在中国科技网及一些专网进行部署和内测, 为网络技术创新提供良好的支持。



图 1 路由器众核数据包处理板



图 2 路由器资源管理系统



## 结题验收的代表性成果

# 机器博士——智能技术集成平台

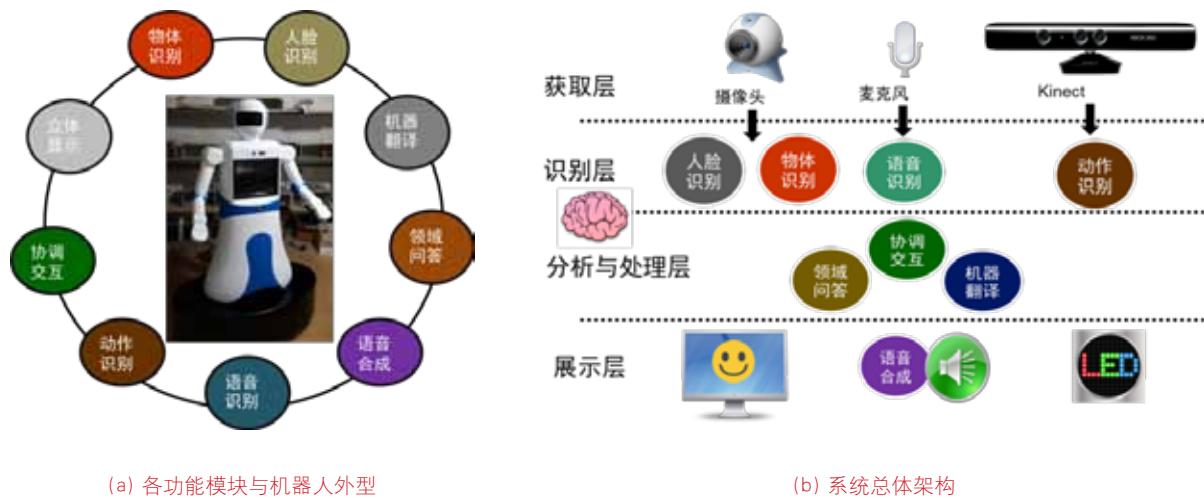
“机器博士 -- 智能技术集成平台”是计算所2011年1月部署的所创新课题，于2014年8月通过结题验收。机器博士系统的研发目标是利用中科院计算所在体系结构方面的优势，集成相对成熟的智能信息（图像、语音、文字等）处理技术，突破内容分析、系统架构、模块集成、交互展示等相关的关键技术，构建机器专家智能系统和展示平台。该系统基于中科院计算所在智能信息处理技术方向上的已有技术积累，对领域问答、机器翻译、物体识别、人脸识别、三维交互、立体显示、智能协调等技术进行了有针对性的研究与开发，并在此基础上进行了系统集成，将各功能模块融合到一个系统中，使机器博士系统具有听觉、视觉等感知功能，能对感知信息进行智能处理与识别，并在此基础上实现了自然语言处理、信息检索、自动问答、语言翻译等功能，形成了不同模块之间的功能跳转，实现了多模态、多形式、多输出的智能交互，同时在软件集成的基础上，构建了硬件系统并设计制作了机器人内部结构和外形，形成了具有较高展示性的系统平台。

机器博士系统的各功能模块与机器人外型如图1(a)所示，系统总体架构如图1(b)所示。系统共分四层，在信息获取层，摄像头用以获取图像信息，麦克风用以获取语音信息，Kinect用以获取图像信息和深度信息，以及人体运动信息等。在识别层，人脸识别技术以及物体识别技术根据获取的图像信息进行识别，对感知的场景进行理解，实现机器博士的视觉功能。语音识别技术根据麦克风获取的音频信息进行语音的识别，对人物对话信息进行理解，实现机器博士的听觉功能。动作识别技术根据Kinect获取的RGB-D信息以及增强的人体骨架与运动信息，对人体动作和手势进行识别，获取人体交互信息，作为机器博士视觉功能的补充。在分析与处理层，首先协调交互技术对语音识别等视听觉功能获取的信息进行分析，实现功能调度与智能协调，领域问答主要侧重于地理知识，同时也支持寒暄与新闻的对话，机器翻译支持中英语言的翻译。在展示层，系统通过语音合成将输出结果以语音的形式进行反馈，通过可交互的触摸屏进行结果的图形化展示，此

外，在机器人头部还有一个 LED 显示屏用于结果的辅助输出以及动画等结果的显示。系统在初始化运行后，首先进入 UI 主界面并进行等待，在此过程中以及后续的操作过程中，系统不断进行人脸检测与识别，若识别出数据库中的人物，将通过 LED 进行显示。根据获取层获取的信息，将通过语音控制、界面交互、概念转换、智能协调等方式，进入各功能模块，进行智能问答、机器翻译、

物体识别、手势识别等各功能的交互。

机器博士作为计算所自主创新开发的机器人系统，体现了计算所关键技术的先进水平，充分展示了计算所在人工智能领域的先进性。在此基础上，项目组正在推进机器博士 V2.0 的研究与开发。希望能够在现有版本的基础上，加强多模态融合反馈能力，研制出拥有更高综合性智能的集成系统。



机器博士系统示意图



## 结题验收的代表性成果

## 天玑学术网

“天玑学术网”是计算所 2011 年 5 月部署的所创新课题，并于 2014 年 8 月通过结题验收。项目旨在解决学术领域所面临的学术交流形态和范围的局限性、以及学术信息高效获取和传播的难题，由计算所天玑团队主导并开发出社会化学术引擎一天玑学术网（soscholar.com），现名学术圈。本项目致力于构建大规模学术资源深度搜索和服务平台，并在此基础上提供用户为中心的互动式学术交流平台。其核心思想是通过对互联网环境下的学术空间进行收集、观察、挖掘与组织，以新颖可视和个性化方式为用户展现学术世界，以实时互动的方式让用户参与在线学术生活。天玑学术网项目从 2011 年 9 月开始，历经三年的建设与发展，在 2014 年 9 月成功通过所创新课题验收。目前天玑学术网已经具备一定规模，形成一个集学术社交、学术应用和学术可视化为一体的学术平台。

天玑学术网在总体架构、学术服务、关键技术、基础性和应用性研究等几个方面均取得一系列成果。

## 总体架构

在总体架构方面，天玑学术网设计并实施“一体系四引擎”。“一体系”指天玑学术网系统总体的体系架构，该体系架构采用分层及分布式策略，包括 9 个服务器子群：负载均衡服务集群、Web 服务器集群、应用服务器集群、静态资源集群、状态管理与缓存管理集群、分布式搜索引擎、公共构件服务群和数据存储集集群。四引擎分别是数据处理引擎、搜索引擎、推荐引擎和分析引擎。该体系结构具备非常好的系统弹性，可以有效应对 1000+ 次 / 秒的并发访问压力。

“四引擎”指的是天玑学术网建设的四个核心引擎：数据处理引擎、搜索引擎、推荐引擎和分析引擎。1) 数据处理引擎包括数据采集流水线和数据融合流水线，数据采集流水线采用分布式框架，支持定向 / 非定向增量采集，目前已经有效采集获取到千万级学术资源。数据融合流水线采用来源信任机制，打造一条多源异构的实体融合流程。2) 搜索引擎融合天玑团队前沿的用户查询分析与理解学术成果，基于并改进实验室 I3Search 索引系统，实现高性能大规模学术数据

检索应用服务。3) 推荐引擎则是探索构建了新型可扩展信息推荐平台的系统架构, 融合了不同类型、不同颗粒度以及不同计算模式的前沿推荐算法包, 支撑在线系统多类型推荐需求。4) 分析引擎采用分布式架构, 基于实验室 Hadoop 和图数据库方面的系统成果, 实现了不同层次大规模分析挖掘算法。

## 学术服务

在学术服务方面, 天玑学术网对学术信息进行深入挖掘, 并融入社交协同因素, 开发多项特色学术服务, 达到了领域深度搜索挖掘与社交协同的有效融合。特色学术服务包括“社会化学术搜索”、“领域调研”、“学术人脉”、“学术评估”、“个性化论文推荐”和“圈内热点”等应用。“社会化学术搜索”, 包括搜索前的“热点关注”和“好友动态”的驱动, 搜索中的他人“点评”和“关注”的引导, 以及搜索后的“收藏”、“标注”、“分享”和“讨论”行为的产生, 三个阶段将学术搜索过程与社交协同有效的融为一体。“领域调研”针对特定领域学术资源进行深入分析挖掘, 展示该领域的发展趋势、领域专家、重要论文和综述论文, 方便学者快速掌握该领域的状况。“学术人脉”通过分析学术论文资源, 展示作者之间的合作关系, 用可视化的方式帮助学者更加清晰的掌握学术合作的脉络。“圈内热点”通过对新浪微博数据的实时采集与分析挖掘, 提取出社交网络中所关注的热点学术资源资讯, 提供给学术网用户。在交互模式方面, 天玑学术网在“学术搜索”里实现了“混合搜索”和“分面搜索”, 帮助用户从多个学术维度快速查找学术资源。天玑学术网还开发了“学术视界”子系统, 以新颖的方式帮助用户方便快捷的浏览学术对象

以及对象之间的关系。

## 关键技术

在关键技术方面, 1) 天玑学术网在“数据采集与处理”方面实现多通道学术资源采集框架, 该框架采用分布式架构, 融合实验室分布式、高性能和高可靠的列存储数据库 ICTBase, 采取解析插件动态化管理机制, 实现定向 / 非定向学术资源采集。天玑学术网实现多源数据融合框架, 该框架采用多源信任机制解决融合冲突问题, 并充分利用实体间内在的关联帮助实体有效的融合, 从而构建多源异构实体融合的流水线。2) 天玑学术网在“学术搜索”方面, 融合实验室前沿的查询理解与处理的学术成果, 形成查询优化、查询扩展和查询推荐等一系列应用模块, 在实验室 I3Search 索引系统基础上增加机器学习排序模块, 通过融合对维度特征, 利用先进的统计学习算法和分面搜索技术, 实现高性能大规模学术数据检索。3) 天玑学术网在“学术推荐”方面, 融合多项前沿的推荐算法(基于内容、协同过滤和排序学习等), 采用多模式混合推荐技术, 基于消息队列的推荐控制流程, 采取算法模块插件式管理机制, 并基于在线学习方式及时捕捉用户兴趣变化, 进而支持大规模推荐需求。4) 天玑学术网在“学术分析与挖掘”方面, 基于 Hadoop 平台实现分布式存储和分析, 基于实验室自主开发的图数据库(GDB)实现大规模关联数据计算, 从而构建不同层次的分析挖掘算法, 形成了学者消歧、学者排名、标签预测、关系挖掘等一系列应用模块。

## 性能指标

天玑学术网从多个数据源采集并融合, 目前



已积累学术资源已达千万级,其中论文1000多万,学者600多万,机构20多万,期刊会议1万多等,超过了预期的学术资源数据目标。目前系统服务器节点21个,存储容量30T+,数据总容量达到10T。在性能上,以最为核心和复杂的“学术搜索”为例,单机版环境下请求处理能力500次/秒,平均响应时间不超过0.8秒,分布式环境下可以轻松满足1000次/秒,平均响应时间不超过1秒的性能指标。系统目前已经积累1.6万左右的Web用户,1400多IOS用户,累计的独立访问IP达到13万多。

## 基础性和应用性研究

在基础性研究方面,天玑学术网在排序学习、个性化推荐和实体挖掘方面取得一系列成果,其中发表在国际会议论文4篇,国内核心期刊论文1篇,研究成果达到国际先进水平。其中,论文“Top-k learning to rank: labeling, ranking and evaluation (SIGIR'12) ”获得了SIGIR2012年最佳学生论文奖。在应用性研究方面,天玑学术网完成两个专利(已受理)和1个软件著作权(已授权)。

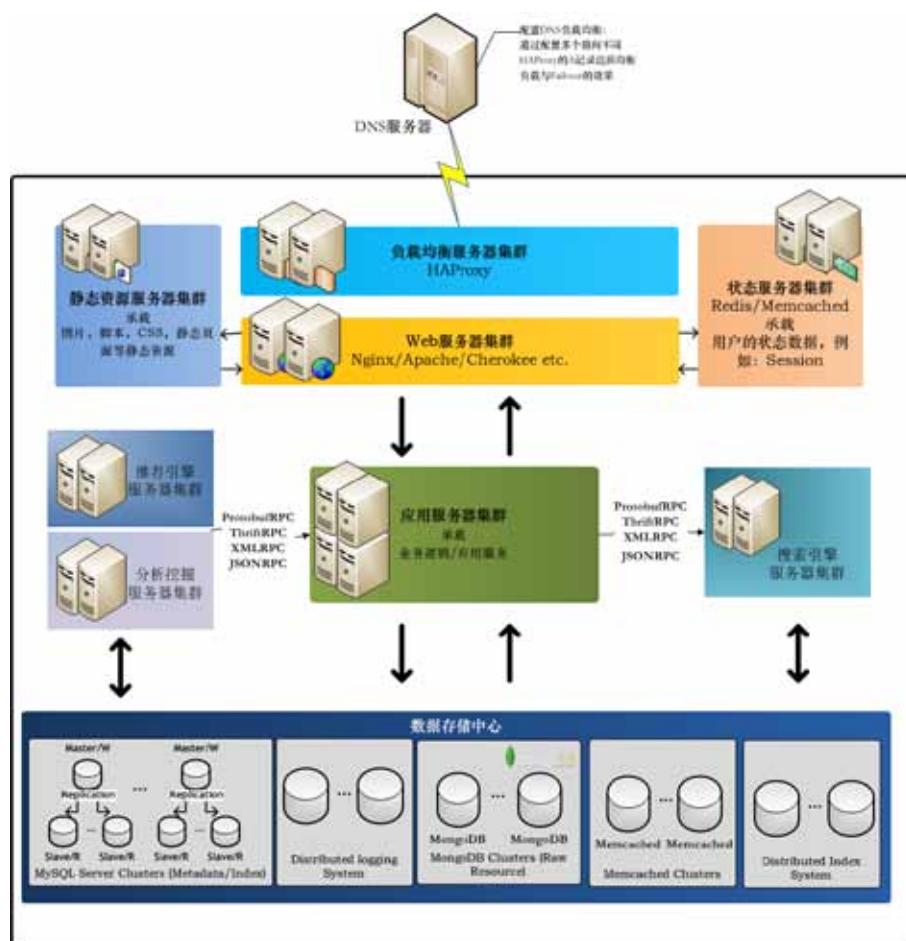


图1 天玑学术网系统架构

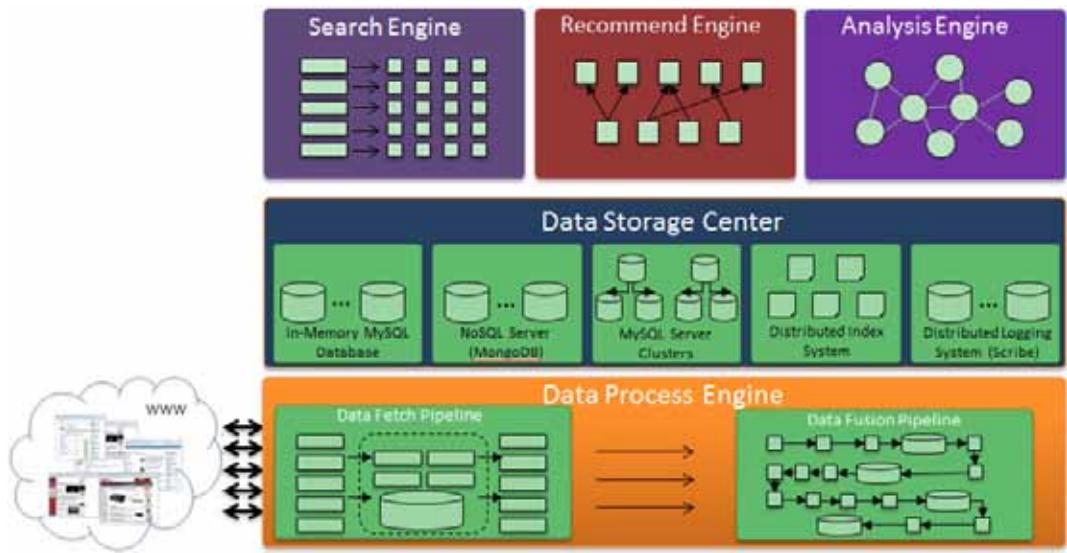


图 2 天玑学术网四大核心引擎

正 气  
大 气  
骨 气



## 科研基地进展

# 计算机体系结构国家重点实验室

获奖及科研项目方面：隶属于计算机体系结构国家重点实验室的集成电路团队的“32位星载容错控制计算机系统关键技术及应用”获得2014年度北京市科学技术奖一等奖（第二完成单位），该项目已应用于65颗以上卫星，占同期国产卫星飞船的85%以上，包括嫦娥、载人飞船、通信、导航、科学试验、遥感等多种空间飞行器，其中最长已在轨工作七年半，累计在轨超过一百余星年。项目成果取得了显著的综合效益，并具有重大的战略意义。

2014年国家自然科学基金资助项目公布，实验室获批资助13项。张云泉研究员、李华伟研究员、陈云霁研究员获批资助重点项目，Olivier Temam教授、姚二林副研究员、陈天石副研究员分别获批面上项目，袁良助理研究员、王蕾助理研究员、董建波助理研究员、江涛助理研究员、马宜科助理研究员分别获批资助青年科学基金项目，孙凝晖研究员、陈云霁研究员获批资助国际（地区）合作与交流项目。

973项目“高通量计算系统的构建原理、支撑技术及云服务应用”，在高通量计算系统的微

结构设计、系统结构设计、可靠设计、云计算服务与应用等方面进行了探索研究，本年度进展良好，超额完成了预期目标，将于2015年进行验收。

创新研究群体科学基金“超并行高效能计算机体系结构与设计方法研究”进入第二个执行期（2013年到2015年）。该研究群体以高效能计算、低功耗计算、高可靠计算为研究方向，在高性能和高效能计算机的超并行体系结构，多核/众核处理器的新型体系结构，以及多核处理器的测试、验证与高可靠性设计方法等六个方面上开展研究，本年度项目进展顺利。

实验室2014年共承担各类研究课题87项，其中973计划1项（课题4项）；863计划1项（子课题4项）；国家自然科学基金项目38项（其中创新群体1项，优秀青年基金1项）；横向项目19项。

学术进展方面：国重未来计算研究组的陈云霁研究员、Olivier Temam研究员、陈天石副研究员领导的团队在类脑计算机方向取得重要的进展。其研制了一种采用人工神经网络架构的机器学习运算装置——寒武纪1号，通过高效

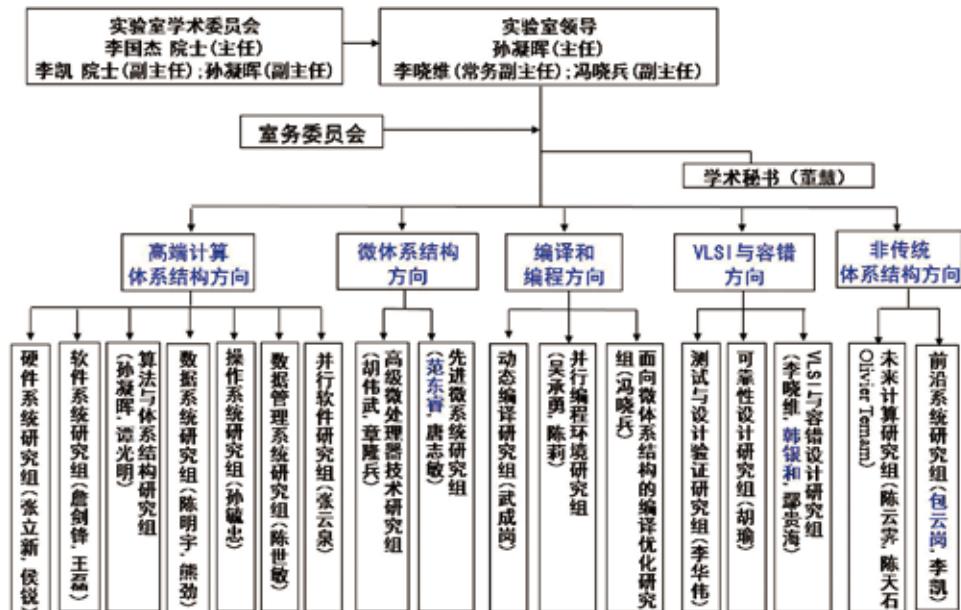


图1 计算机体系结构国家重点实验室组织结构图

的分块处理和访存优化，能高效率处理任意规模、任意深度的神经网络，以不到传统处理器1/10的面积和功耗达到了100倍以上的神经网络处理速度，性能功耗比提升了1000倍。相关论文 (Tianshi Chen, Zidong Du, Ninghui Sun, Jia Wang, Chengyong Wu, Yunji Chen, and Olivier Temam, "DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning") 2014年3月发表于本领域最好的国际会议之一 ASPLOS (ACM International Conference on Architectural Support for Programming Languages and Operating Systems)，并获得了最佳论文奖，这是亚洲科研机构首次获此奖。2014年12月，后继工作 DaDianNao: A Machine-Learning Supercomputer 再次获得本领域顶级会议 MICRO 的最佳论文奖，这是 MICRO47年来美国之外国家首获此奖。

实验室持续发表了 ISCA-2014、MICRO-2014、ASPLOS-2014、DAC-2014、ICCAD-2014 顶级会议论文，在 IEEE TVLSI 和 ACM TACO 等顶级期刊上均有学术论文持续发表。本年度实验室研究人员发表论文 85 篇，其中 EI 收录 41 篇，SCI 收录 20 篇，在国际、国内学术会议上做报告 49 人次，主办承办大型学术会议 3 次，暑期夏令营 1 次，大中小学生参观访问共计 3 次。获授权国家发明专利 6 项。

承办重要国际会议及学术交流方面：2014年11月16日-20日，国重集成电路团队在杭州成功组织召开了第23届 IEEE 亚洲测试学术会议 (IEEE 23rd Asian Test Symposium, 简称 ATS' 14) 及第15届 IEEE 寄存器传输级及高层国际研讨会 (简称 WRTL' 14)。本次会议由 IEEE Computer Society, Test Technology Technical Council 和中国科学院计算技术研究所 (简称计算所) 共同主办。集成电路实验



图 2 国重主任向学术委员会汇报

室李华伟研究员担任 ATS' 14 大会主席, 韩银和研究员担任 ATS' 14 程序主席, 张磊副研究员担任 WRTLT' 14 程序主席。本次 ATS' 14 和 WRTLT' 14 会议共吸引了 140 余位境内外专家学者参加, 其中境外代表 80 余位。

2014 年 6 月 25 日至 26 日, 国重组织承办了第 5 届亚太计算机系统国际研讨会 ACM APSys-2014 (<http://acs.ict.ac.cn/apsys2014/>)。包云岗副研究员出任大会共同主席。本届 APSys 大约有 100 位与会人员, 来自 MIT、Yale、EPFL、清华大学、北京大学、上海交通大学等国内外研究机构的研究人员。本届 APSys 与 ICML 联合召开, 吸引了更多对计算机系统、机器学习、大数据分析等感兴趣的人员参加。

2014 年 9 月 11 日至 12 日, 国重组织承办了第 4 届面向大数据的体系结构与系统国际研讨 会 ASBD 2014 (<http://acs.ict.ac.cn/asbd2014/>)。张立新研究员担任大会主席, 包云岗副研究员出任大会共同主席。本届 ASBD 大约有 100 位与会人员。华为公司在 ASBD 研讨会上发布了计算所 - 华为联合实验室的最新研究成果“第三代数据中心”白皮书, 举得很大反响。

本年度内, 国内外学者来实验室做学术交流

报告共 24 次, 就研究方向、课题合作、学术探讨、联合参会等开展了深入的实质性合作; 在国际、国内学术会议上做报告 49 人次, 公众开放活动 3 次。

国重开展的计算机系统结构先进课程(系列讲座), 作为 2014 年人社部北京分院国家专业技术人才继续教育基地培训点培训项目之一, 到目前为止, 进展顺利, 并被北京分院在官方网站上进行了报道。2014 年, 授课和研讨交流 24 人次, 共计近 50 个学时。其中海外专家学者 8 人次, 国内知名学者 13 人次, 知名企业界学者 3 人次, 同比去年增长了 60%, 有近 600 人次专业技术人员, 科研人员, 研究生参与了系列讲座, 与专家学者们互动交流。

2014 年度实验室对外发布了 21 项申请指南, 通过组织专家进行评审, 择优资助十项开放课题的申请。2014 年 11 月, 实验室召开开放课题学术交流会, 组织了正在执行期的开放课题承担者来实验室进行学术探讨和交流。来自清华大学、无锡江南所、中国农业大学、北京交通大学、天津大学等八家单位的老师分别就自己的研究内容、研究进展以及跟实验室的合作等进行了汇报, 并对共享实验室的科研设备、科研平台等情况进行了说明。

人才培养方面: 2014 年, 实验室常务副主任李晓维研究员入选 2014 年“科技北京”百名领军人才培养工程。范东睿研究员获得中科院卓越青年科学称号。韩银和、谭光明副研究员晋升为研究员。胡杏、刘磊、刘少礼、叶靖荣获院长优秀奖。

研究平台建设方面: 2014 年, 实验室建成验收了三大基础科研平台和研发了四大自研平台, 这些设备构建成实验室公共计算平台。各设备全

年不间断工作，为本学科多个研究方向的项目研究提供实验验证、仿真和测试环境。继 2013 年 12 月 24 日第二平台建成验收后，计算机体系结构国家重点实验室于 2014 年 7 月 11 日召开了第一平台和第三平台——端云软硬件千核平台和新型微体系结构设计平台的建成验收会。在院拨国重经费的支持下，三大基础科研平台建成，运行良好。此外，四个特色自研平台进展良好。由鄢贵海副研究员负责的“高效能可重构大数据处理加速引擎研究平台”、由陈云霁研究员负责的“机

器学习加速器的结构和编程研究平台”、由陈世敏研究员负责的“面向测量信息的日志记录分析平台”、由张云泉研究员负责的“众核处理器并行编程方法研究平台”，均按照计划进展良好。

国重建设方面：实验室主任：孙凝晖研究员（兼）；副主任：李晓维研究员（常务）、冯晓兵研究员；学术委员会主任：李国杰院士。实验室包括十七个研究组（如图 1），部署在五大研究方向——高端计算体系结构、微体系结构、编译和编程、VLSI 与容错计算、非传统体系结构。



图 3 国重学术委员会会议合影



## 科研基地进展

# 移动计算与新型终端北京市重点实验室

移动计算与新型终端北京市重点实验室依托于中国科学院计算技术研究所组建，2011年5月由北京市科委正式认定。2014年，实验室在依托单位中国科学院计算技术研究所的支持下，在实验室全体人员的共同努力下，在北京市科委组织的绩效考评中获得一致好评。2015年2月2日第二届学术委员会会议上，任命李锦涛研究员继续担任实验室主任，李国杰院士继续担任实验室学术委员会主任，并认命了第二届学术委员会委员。

随着移动互联网和宽带移动通信的发展，移动化的社交游戏、手机网络浏览、智能手机支付、手机增强三维现实显示、物联网及云端计算等成为了信息技术应用发展的大趋势，而移动计算及具备超强数据处理能力、多样化操作模式的一体化新型终端则是实现上述应用的关键。2014年，实验室继续以移动计算与新型终端发展方向的引领性关键基础理论与及其技术实现方法为核心开展研究，积极进行成果转化，切实为移动计算与终端产业作出贡献。

## 科研项目方面：

2014年承担了包括国家科技重大专项、国家自然科学基金、北京市重大科技项目、北京市自然科学基金、科技部国际合作项目及横向课题在内的40余项课题的研究工作。其中新增项目包括：国家科技重大专项2项、国家863项目2项、国家自然科学基金2项、北京市自然科学基金2项等，为移动计算和新型终端领域的发展持续提供创新方法和关键技术，解决本领域急待解决的重大关键问题，本年度到账经费超过一千万元。

## 科研成果方面：

(1) 实验室完成了完全自主知识产权的国内首颗自主4G小型基站基带芯片的研制，按照中国移动相关测试规范要求进行了内场测试，并将在工信部电信研究院组织的内外场测试基础上开展公网以及行业专网的试用推广。同时，成功研制了低成本、沉浸式、易操作、高保真的基于人机物三元融合端计算技术的远程无障碍交流平台-远程维权服务平台，该平台在人一人交互、人一机交互和机一物协同等方面做了一系列的理论创新和技术突破。还研发了跌倒检测的原型样

机，并开发了与之配套的移动终端应用程序，并对设备的性能进行了初步测试。此外，实验室研发的面向智力障碍者和老年人的智能可穿戴设备“防走失定位卡”获2014年第四届北京市残疾人辅助器具“金点子”创意发明活动三等奖，并和导盲避障系统一起在2014年11月27-29日中国国际福祉博览会展出。

(2) 在虚拟3D展示关键技术、原型系统等方面，突破了基于立体图相对和深度信息融合的高精度深度图构建技术，自主设计并搭建了基于多相机的图像采集系统，还开发了基于HTML5和Flash的现实终端软件以及自助式展示终端，可以实现在智能手机、智能电视以及查询机等设备上的三维产品展示应用。并在智能感知涉及的目标检测方面，针对行人目标检测展开研究，提出了基于多特征融合和级联分类器的行人检测方法。在视觉目标检测技术、低功耗高精度室内外场景识别软件、基于深度学习和半监督学习的室内定位方法上都做了研究，室内定位系统还被应邀指定为2014第四届中国(宁波)智慧城市技术与应用产品博览会官方场馆导航系统，获得了与会领导和使用者的好评。

(3) 本年度在超级基站原型机研制方面取得了实质突破，经过软硬件开发及联合测试，目前系统已实现CS通话、PS业务下载功能，且语音质量清晰，业务下载流畅，初步通过了功能性验证，并在专网系统中得到了应用。同时，对于超级基站实训平台进行了需求升级、功能及软硬件架构设计升级。基于超级基站架构，针对潮汐效应带来的负载差异性和负载动态变化，以及资源在虚拟映射时的资源利用率等问题上进行了研究，不仅能够通过优化算法得到全局最优部署方案，更通过对业务的预测和休眠模型的建立，降

低了休眠检测与决策的系统开销，减少了频繁切换带来的问题，提高了整体系统的稳定性。

学术进展方面：实验室先后在相关领域的国际权威学术期刊如IEEE Trans. on Communication、IEEE Trans. on Wireless Communication等以及一系列国际会议上发表论文100余篇，相关研究工作和成果已获得国际同行的认可。经过2014年在新型终端、智能业务、移动计算等方面的研究和积累，本年度共新发表论文30篇；新申请及新授权专利、软件著作权共计35项。

#### 队伍建设及人才培养方面：

实验室始终把队伍建设 and 人才培养作为实验室建设的首要任务，在关键方向上有目的、有重点地培养和引进人才。为了能够更好的稳定现有人才，实验室推行了“经费有保障、指导有专家，管理有制度”的体系完备的创新人才培养新路子。

本年度，在员工培养方面，实验室提升副研究员3名，高级工程师1名，在各个科研方向和岗位上提拔了新一代骨干力量。在学生培养方面，实验室共培养出博士研究生3名，硕士研究生19名，成为国家的栋梁之才。

人才引进方面通过特聘教授的形式，已聘请一批国际知名专家和海外优秀人才。本年度实验室申请到中国科学院与发展中国家科学院院长奖学金计划，支持尼泊尔学生Sumarga和巴基斯坦学生AHMAD FAIZAN在实验室从事博士研究，为发展中国家人才培养做出努力。

#### 国际合作方面：

移动计算与新型终端北京市重点实验室一直以“创造高端国际交流机会，带动促进北京市兄



弟单位共同对外发展”为理念，努力提升实验室以及北京市在该领域的国际影响力与国际化接轨。

2014年7月17日-18日，移动计算与新型终端北京市重点实验室为了促进通信和计算机领域国内外科研人员的交流合作，在北京市自然科学基金委和北京市国际合作基地的支持下，依托中国科学院计算技术研究所，学术委员会主任李国杰和委员李幼平两位院士联合发起了《通信与计算机科学技术融合》国际学术研讨年会。专家们探讨在通信与计算机科学技术的交叉融合的背

景下，通信与计算机之间的相互借鉴、相互促进和融合的思路和技术路线，并就通信与计算结合的科研方向、产业发展提出指导性意见。

除此之外，实验室还主办了 GameNets 国际会议和第四届全国可穿戴计算学术会议，邀请多位国际知名专家学者来访进行学术交流。促进实验室人员和北京市通信业界吸收国际上新颖的思维方式、研究成果、研究方法，全面提升研发水平，提高实验室国际影响力，推广具有自主知识产权的创新思想和技术，与世界一流水平接轨。



第一届《通信与计算机科学技术融合》国际学术研讨年会合影

## 科研基地进展

# 中科院智能信息处理重点实验室

2014 年度，智能信息处理重点实验室继续依托计算所在体系结构和网络研究方面的优势，以大规模多媒体数据的处理和利用为核心，在知识表示与推理、模式分类与学习等方面开展深入的研究，产出一流成果，培养基础人才，培育前瞻思想。实验室将研究方向凝练为“智能计算”、

“视觉计算”和“生物信息学”三个大的学科方向，并从信息的获取、表示、理解与转换等几个方面进行布局，进一步明确了包括自然语言处理、知识网格、大数据挖掘与机器学习、智能科学、视觉信息处理与学习、智能人机接口、大规模知识处理及智能应用、生物信息学在内的研究方向，构成了一个互为依托的有机整体。以期能够更好地满足国家在智能信息处理方向的重大需求、服务国民经济建设和建设和谐社会。

在基础研究方面，本年度实验室继续保持了良好的发展势头，完成了多项创新性研究成果，相关成果累计发表论文 100 余篇，其中 IJCV、IEEE Trans. on IP/ KDE/CSVT 等国内外刊物论文 30 余篇，CVPR、ECCV、NIPS、AAAI、ICDE 等一流国际会议论文 50 余篇。新申请专利十余项。

在“智能计算”领域，继续坚持机器翻译、智能科学、知识网格和大数据挖掘等学科方向。自然语言处理研究组提出了基于分析 - 生成模式的统计机器翻译模型，借助句法信息的翻译质量评估方法，基于句法信息的结构化语言模型，以及文档级术语翻译建模等。在机器翻译方面发表领域顶级会议论文 6 篇（其中 4 篇 COLING，1 篇 EMNLP，1 篇 AAAI 论文）。语言分析方面，研究组在自然语言标注标准的自动适应问题上展开系统性研究，提出了高效的异种知识库融合与转换方法，在保持分析速度不变的条件下可以显著提升分析精度。相关论文被领域顶级刊物 Computational Linguistics 录用，刘群研究员和姜文斌副研究员针对该工作在领域顶会 COLING 2014 做大会特邀报告。在科技奖项方面，本年度刘群研究员获得了中国电子学会科学技术奖一等奖。在大规模数据挖掘和机器学习方面，针对互联网上复杂事件关联，且事件在不断地更新和涌现，设计了一种紧凑、高效的数据结构来存储事件场景，提出了一种新的在线频繁场景挖掘算法。该成果被 CCF A 类会议 ICDE 2015 录用为长文，



这也是首篇计算所为第一完成单位的 ICDE 研究类长文。结合主成分分析 (PCA) 和大间隔学习, 提出了一种基于后验正则化的贝叶斯方法, 被 CCF A 类会议 AAAI 2015 录用为长文。在机器学习算法以及大数据并行数据挖掘技术研究方面发表录用 26 篇论文, 其中 CCF A 类 2 篇, CCF B 8 篇, SCI 论文 7 篇。

在“视觉计算”领域, 坚持智能交互、视觉模式分析、多媒体等方向。人脸识别研究组继续围绕人脸检测、面部特征点定位、人脸识别三个方向开展研究工作, 提出了漏斗形的多姿态人脸检测框架、联合深度图像特征学习与混合统计量集合建模的方法等, 发表了 17 篇学术论文, 其中 CCF A 类论文 6 篇 (包括国际刊物 IJCV 和 IEEE T IP 论文各 1 篇, 国际会议 CVPR 论文 4 篇), 领域顶级会议 ECCV 论文 2 篇, 领域顶级会议 NIPS 论文 1 篇。此外, 利用深度学习技术在姿态、表情分析方面取得了较好的效果, 提出的融合深度特征与多核学习方法取得了 ACM ICMI 2014 国际会议上组织的 EmotiW 表情识别竞赛的冠军。在视觉编码方面, 提出了一种深度动态编码器 (deep dynencoders) 模型, 用来

建模视频序列, 可应用在不同的视频任务上, 包括动态纹理合成、视频分类和视频分割等。还提出了一种新的图像超分辨率方法——深度网络级联 (deep network cascade, DNC)。智能人机交互研究组在人机交互方向继续围绕手势、手语识别开展研究工作。课题组提出基于格拉斯曼协方差矩阵 (GCM) 的手语描述方法、层级 GCM (HGCM) 手语表示方法等。此外, 还构建并发布了标准中国手语词汇集, 联合微软亚洲研究院、北京联合大学共同成立了基于 Kinect 的手语识别工作组。

在科研项目承担与争取方面, 实验室本年度获批 973 课题 1 项 (负责人为陈熙霖研究员)、面上 / 青年基金项目多项。此外, 还持续得到了百度、高通、Intel 等产业公司的支持。2014 年底可用经费 3200 万元, 2014 年收入 1000 余万元。

在实验室管理方面, 实验室召开了学术委员会、学术规划会、室务会和学术年会, 特别是在公共经费使用方面, 形成了透明的制度, 通过为导师缴纳学生培养费、支持博士生学术交流、按学术论文进行经费拨付等方式, 尽可能为研究人员提供稳定的支撑。

## 科研基地进展

# 中科院网络数据科学与技术 重点实验室

2014 年度，中科院网络数据科学与技术重点实验室继续坚持已有定位：推动网络数据科学发展，突破 ZB 级网络数据感知、传输、存储、管理与分析体系架构，研究网络数据界的溯源、定位、预测与控制方法，支撑安全大数据、情报大数据、金融大数据、商业与媒体消费大数据等系列应用。在基础研究、大项目争取、学术交流人才引进和推动大数据研究与发展等方面都取得了进展。获得钱伟长中文信息处理科技奖一等奖和电子学会科技进步奖二等奖。

具体包括：在基础研究方面，围绕着网络数据研究，发表论文近 50 篇，专著 1 部。其中，发表在 PNAS 等 A 类期刊上的论文有 5 篇，发表在 WWW、SIGIR 等 A 类会议上的论文有 4 篇。同时我们组织了大数据分析的国际 workshop，撰写了一系列与大数据相关的前瞻性论文，完成了大数据学科的布局与推进。

在重大项目争取方面，获得国家自然科学杰出青年基金项目“web 信息检索与数据挖掘”。

国家自然科学基金项目 4 项，国家 973 计划子课题 2 项，国家 863 计划子课题 1 项，院先导专项 1 项。

在学术交流人才引进方面，“中科院网络数据科学与技术创新国际团队”启动试运行论证会顺利召开。经第十三次中国科学院人才工作领导小组会议审定，依托我所的“中科院网络数据科学与技术创新国际团队”顺利通过院立项评审，该创新国际团队共有成员 16 名，国内成员 8 名、海外成员 8 名。该团队将立足于国家网络空间战略需求与大数据产业发展瓶颈，研究相关关键技术，推动网络大数据科学与技术学科方向的大力发展，同时打造一支具有国际影响力的研究队伍，取得国际领先的基础研究成果，为产业界网络大数据的深度挖掘与社会化服务提供支持。

在推动大数据学科与产业发展方面，牵头组织了的国内规模最大、最具影响的大数据领域技术盛会——2014 中国大数据技术与学术大会 (Big



Data Technology Conference 2014, BDTC 2014)。大会主旨在于推进大数据科研、应用与产业的发展。本次大会以更加国际化的视野,帮助与会者了解海内外大数据技术的发展趋势;从技术与实践角度出发,探讨“大数据生态系统”、“大数据技术”、“大数据应用”、“大数据基础设施”等领域的新技术应用和实践经验。大会特邀近百位来自全球大数据产业界知名学者、企业领军人

物、行业资深专家、一线实践技术代表,内容涵盖 Hadoop、YARN、Spark、Tez、HBase、Kafka、OceanBase 等开源软件的最新进展, NoSQL/NewSQL、内存计算、流计算和图计算技术的发展趋势, OpenStack 生态系统对于大数据计算需求的思考,以及大数据下的可视化、机器学习 / 深度学习、商业智能、数据分析等的最新业界应用。



大数据大会现场

# 学术活动

近年来，计算所更加重视对外合作交流，邀请国内外专家学者来所做所级学术报告，积极参加和主办各类学术会议。2014年计算所科研人员参加各类国际学术会议并做报告85次，邀请国内外专家学者做学术报告120次，同时，还主承办了大型国内/国际学术会议8次，提升了计算所在相关学科领域的学术影响力。

## □ 学术报告列表

序号	报告人	报告题目	报告人单位	报告日期
1	Chia-Wen Lin	Video Retargeting: Algorithms, Applications, and Quality Assessment	清华大学（新竹）	2014.1.14
2	裴健	Social Network Analysis by Compression: Not Only Space Saving, But Also Insight Gaining	Simon Fraser University	2014.1.15
3	张晓东	数据表在大数据处理中分布存储的基本结构与应用	The Ohio State University at Columbus	2014.1.15
4	李响	神经网络语言模型	中科院计算所	2014.1.22
5	王明轩	深度学习在NLP中实践	中科院计算所	2014.1.26
6	马坚伟	讨论压缩感知在遥感成像和地震勘探中应用	哈尔滨工业大学	2014.2.18
7	陈宏申	Learning Continuous Phrase Representations for Translation Modeling	中科院计算所	2014.2.28
8	孟凡东	RNN Encoder-Decoder for MT	中科院计算所	2014.3.15
9	Sandip Kundu	Adaptive Thread to Core Assignment via Online Program Phase Classification in Asymmetric Multicore Processors(AMP)	University of Massachusetts Amherst	2014.3.19



序号	报告人	报告题目	报告人单位	报告日期
10	郭斯杰	Book Keeper: Durability as a Service	Twitter	2014.3.19
11	Alex Hauptmann	Care Media: Visual Health Monitoring of Behavior	Carnegie Mellon University	2014.3.21
12	赵世奇	研究生成长报告	百度公司	2014.3.25
13	刘江鸣	联合 NE 与依存句法分析	北京交通大学	2014.3.27
14	Sandip Kundu	Improving Yield and Reliability of Chip Multi-processors	University of Massachusetts Amherst	2014.4.9
15	张晓东	用新的计算模式创建大数据处理系统	The Ohio State University at Columbus	2014.4.14
16	Peerachet	Generating Sequences with Convolutional Neural Network	泰国科技部国家电子和计算机技术中心 (NECTEC)	2014.4.15
17	马晓松	Understanding and Adapting to Heterogeneous Concurrent I/O Workloads on Shared Storage	North Carolina State University	2014.4.16
18	张文	实验探索从平行语料中抽取子句切分, 改进机器学习	中科院计算所	2014.4.22
19	Alexey Nesvizhskii	Combined Transcriptome and Proteome Analysis: Methods and Applications	University of Michigan	2014.4.23
20	Alexey Nesvizhskii	Reconstruction of Protein Interaction Networks Using Affinity Purification Mass Spectrometry Technology	University of Michigan	2014.4.23
21	马晓松	Understanding and Adapting to Heterogeneous Concurrent I/O Workloads on Shared Storage	Qatar Computing Research Institute	2014.4.26
22	马世龙	面向航空任务电子系统的软件自动化测试与检测研究	北京航空航天大学	2014.4.30
23	于剑	贵类公理	北京交通大学	2014.5.6
24	Tarik Taleb	Mobile Networks as a Cloud Service	NEC Europe Ltd	2014.5.6
25	Xiao-Ming Fu	Microblogging Services with Divergent Traffic Demands	Georg-August-University of Goettingen	2014.5.8
26	Yu-Xiang Chen	3D Image Analysis for in Situ Structural Biology using Cryo-Electron Tomography	Max Planck Institute of Biochemistry	2014.5.9

序号	报告人	报告题目	报告人单位	报告日期
27	张岳	基于 Shift-Reduce 的依存分析	Singapore University of Technology and Design	2014.5.11
28	张金超	Tagging the Web: Building a Robust Web Tagger with Neural Network	中科院计算所	2014.5.13
29	Stevens Lumetta	The CompGen Platform	University of California,Berkeley	2014.5.21
30	邢波	大规模机器学习系统的研究和应用	Carnegie Mellon University	2014.5.21
31	Frank-Z. Wang	How Will Computers Evolve Over the Next 10 Years	University of Kent	2014.5.23
32	Xiao-Hui Bei	Balancing Efficiency and Fairness in Resource Allocation	Nanyang Technological University	2014.5.30
33	Stephen Hanly	5G 前沿通信技术	Macquarie University	2014.5.30
34	Guo-Liang Wang	Log-concavity of Combinations of Sequences and Applications to Genus Distributions	University of Haifa	2014.6.4
35	Shao-Lei Ren	How Can Multi-Tenant Data Center Become Sustainable	Florida International University	2014.6.10
36	Min-Ming Li	Algorithmic Mechanism Design on Cloud Computing and Facility Location	香港城市大学	2014.6.13
37	Yi Li	Convolutional Neural Networks for Two Challenging Problems in Computer Vision: Image Quality Assessment and Visual Tracking	Australia's ICT Research Centre of Excellence(NICTA )	2014.6.13
38	于浩	Scale-out Energy-efficient Big-data Server Thousand Cores In-Memory With Reconfigurable I/Os	Nanyang Technological University	2014.6.16
39	蔡洽吾	拼写校正分析	中科院计算所	2014.6.17
40	Ming-C. Lin	Virtual Urbanscape: Recent Advances, Challenges & Opportunities	University of North Carolina	2014.6.17
41	邢晶	高性能计算和存储相关的关键技术	中科院计算所	2014.6.17
42	Ji-Hong Kim	Lifetime Improvement of NAND Flash-based Storage Systems	Seoul National University	2014.6.23



序号	报告人	报告题目	报告人单位	报告日期
43	Peng-Hui Yao	A Parallel Repetition Theorem for Entangled Two-player One-round Games under Product Distributions	National University of Singapore	2014.6.25
44	Zhi-Gang Deng	Computer Graphics and Animation Research: Recent Advances at UH	University of Houston	2014.6.26
45	黄广斌	超限学习机——迈向人脑学习机制的一小步以及大数据分析的基础	Nanyang Technological University	2014.6.27
46	Yuan-Yuan Zhou	Configuration Management in Data Centers	University of California, San Diego	2014.6.30
47	尤海航	关于高性能计算机的性能评估方法	中科院计算所	2014.7.1
48	Sheldon X.-D. Tan	Physics-Based Full-Chip Electron-Migration Analysis and System-level Reliability Management	University of California, Riverside	2014.7.4
49	Shi-Yan Hu	Interconnect Synthesis for Emerging Nanometer VLSI Design	Michigan Technological University	2014.7.7
50	宋朝明	Understanding Complex Systems: Networks and Dynamics	University of Miami	2014.7.7
51	Hai Li	Design Challenges in MLC STT-RAM Caches	University of Pittsburgh	2014.7.8
52	Michael Hsiao	On Improving Diagnosability for LBIST	Virginia Polytechnic Institute and State University (Virginia Tech)	2014.7.10
53	Hong-Wei Zhang	Predictable Wireless Networking for Real-Time Sensing and Control: Connected and Automated Vehicles and Beyond	Wayne State University	2014.7.10
54	Yi-Ran Chen	Emerging Nonvolatile Memories: Designs and Applications in Conventional and Neuromorphic Computing	University of Central Florida	2014.7.11
55	梁云	关于 GPU 性能优化技术	北京大学	2014.7.16
56	Ming-Hsuan Yang	Learning to Parse Scenes for Image Understanding	University of California, Merced	2014.7.16

序号	报告人	报告题目	报告人单位	报告日期
57	Hua Gang	Probabilistic Elastic Part Model: A Pose-Invariant Representation for Face Recognition	Stevens Institute of Technology	2014.7.21
58	毛国强	Spectrum and Energy Efficient Technologies for Heterogeneous and Dense Networks	University of Technology, Sydney	2014.7.21
59	Hua-Cheng Yu	Finding Four-Node Subgraphs in Triangle Time	Stanford University	2014.7.23
60	Gang Hua	Probabilistic Elastic Part Model: A Pose-Invariant Representation for Face Recognition	Stevens Institute of Technology	2014.7.23
61	陈锐志	Contextual Thinking: Transforming Mobile Phones from Smart to Cognitive	Texas A&M University Corpus Christi	2014.8.11
62	Shui-Wang Ji	Machine Learning in Vision and Brain Analytics	Old Dominion University	2014.8.12
63	Babak Falsafi	Big Data and Dark Silicon: Taming Two IT Inflection Points on a Collision Course	Swiss federal Institute of Technology in Lausanne (EPFL)	2014.8.14
64	John Sampson	Scaling Diversity Alongside Dark Silicon with Coprocessor-Dominated Architectures	Pennsylvania State University	2014.8.18
65	Jason Mars, Ling-Jia Tang	Improving Efficiency and Utilization of Datacenters with Whare-Map and Bubble-Flux	University of Virginia, University of Michigan	2014.8.19
66	徐小文	超大规模数值模拟应用对传统数值算法的新要求	北京应用物理与计算数学研究所	2014.8.20
67	宋风龙	DPU 处理器结构设计	中科院计算所	2014.9.3
68	Uri Zwick	The Random-Facet Pivoting Rule	Tel Aviv University	2014.9.4
69	Jose Luis Vazquez-Poletti	Cloud Computing: Where Infrastructure Adapts to the Application	Complutense University of Madrid	2014.9.12
70	崔鹏	Social-Sensed Multimedia Computing	清华大学	2014.9.17
71	David Woodruff	Lower Bounds for Data Streams - A Survey and Open Questions	IBM Almaden Research Center	2014.9.18



序号	报告人	报告题目	报告人单位	报告日期
72	Geng-Fa Fang	Licensed Shared Access: Evolution Towards Dynamic Spectrum Sharing in 5G Networks、Radio Environment Maps for Licensed Shared Access networks	Macquarie University	2014.9.19
73	刘洋	命名实体识别经典工作	中科院计算所	2014.9.20
74	James. C. Hoe	CoRAM: An FPGA Architecture for Computing	Carnegie Mellon University	2014.9.24
75	马青松	基于整数线性规划的依存分析	中科院计算所	2014.9.27
76	刘金刚	相对标准的基本原理	首都师范大学	2014.10.9
77	杨似彤	译后编辑经典工作综述	中科院计算所	2014.10.13
78	Larry Davis	Video Understanding	University of Maryland	2014.10.13
79	Yang Wang	Separating Data from Metadata for Robustness and Scalability	The Ohio State University at Columbus	2014.10.21
80	Yi-Er Jin	Computer System Protection through Hardware-Software Collaboration	University of Central Florida	2014.10.22
81	麻晓娟	Towards Crowd-sourced Semantic-based Multimodal User Interfaces	华为技术有限公司	2014.11.4
82	李灵军	Developing Mass Spectrometry-based Molecular Imaging and Proteomics Strategies for the Studies of Neurological Diseases	University of Wisconsin-Madison	2014.11.12
83	葛瑛	Top-down Proteomics in Health and Disease: Challenges and Opportunities	University of Wisconsin-Madison	2014.11.12
84	关慎恒	Characterization of Large Intact Protein Complexes by Native Mass Spectrometry	The University of California, San Francisco	2014.11.12
85	刘晓文	De Novo Protein Sequencing by Combining Top-down and Bottom-up Tandem Mass Spectra	Indiana University-Purdue University Indianapolis (IUPUI)	2014.11.12
86	黄超兰	Comparison of Ultradefinition (UD) MSE and WiSIM-DIA Techniques in the Study of Middle-down Proteomics	上海蛋白质中心	2014.11.12
87	谭敏佳	Lysine Glutarylation is a Protein Posttranslational Modification Regulated by SIRT5	中科院上海药物所	2014.11.12

序号	报告人	报告题目	报告人单位	报告日期
88	应万涛	Site-specific Glycoform Assignment through Glycoproteomics Approaches	北京蛋白质组研究中心 (BPRC)	2014.11.12
89	刘铭琪	Deciphering Glycoform of Site-specific Glycosylation based on High-throughput HCD/CID-MS/MS and MS3	复旦大学	2014.11.12
90	叶明亮	Glycoproteomics Analysis and Data Processing	中科院大连化学物理研究所	2014.11.12
91	汤海旭	Characterization of Glycan Microheterogeneity in Glycoproteins Using Liquid Chromatography Coupled Tandem Mass Spectrometry	Indiana University-Purdue University Indianapolis (IUPUI)	2014.11.12
92	林诚	De Novo Glycan Sequencing by Tandem Mass Spectrometry	Boston University	2014.11.12
93	黄蔚	Functional Glycomics: Unveiling the Role of Protein Glycosylation	中科院上海药物所	2014.11.12
94	高友鹤	Informatics from a Biologist's Point of View	北京协和医院	2014.11.12
95	陆豪杰	Advancements in Tandem-MS based Proteome Quantification	复旦大学	2014.11.13
96	刘小云	Integrated Proteomic and Metabolomic Analyses of Intracellular <i>Salmonella Typhimurium</i> Reveal Severe Iron Starvation During Infection of Epithelial Cells	北京大学	2014.11.13
97	丁琛	Towards Building an Ideal Proteomics Platform	北京蛋白质组研究中心 (BPRC)	2014.11.13
98	单亦初	Isobaric Labeling Assisted Proteome Identification and Quantification Based on Database Search and Denovo Sequencing	中科院大连化学物理研究所	2014.11.13
99	汪迎春	Systematically Ranking the Tightness of Membrane Association for Peripheral Membrane Proteins	中科院遗传发育所	2014.11.13
100	钟鸿英	Mass Spectrometric Quantification of Interactions between Proteins and Drugs	华中师范大学	2014.11.13
101	迟浩	Fast and Comprehensive Analysis of High Resolution MS/MS Data	中科院计算所	2014.11.13



序号	报告人	报告题目	报告人单位	报告日期
102	王春光	Preliminary Application of Mass Spectrometry on the Study of Conotoxins	同济大学	2014.11.13
103	周虎	Technology Development for Membrane Protein Analysis and its Applications	中科院上海药物所	2014.11.13
104	薛宇	The Reconfiguration of the Human Kinase-Substrate Phosphorylation Network by Genetic Polymorphisms Determines the Cancer Susceptibility	华中科技大学	2014.11.13
105	王恒樑	The Influence of Temperature on the Virulence of Shigella Flexneri	军事医学科学院	2014.11.13
106	万晓华	Iterative Methods in Large Field Electron Microscope Tomography	中科院计算所	2014.11.13
107	张红雨	Exploring the Origins of Proteins by ATP Selection in a Random Peptide Library Consisting of Reduced Amino Acids	华中农业大学	2014.11.13
108	Christophe Dubach	关于异构系统上性能可移植性的报告	University of Edinburgh	2014.11.18
109	Zhao-Hui Wei	On the Power of Linear Programming and Semidefinite Programming over Polytopes	Nanyang Technological University	2014.11.20
110	贾韬	Collective Behavior in the Evolution of Scientific Research Interests	Rensselaer Polytechnic Institute	2014.11.25
111	翟成祥	Towards a Game-Theoretic Framework for Information Retrieval	University of Illinois Urbana, Champaign	2014.11.25
112	向小佳	LockFree 数据结构动态分析工具 & 支持 TransactionMemory 的 CPU Simulator(X86) 的设计与实现	中科院计算所	2014.12.1
113	Xiao-Yu Wang	Object Detection with Regionlets	NEC Laboratories America	2014.12.2
114	张怡颖	From Design Principles and Disruptive Technologies to Real Systems: Making Computers Remember Faster and More Reliably	University of California, San Diego	2014.12.15
115	Si-Wei Lyu	Multi-View Multi-Target Tracking and 3D Realtime Visualization of Tracking Results	University at Albany, State University of New York	2014.12.15

序号	报告人	报告题目	报告人单位	报告日期
116	Calton Pu	An Experimental Study of Latency Long Tail Problem: Impact of Very Short Bottlenecks in Cloud Environments	Georgia Institute of Technology	2014.12.19
117	杨双红	Large Scale Topic Modeling on Twitter	Twitter	2014.12.19
118	张峰	A System Researcher's Perspective on Deep Learning	New York University	2014.12.24
119	Shui-Cheng Yan	NUS-Purine: A Bi-graph based Deep Learning Framework	National University of Singapore	2014.12.26
120	原进宏	Cloud RAN Based on Network Coding Approach	The University of New South Wales	2014.12.26

人 人 是 才  
人 人 幸 福



## □ 重要学术会议

### 第 5 届亚太计算机系统国际研讨会

2014 年 6 月 25 – 26 日, 计算机体系结构国家重点实验室组织承办了第 5 届亚太计算机系统国际研讨会 ACM APSys-2014 (<http://acs.ict.ac.cn/apsys2014/>)。包云岗副研究员出任大会共同主席。本届 APSys 大约有 100 位与会人员, 来自 MIT、Yale、EPFL、清华大学、北京大学、上海交通大学等国内外研究机构。本届 APSys 与 ICML 联合召开, 吸引了更多对计算机系统、机器学习、大数据分析等感兴趣的人员参加。

### 《通信与计算机科学技术融合》国际学术研讨年会

随着以云化大数据计算、传输、存储与个性化应用服务为核心特点的移动互联网应用的爆炸式增长, 计算技术、宽带通信技术和智能服务技术呈现出了显著的相互融合, 相互促进的趋势。预计未来十年内, 通信与计算服务需求还将增长 1000 倍, 通信与计算技术的融合是最有可能支撑未来服务需求的必由之路。

为了促进通信和计算机领域科研人员的交流合作, 在李国杰和李幼平两位院士倡导下, 中国科学院计算技术研究所牵头发起组织《通信与计算机科学技术融合》国际学术研讨年会。每年定期邀请国内外通信与计算机领域的知名专家学者, 一起研讨通信与计算机科学技术的交叉融合, 通过通信思维与计算机思维的碰撞探索移动通信技术新的发展途径。

2014 年 7 月 17-18 日, 第一届《通信与计算机科学技术融合》国际学术研讨年会在北京香山饭店隆重召开。本次研讨会以绿色通信为主题,

旨在开放式地探讨在大数据时代, 未来无线通信网络如何融合高通量计算等前沿技术, 大幅降低系统能耗, 走出一条发展新路。邬贺铨、李国杰、李幼平、陈俊亮等 8 位资深院士以及清华大学王京教授、中国电信韦乐平总工程师、北京邮电大学张平教授、北京交通大学张宏科教授、英国肯特大学王江舟教授等 20 多位国内外通信与计算机领域知名专家学者参加了本次会议。本次研讨会得到了北京市科委的大力支持。



会议由中国科学院计算技术研究所无线中心主任石晶林研究员主持。大会主席李国杰院士和中国科学院计算技术研究所党委书记兼副院长李锦涛分别致辞。随后, 石晶林研究员以“大数据时代绿色的通信与信息服务的思考”的报告拉开了本次研讨会的序幕。陈俊亮院士阐述了以内容为中心的未来网络架构及其面临的机遇和挑战; 李幼平院士提出了基于超级基站架构的天播地存服务模式, 充分利用卫星的广播特性, 利用存储帮助路由, 边缘帮助核心, 可极大地减少核心网压力, 并为用户提供个性化服务; 中国电信韦乐平专家的报告首先分析比较了 CT 和 IT/Internet 两大行业特点, 总结了电信业在长期学习和应用计算机技术成果中的根本问题, 说明电信业实施“去电信化”的必要性, 最后指出 SDN 是“去

电信化”的重要突破方向及其主要挑战；北京邮电大学张平教授提出了构建信息生态系统的概念，认为不同的业务具有不同的生态，在不同的阶段和高度上，其特性不同，把握好通信、计算、存储三者之间的关系需要充分考虑不同的生态。北京交通大学张宏科教授阐述了当前网络的不足，提出了智能协同网络模型。来自日本东北大学的加藤教授和英国肯特大学的王江舟教授分别报告了D2D技术进展以及异构网络部署面临的问题和解决方案；悉尼科技大学的毛国强教授从复杂系统的角度，提出解决通信问题本身不应只在通信领域，要针对特定的领域，站在不同的高度，分析业务对于通信能力的需求，从而从根本上解决通信难题；来自新加坡科研局部A\*Star的孙素梅研究员和来自澳大利亚联邦科学与工业组织的黄小晶研究员分别报告了大规模天线技术的和E波段链路回传技术面临的主要问题和科研、产业界的进展。



在本次研讨会上专家们畅所欲言，进行了激烈的讨论和思想碰撞，探讨了在通信与计算机科学技术的交叉融合的背景下，通信与计算机之间的相互借鉴，相互促进和融合的思路和技术路线，并对通信与计算结合的科研方向、产业发展提出了指导性意见。

## 第4届面向大数据的体系结构与系统国际研讨会

2014年9月11日至12日，计算机体系结构国家重点实验室组织承办了第4届面向大数据的体系结构与系统国际研讨会ASBD 2014 (<http://acs.ict.ac.cn/asbd2014/>)。张立新研究员担任大会主席，包云岗副研究员出任大会共同主席。本届ASBD大约有100位与会人员。华为公司在ASBD研讨会上发布了计算所-华为联合实验室的最新研究成果“第三代数据中心”白皮书，取得很大反响。

## 第23届IEEE亚洲测试学术会议/第15届IEEE寄存器传输级及高层国际研讨会

2014年11月16日-20日，计算机体系结构国家重点实验室在杭州成功组织召开了第23届IEEE亚洲测试学术会议(IEEE 23rd Asian Test Symposium, 简称ATS'14)及第15届IEEE寄存器传输级及高层国际研讨会(简称WRTL'14)。本次会议由IEEE Computer Society, Test Technology Technical Council和中国科学院计算技术研究所(简称计算所)共同主办。实验室李华伟研究员担任ATS'14大会主席，韩银和研究员担任ATS'14程序主席，张磊副研究员担任WRTL'14程序主席。本次ATS'14和WRTL'14会议共吸引了140余位境内外专家学者参加，其中境外代表80余位。全球领先的电子设计自动化(EDA)企业(Mentor、Synopsys、Cadence等)、测试设备供应商(Advantest、Avago等)和集成电路设计公司(Intel、AMD、威盛、华为等)等都派有技术研究人员和工程师参加会议。这两个国际会议的圆满召开，为测试领域的国内外学术界和企业界交流讨论



提供了有效的平台，展现了中国学者在集成电路测试领域的国际影响力，对于促进世界尤其是亚洲地区电子测试技术的研究和发展有重要意义。

### 第 20 届全国信息存储技术学术会议 NCIS2014

先进计算机系统研究中心组织并承办了第 20 届全国信息存储技术学术会议 (NCIS2014) 暨 National Conference of Information Storage 2014 (会议编号: CCF-TC-14 - 26N)。NCIS2014 由中国计算机学会主办，中国计算机学会信息存储技术专业委员会及中国科学院计算技术研究所承办，清华大学及北京大学协办。会议于 2014 年 9 月 11-12 日在北京召开，由国防科技大学方粮教授和中国科学院计算技术研究所孙凝晖研究员担任大会主席，清华大学舒继武教授和北京大学代亚非教授担任程序委员会主席。CCF 专委工委常务委员、太极计算机股份有限公司总裁刘淮松先生代表中国计算机学会出席会议，并在开幕式上致辞。

本届会议邀请了美国俄亥俄州立大学张晓东教授、美国匹兹堡大学陈怡然博士、中国科学院计算技术研究所陈世敏研究员及张立新研究员、香港中文大学李柏晴博士 5 位知名专家学者做大



会报告。会议还举办了青年学者论坛，邀请中科院计算技术研究所高级工程师王磊、华中科技大学副教授华宇、百度高级架构师欧阳剑、清华大学副教授张广艳、北京大学特聘研究员孙广宇、微软亚洲研究院洪春涛博士 6 位知名青年学者介绍各自的科研工作，以及他们对如何做出出色研究工作的经验和心得。会议还举办了博士生论坛，邀请清华大学陆游游、华中科技大学付恣和北京航天航空大学康俊彬三位在读博士生介绍了各自在知名国际会议上发表的相关科研工作。此外，会议还邀请了曙光、华为、百度、Intel 和 Memblaze 等企业界的同仁介绍各自的存储产品技术。

本次大会共收到投稿 70 篇，录用 42 篇，录



取率为 60%。录用的论文将以论文集的形式发表到《计算机研究与发展》增刊。大会从录用论文中评选出的 5 篇优秀论文，它们将被推荐到《计算机研究与发展》正刊。大会共举办两天，共有 132 人参加会议，其中注册参会 97 人，厂商代表 16 人。

### 第四届全国可穿戴计算学术会议暨首届“中国（国际）智能可穿戴技术与产业论坛”

2014 年 7 月 26 日 -28 日，普适计算研究中心在成都组织召开了第四届全国可穿戴计算学术会议暨首届“中国（国际）智能可穿戴技术与产业论坛”，会议邀请国内外可穿戴计算相关领域的 10 名著名学者作了大会报告，26 名专家作了专题报告，15 位研究学者宣读了专题论文，涉及国内外可穿戴计算技术、热点产品的现状与潮流方向，相关专利分析以及可穿戴计算在健身、健康、娱乐、工业和其它领域的典型应用等。会议开展了“可穿戴与大数据”、“面向智能可穿戴的材料”、“智能可穿戴终端、身体传感网络”、“交互、设计、用户体验、多方向交叉”、“可穿戴计算发展战略研讨”等专题讨论，还包括了企业展示、创客产品交流、Poster 交流以及可穿戴服装展示等精彩环节。

大会演讲的海内外著名专家包括德国科学与工程院院士及执行理事会理事、德国不莱梅大学 Otthein Herzog 教授，德国人工智能研究中心 Paul Lukowicz 教授，加拿大工程院院士、加拿大西安大略大学姜晶教授，国际医学与生物工程院院士、美国医学与生物工程学院院士、香港中文大学张元亭教授（IEEE fellow）、原南洋理工大学生物工程系系主任、美国医学与生物工程学院院士、英国皇家化学学会会士、西南大学李长明教授，中科院软件所戴国忠教授，哈尔滨工业大学杨孝宗教授，还有来自香港中文大学、中科院计算所、中科院软件所、中科院自动化所、中科院深圳先进技术研究院、中科院苏州纳米技术与纳米仿生研究所、清华大学、北京大学、浙江大学、兰州大学、西南大学，东南大学，北京科技大学，重庆大学、电子科技大学等科研学术机构的教授、千人计划学者以及 973 首席科学家等。此外，还有中国社会福利协会兼养老服务信息化建设工作委员会、民政部康复中心、国家体育总局、国家知识产权局、江苏省物联网研究发展中心、工信部通信研究院等单位的专家，和 Intel、华为、科大讯飞、深圳中兴物联网科技、JXJ、云合网等企业领导人，共计约 500 多人参加了会议。





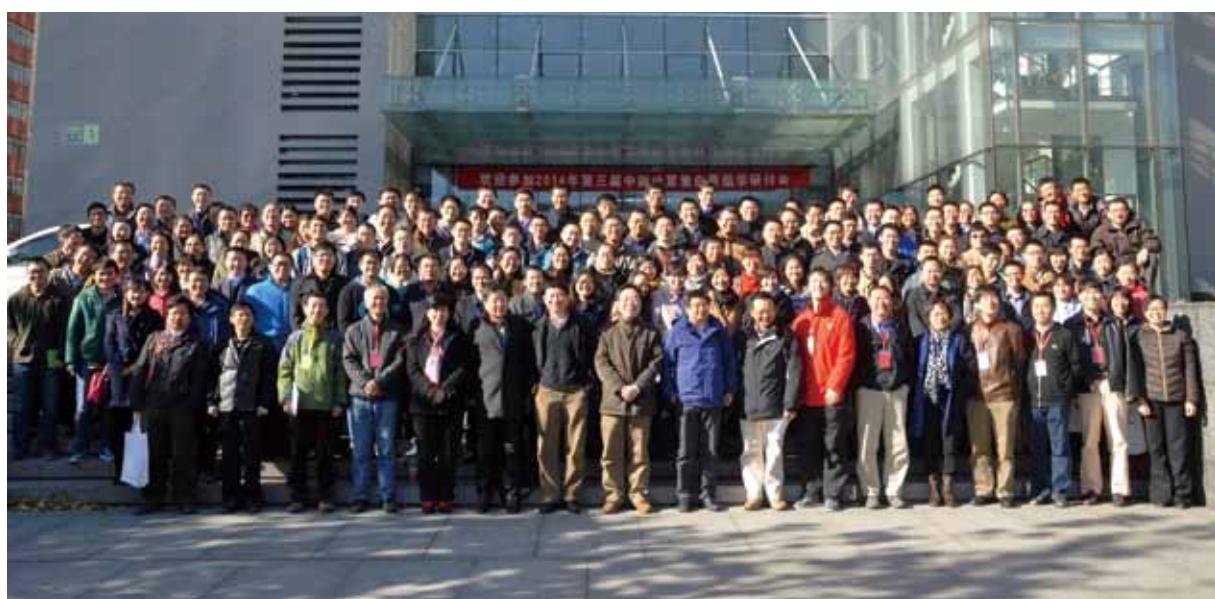
### 第三届中国计算蛋白质组学研讨会

2014年11月12-13日，第三届中国计算蛋白质组学研讨会(CNPC-2014, China Workshop on Computational Proteomics)在计算所盛大召开。分析测试及科学仪器行业全球访问量第一的网站Antpedia会后报道中提到了本届会议“可称得上是蛋白质组学界的APEC”(<http://www.antpedia.com/news/99/n-476499.html>)。这已是计算所的pFind团队发起并组织筹办的第三届会议。

CNPC会议的主题是“计算蛋白质组学”，旨在推动计算技术在中国的蛋白质组研究中发挥出更加切实的作用，更好地促进国内的学术交流。CNCP的办会方针是“学术优先、其他从简”，办会方式是不征文、不收费，只设邀请报告、只设一个进程，学术平等。CNCP的报告人由资深学者和往届报告人共同推荐产生，新人优先。CNCP会议的特色是办成“小人物”的学术盛会，所谓“小人物”是指会议邀请的报告人暂时还是

一个年轻的“小人物”，比如一个刚毕业的博士、在站或刚出站的博士后、刚回国的年轻PI等等，但在较短的时间内已经取得了高水平的研究成果，比如发表了高水平的论文，未来的发展具有成为“大人物”的高潜力；另外，报告的内容主体上都是最近一到两年内由报告人亲自完成的工作，包括未发表的最新内容。CNCP会议通过技术手段做到了严格控制报告时间，每个报告30分钟，实际效果得到了代表们的赞誉。

CNPC-2014会议邀请到了26位来自海内外的一线专家学者与会报告，来自北京、上海、辽宁、天津、美国、香港等地的学者180余人出席了此次盛会。本届会议内容丰富、形式多样，全面的展示并交流了中国的计算蛋白质组学发展成果。根据研究内容，11月12日和13日两天安排了四组共26个报告，这四组分别为整体蛋白质的鉴定计算、蛋白质组学的应用、定量蛋白质组学和糖蛋白质组学。在这26个报告中，有50%（13个报告）内容还没有正式发表，代表了领域内最新的研究成果。





## 2014 中国大数据技术与学术大会

2014 中国大数据技术大会 (Big Data Technology Conference 2014, BDTC 2014) 暨第二届 CCF 大数据学术会议于 12 月 12-14 日在北京新云南皇冠假日酒店顺利召开。本次大会由中国计算机学会 (CCF) 主办, CCF 大数据专家委员会承办, 由中科院计算所网络数据科学与技术重点实验室与 CSDN 具体组织, 主旨在于推进大数据科研、应用与产业的发展。

本次大会历时三天, 以更加国际化的视野, 帮助与会者了解海内外大数据技术的发展趋势; 从技术与实践角度出发, 探讨“大数据生态系统”、“大数据技术”、“大数据应用”、“大数据基础设施”等领域的新技术应用和实践经验; 通过创新大赛和培训课程等方式, 深度剖析大数据创



业热点和分享行业实战经验。为了更好地讨论大数据技术生态系统的现状和发展趋势, 交流大数据技术实践经验, 进一步推进大数据技术创新与应用, 展示国内外大数据领域的最新成果。

BDTC 2014 特邀近百位来自全球大数据产业界知名学者、企业领军人物、行业资深专家、一线实践技术代表, 内容涵盖 Hadoop、YARN、Spark、Tez、HBase、Kafka、OceanBase 等开源软件的最新进展, NoSQL/NewSQL、内存计算、流计算和图计算技术的发展趋势, OpenStack 生态系统对于大数据计算需求的思考, 以及大数据下的可视化、机器学习 / 深度学习、商业智能、数据分析等的最新业界应用。

在大会开幕式上, 中科院计算所网络数据科学与技术重点实验室主任程学旗研究员作为 CCF 大数据专家委员会秘书长发布了《中国大数据技术与产业发展白皮书 (2014)》和《2015 大数据十大发展趋势预测》, 代表了百余位业内专家在大数据典型应用现状、大数据技术体系现状、大数据 IT 产业链与生态环境、大数据人才资源、大数据发展趋势与建议等方面的深入思考, 为企业制定大数据战略规划提供了极具价值的参考建议。

作为中国大数据领域最具价值的 IT 盛会, 历经七届沉淀, 中国大数据技术大会已经成为中国最具影响、规模最大、参会者人数最多的大数据领域技术盛会。2014 中国大数据技术大会对于产业的发展与推进有着重要意义。2014 中国大数据技术大会, 六十余场主题演讲、技术论坛和专业培训, 数千名业内人士与会齐聚的深度实践之旅。作为大数据技术与应用深度结合的新起点, BDTC 2014 已经成为产业界、科技界与政府部门密切合作的新平台, 为推动我国大数据的产学研用做出了重大贡献。



# 知识产权

2014年,计算所申请发明专利214项,获得新授权发明专利68项,授权专利总数达1095项。专利数量在严格控制质量的前提下稳定增长,并且在专利布局和组合申请方面继续开展深度规划和挖掘,改变单点申请的局面。2014年,计算所获软件登记证书36项,制修订标准提案7项。

本年度深入推进专利价值分析和分级分类管理工作,通过专利评价、评级,科学合理判断专利使用价值,找出适合市场需求主体的专利,为促进专利转移转化进行有益尝试。通过开展动态专利分级分类管理工作,使专利价值分析工作贯穿专利管理的全过程。2014年转让专利4项,授权许可专利22项,成交金额127万元。

2014年计算所发表代表性论文94篇。

## □ 授权专利列表

序号	授权号	专利名称	发明 / 设计人
1	ZL200810226029.6	一种协议数据单元构造和解析的方法及装置	杜洁; 董江涛; 石晶林
2	ZL200910244005.8	一种基于规则的分布式推理方法及系统	叶剑; 李锦涛; 史红周; 朱珍民; 杜静
3	ZL201010196406.3	一种面向在线多媒体信息挖掘的数据缓存方法和系统	张冬明; 张勇东; 刘峰; 舒敏; 刘毅; 顾晓光; 安茂波; 云晓春; 李锦涛
4	ZL201010138515.X	创建多核处理器虚拟机系统可信环境的方法及装置	杜磊; 孙毓忠; 宋擒豹
5	ZL201010126867.3	一种路径结果文字描述方法及装置	赵婷; 方金云
6	ZL201010211419.3	一种基于多视觉信息的目标定位方法和系统	罗海勇; 张波; 赵方; 朱珍民; 何哲
7	ZL201010292031.0	一种蛋白质鉴定的大规模分布式并行加速方法及其系统	王乐珩; 王文平; 迟浩; 吴妍洁; 周郴; 付岩; 孙瑞祥; 贺思敏

序号	授权号	专利名称	发明 / 设计人
8	ZL201010287612.5	检测网络安全性的方法及系统	王元卓; 李金明; 喻民; 林思明; 程学旗
9	ZL201010546475.2	一种用后缀数组加速大规模蛋白质鉴定的方法及其系统	周郴; 迟浩; 王乐珩; 李由; 吴研洁; 付岩; 孙瑞祥; 贺思敏
10	ZL201010290544.8	一种基于胖树结构的动态容错方法和系统	胡农达; 王达伟; 安学军; 孙凝晖
11	ZL201010620376.4	脉象检测系统	张静静; 王睿; 陆世龙; 崔莉
12	ZL201010589387.0	一种路灯杆的数字化管理与智能交互装置、方法及其系统	钱跃良; 张博宁; 褚诚缘; 谢萦; 王向东
13	ZL201010292032.5	一种蛋白质鉴定的大规模分布式并行加速方法及系统	王乐珩; 王文平; 迟浩; 吴妍洁; 周郴; 付岩; 孙瑞祥; 贺思敏
14	ZL201010292060.7	一种蛋白质鉴定的大规模分布式并行加速方法及其系统	王乐珩; 王文平; 迟浩; 吴妍洁; 周郴; 付岩; 孙瑞祥; 贺思敏
15	ZL201110399845.9	多址接入信道下自适应编码中继系统与方法	周恩; 周一青; 石晶林
16	ZL201110033439.0	一种分布式文件系统及其访问方法	廖浩均; 韩冀中; 戴娇; 周薇; 路远征
17	ZL201110099439.0	一种话音传输控制系统及方法	孙毅; 宋翊麟
18	ZL201110432327.2	基于社会信息的网络视频在线地理定位系统	宋一丞; 曹娟; 夏添; 张勇东; 李锦涛
19	ZL201110190662.6	一种基于 DDR SDRAM 的栈式数据缓存装置及其方法	王展; 曹政; 陈飞; 王凯; 安学军; 孙凝晖
20	ZL201110161094.7	一种对 WuManber 算法进行测试的方法和系统	张宇; 刘萍; 刘燕兵; 郭莉; 陈训逊; 李楠宁; 薛晨
21	ZL201210046913.8	一种基于 DDR SDRAM 的栈式数据缓存装置及其方法	王展; 曹政; 陈飞; 王凯; 安学军; 孙凝晖
22	ZL201210048441.X	一种智能电视的交互方法与系统	钱跃良; 王向东; 刘宏
23	ZL201210193655.6	面向社区的发布订阅系统重聚集方法及重聚集系统	李伟; 虎嵩林
24	ZL201210048483.3	一种在线迭代编译的优化方法和优化系统	陈洋; 房双德; 吴承勇; Olivier Temam(奥里维·特曼); Lieven Eeckhout (利文·艾可霍特)



序号	授权号	专利名称	发明 / 设计人
25	ZL201210096454.4	一种支持单根 I/O 虚拟化用户级接口控制装置及其方法	曹政；刘飞龙；刘小丽；安学军；张佩珩；孙凝晖
26	ZL201210160593.9	用于多根 I/O 虚拟化共享系统的 I/O 资源管理方法	刘小丽；曹政；安学军；张佩珩；孙凝晖；王展；苏勇
27	ZL201210160431.5	全虚拟化环境下启动加载器的可信验证方法和系统	高云伟；薛栋梁；邬小龙；孙毓忠
28	ZL201210160617.0	一种针对硬件虚拟化的内存监控方法和系统	田昕晖；高云伟；邬小龙；孙毓忠
29	ZL201210159873.8	一种机群文件系统分布式元数据一致性保证方法和系统	张军伟、郑彩平、邵冰清、王利虎、刘振军、刘浏
30	ZL201210212340.1	一种虚拟路由器的数据转发平面的构建方法与系统	关洪涛；张建华；谢高岗；贺鹏；王永功；李坤丽
31	ZL201210176976.5	一种基于局部虚拟化的虚拟网构建方法和系统	关洪涛；张建华；谢高岗；刘俊杰；王永功；贺鹏
32	ZL201210206777.4	一种面向需求侧响应基于预测的三相负载调度方法及装置	罗海勇、张雨晨、朱珍民、王向东
33	ZL201210347911.2	一种可控的火焰动画生成方法及其系统	朱登明；冯笑冰；魏毅；王兆其
34	ZL201110252255.3	用于网络安全设备性能测试的测试流量合成方法及装置	肖军；张永铮；云晓春
35	ZL201110303491.3	对 GPS L5 信号的捕获方法及装置	崔绍龙；姚相振；覃新贤；方金云
36	ZL201210003524.7	6LoWPAN 网络密钥建立方法	杨德兴；史红周；刘钦明
37	ZL200910076428.3	一种跨领域的文本情感分类器的训练方法和分类方法	谭松波；程学旗
38	ZL201110253544.5	一种面向分布式工作流的数据流调度方法和系统	王桂玲；张鹏；刘晨；徐学辉；季光；韩燕波
39	ZL201110339200.6	一种数字信息推荐预测模型的训练方法和系统	鲁凯；王斌；史亮；李文娜；李锐；徐飞
40	ZL201110339736.8	一种数字信息推荐预测模型的训练方法和系统	鲁凯；王斌；史亮；李文娜；李锐；徐飞
41	ZL201110147921.7	针对群体绘制的效率控制方法	张言丰；毛天露；蒋浩；王兆其

序号	授权号	专利名称	发明 / 设计人
42	ZL201110220842.4	汉语语音识别中重复出现词识别错误的自动修正方法	李新辉；王向东；钱跃良；林守勋
43	ZL201010257428.6	一种用于环境监测的传感器网络及数据传输方法	李栋；宋磊；张乐；黄庭培崔莉；赵泽；陆世龙；刘强
44	ZL201010171308.4	一种在无线 MESH 网络中基于冲突避免的调度方法	胡金龙；周俊；张晶；唐槐；石晶林
45	ZL201010183170.X	人体脉搏信息采集装置及人体健康状况监护装置	王睿；陆世龙；崔莉；宫继兵
46	ZL201110269248.4	传感器网络节点能量自供给方法和装置	张静静；赵泽；陈海明；崔莉
47	ZL201110286512.5	IP 查找方法和装置以及路由更新方法和装置	罗腊咏；谢高岗；谢应科
48	ZL201210010315.5	微博搜索排名方法及微博搜索引擎	程学旗；陈根宝；李静远；王元卓；邢国亮；方滨兴
49	ZL201110435850.0	保障 RRC 连接状态的 UE 的 MBMS 服务连续性的方法	周一青；周黎阳；关娜；田霖；石晶林；史岗
50	ZL201110441775.9	海量结构化数据复杂查询任务的分布式查询方法和系统	吴广君；李超；王树鹏；云晓春；王勇；李斌斌
51	ZL201110442091.0	海量结构化数据复杂查询任务的分布式查询方法和系统	吴广君；李超；王树鹏；云晓春；王勇；李斌斌
52	ZL201110443604.X	基于分布熵的局部敏感哈希高维索引方法	张伟高科张勇东李锦涛
53	ZL201110270591.0	用于并行多处理器系统的通信行为获取装置	王展；曹政；刘锐；苏勇；刘小丽；刘厚贵；靳庆贺；安学军；孙凝晖
54	ZL201210125122.4	虚拟人组合任务规划方法	李淳范；宗丹；夏时洪；王兆其
55	ZL201210166934.3	用于计算机辅助翻译的候选短语查询方法及辅助翻译系统	刘群；王洋；刘洋；骆卫华；吕雅娟
56	ZL201110195375.4	一种视频质量评价方法	黄庆明；许倩倩；蒋婷婷；苏荔；蒋树强
57	ZL201210222936.X	依存映射方法及系统	刘凯；姜文斌；吕雅娟；刘群
58	ZL201110424401.6	内核态虚拟网络设备的建立方法、及其包发送和接收方法	林思明；周洲仪；程学旗；张冬；李金明；王元卓；梁英



序号	授权号	专利名称	发明 / 设计人
59	ZL201310127459.3	在线社会网络中网络群体的检测方法及系统	程学旗、王元卓、于建业、李静远
60	ZL201210227973.x	一种基于依存句法树的翻译规则抽取方法和翻译方法	谢军；米海涛；刘群
61	ZL201010517817.8	一种物-物智能交互装置及其方法、交互系统	钱跃良
62	ZL201110333844.4	一种服务器瞬间拥塞控制方法和系统	肖军；张永铮；云晓春
63	ZL201210088773.0	移动阅读中的社交关系挖掘方法及装置	叶剑；朱珍民；张磊；李艳兵；唐熊；肖灿；杜静
64	ZL201110451208.1	内存总线的信号采集装置	崔泽汉；陈明宇；包云岗；朱晏；张金勇
65	ZL201210007365.8	SRAM 型 FPGA 的低功耗设计方法	黄柯衡；胡瑜；李晓维
66	ZL201110322252.2	维护集群缓存一致性的方法及系统	司成祥；许鲁
67	ZL201310077524.6	微博中用户间潜在关注关系的发现方法及装置	程学旗、贾岩涛、王元卓、于建业、李静远、冯凯
68	ZL201310112260.3	一种基于射频信号强度的 AP 位置估计方法	罗海勇、耿皓、赵方、朱珍民

从 优 秀 到 卓 越

## □ 软件登记列表

序号	登记号	软件名称
1	2014SR013253	VIPL 结合颜色 \ 形状信息的行人检测软件 P.Detection
2	2014SR004210	VIPL 排队人数估计系统
3	2014SR004299	VIPL 人脸检测软件
4	2014SR013254	VIPL 人体动作识别软件 P.Action
5	2014SR013256	VIPL 行人检索软件
6	2014SR003789	面向移动互联网的业务托管和运行平台
7	2014SR014165	三维虚拟人交互游戏系统
8	2014SR130071	高精度串联质谱肽段母离子检测软件
9	2014SR130087	蛋白质质谱分析的开放式搜索软件
10	2014SR014171	skyFS 机群文件系统软件
11	2014SR017556	增强型搜索引擎服务系统
12	2014SR017551	虚拟机功耗测量系统
13	2014SR057743	天玑微博舆情分析软件系统
14	2014SR057251	学术移动助手软件
15	2014SR056355	矢量叠加并行计算程序软件
16	2014SR056288	电子断层三维重构软件
17	2014SR056223	细粒度法律检索服务软件
18	2014SR057912	法律信息资源规范化处理软件
19	2014SR056291	矢量缓冲并行计算程序软件
20	2014SR057094	面向湖库区的多维异构数据实时计算显示系统
21	2014SR057436	面向湖库区的多维异构数据集成展示系统
22	2014SR057259	地理计算并行算法性能评测软件
23	2014SR057745	蛋白质基因组注释与评价工具
24	2014SR057734	面向业务区分优化的移动管理系统
25	2014SR113072	Flowx 网络数据流导出系统



序号	登记号	软件名称
26	2014SR113002	群体社会行为仿真实验平台
27	2014SR113333	人口撤离方案交互设置系统
28	2014SR113327	人口撤离方案模拟系统
29	2014SR113031	cHPPNetSim 并行网络模拟器
30	2014SR113336	基于 D2D 的协作多播性能仿真系统
31	2014SR113005	复杂系统风险评估专家系统
32	2014SR128782	蛋白质质谱分析的鉴定搜索引擎前处理软件
33	2014SR130089	基于质谱数据的完整蛋白质鉴定软件
34	2014SR128264	天玑 YouTube 数据实时采集与分析系统
35	2014SR128277	天玑互联网新应用数据采集与特征分析系统
36	2014SR163925	大数据挖掘云引擎系统

## □ 标准列表

序号	标准号	中文名称	标准范围
1	GB/T 30996.1-2014	信息技术 实时定位系统 第 1 部分：应用程序接口	国家标准
2	GB/T 29261.5-2014	信息技术 自动识别和数据采集技术 词汇 第 5 部分：定位系统	国家标准
3	GB/T 31101-2014	信息技术 自动识别和数据采集技术 实时定位系统性能测试方法	国家标准
4	YD/T 2710-2014	IPv6 路由协议 适用于低功耗有损网络的 IPv6 路由协议 (RPL)	行业标准
5	YD/T 1615-2013	IPv6 技术要求 基于网络的流切换移动	行业标准
6	YD/T 0973-2013	IPv6 技术要求 代理移动 IPv6 路由优化	行业标准
7	YD/T 2705-2014	持续数据保护 (CDP) 灾备技术要求	行业标准

## □ 代表性论文列表

### 会议论文：

#### CCFA类：

1. Yunji Chen, Tao Luo, Shaoli Liu, Shijin Zhang, Liqiang He, Jia Wang, Ling Li, Tianshi Chen, Zhiwei Xu, Ninghui Sun, and Olivier Temam, “DaDianNao: A Machine-Learning Supercomputer” ,IEEE/ACM International Symposium on Microarchitecture (MICRO), 2014. (Best Paper Award)
2. Tianshi Chen, Zidong Du, Ninghui Sun, Jia Wang, Chengyong Wu, Yunji Chen, and Olivier Temam, “ DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning” , ACM ASPLOS, 2014. (Best Paper Award)
3. Tianshi Chen, Qi Guo, Ke Tang, Olivier Temam, Zhiwei Xu, Zhi-Hua Zhou, and Yunji Chen, “ArchRanker: A Ranking Approach to Design Space Exploration” , ACM/IEEE ISCA, June 2014
4. Qing Yi, Qian Wang and Huimin Cui, “Specializing Compiler Optimizations Through Programmable Composition For Dense Matrix Computations” , IEEE/ACM MICRO, 2014
5. Dawei Zang, Zheng Cao\*, Zhan Wang, Xiaoli Liu, Lin Wang, Ninghui Sun. “Decentralized NIC-switching Architecture Using SR-IOV PCIe Network Device,” Micro IEEE, vol. 34 (5), pp. 42-50, 2014. IF. 1.812
6. Lei Wang, Chunjie Luo, Yongqiang He, Jianfeng Zhan, Kent Zhan, Xiaona Li, Yuqing Zhu, Shujie Zhang, Qiang Yang, Bishu Qiu, Zhen Jia, “BigDataBench: a Big Data Benchmark Suite from Internet Services” ,The 20th IEEE International Symposium On High Performance Computer Architecture (HPCA-2014), February 15-19, 2014 ,Page(s):488 – 499.
7. Xueqi Cheng, Yanyan Lan, Jiafeng Guo, Xiaohui Yan, BTM: Topic Modeling over Short Texts, IEEE Transactions on Knowledge and Data Engineering, vol. 99.
8. Suqi Cheng, Huawei Shen, Junming Huang, Wei Chen, Xueqi Cheng, IMRank: influence maximization via finding self-consistent ranking, The 37th International ACM SIGIR conference on research and development in Information Retrieval.



9. Yadong Zhu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng, Shuzi Niu, Learning for Search Result Diversification, The 37th International ACM SIGIR conference on research and development in Information Retrieval, New York, USA, 293-302.
10. Zhuo Zhang, Zhibin Zhang, Yunjie Liu and Gaogang Xie, ProWord: An Unsupervised Approach to Protocol FeatureWord Extraction, Annual IEEE International Conference on Computer Communications, Canada, 2014-4-27, 1393-1401.
11. Zhenyu Li, Gaogang Xie, Jiali Lin, Yun Jin, Mohamed-Ali Kaafar, Kave Salamatian, On the Geographic Patterns of a Large-scale Mobile, Video-on-Demand System, Annual IEEE International Conference on Computer Communications, Canada, 2014-4-27, 397-405.
12. Tong Yang, Gaogang Xie, YanBiao Li, Qiaobin Fu, Alex X. Liu, Qi Li, Laurent Mathy, Guarantee IP Lookup Performance with FIB Explosion, ACM SIGCOMM, USA, 2014-8-17, 39-50.
13. Wan, Ji, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study." In Proceedings of the ACM International Conference on Multimedia. Orlando, FL, USA, November 03 - 07, 2014. pp. 157-166.
14. Xiaopeng Yang, Tao Mei, Yongdong Zhang; Rescue Tail Queries: Learning to Image Search Re-rank via Click-wise Multimodal Fusion. ACM Multimedia 2014: 537-546,
15. Meina Kan, Shiguang Shan, Hong Chang, Xilin Chen, Stacked Progressive Auto-Encoder (SPAЕ) for Face Recognition Across Poses, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 美国, June 23-27
16. Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xilin Chen, Learning Euclidean-to-Riemannian Metric for Point-to-Set Classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 美国, June 23-27
17. Mengyi Liu, Shiguang Shan, Ruiping Wang, Xilin Chen, Learning Expressionlets on Spatio-Temporal Manifold for Dynamic Facial Expression Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 美国, June 23-27
18. Risheng Liu, Junjie Cao, Zhouchen Lin, Shiguang Shan, Adaptive Partial Differential Equation Learning for Visual Saliency Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 美国, June 23-27

19.Zhenyu Chen, Yiqiang Chen, Shuangquan Wang, Lisha Hu, Xinlong Jiang, Xiaojuan Ma, Nicholas D. Lane, Andrew T. Campbell. ContextSense: Unobtrusive Discovering Incremental Contexts using Dynamic Bluetooth Data. UbiComp Adjunct, Seattle, USA, Sep. 13-17, 2014: 23-26

20.Yang Gu, Yiqiang Chen, Junfa Liu, Xinlong Jiang: SAP dissimilarity based high performance Wi-Fi indoor localization. UbiComp Adjunct, Seattle, USA, Sep. 13-17, 2014: 55-58

21.Hua-Wei Shen, Dashun Wang, Chaoming Song, and Albert László Barabási, Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes, Proceedings of the 28th AAAI Conference on Artificial Intelligence.

22.Yue Yin, Bo An, Manish Jain, Game-theoretic resource allocation for protecting large public events, 28th AAAI Conference on Artificial Intelligence, 加拿大, Jul.27-31, 826-834

#### CCF B 类:

23.Xing Hu, Yi Xu, Jun Ma, Guoqing Chen, Yu Hu, Yuan Xie, "Thermal-sustainable power budgeting for dynamic threading" , IEEE/ACM DAC, pp.1-6, June, 2014

24.Yinhe Han, Ying Wang, Huawei Li, Xiaowei Li, "Data-Aware DRAM Refresh to Squeeze the Margin of Retention Time in Hybrid Memory Cube" , IEEE/ACM ICCAD, pp.295-300, November, 2014

25.Xin He, Guihai Yan, Yinhe Han, and Xiaowei Li, "SuperRange: Wide Operational Range Power Delivery Design for both STV and NTV Computing," Proc. Design Automation and Test in Europe (DATE), Dresden, , Paper 6.4-2, March, 2014

26.Jian Wang, Huawei Li, Tao Lv, Tiancheng Wang, and Xiaowei Li, "Functional Test Generation Guided by Steady-State Probabilities of Abstract Design," Proc. Design Automation and Test in Europe (DATE), Dresden, Paper IP5-13, March 2014

27.Bing Li, Shuchang Shan, Yu Hu, Xiaowei Li, "Partial-SET: Write Speedup of PCM Main Memory," Proc. Design Automation and Test in Europe (DATE), Dresden, Paper IP1-20, March 2014

28.Lunkai Zhang, Dmitri Strukov, Hebatallah Saadeldeen, Dongrui Fan, Mingzhe Zhang, Diana Franklin.SpongeDirectory: Flexible Sparse Directories Utilizing Multi-Level Memristors, PACT, 2014, page 61-73.



- 29.Yulong Luo, Guangming Tan: Optimizing stencil code via locality of computation. International Conference on Parallel Architectures and Compilation, PACT '14, Edmonton, AB, Canada, August 24-27, 2014. ACM 2014, ISBN 978-1-4503-2809-8: 477-478
- 30.Yaqin Zhou, Qiuyuan Huang, Fan Li, Xiang-yang Li, Min Liu, Zhongcheng Li, Zhiyuan Yin, Almost Optimal Channel Access in Multi-Hop Networks With Unknown Channel Variables, Distributed Computing Systems(ICDCS), 2014 IEEE 34th International conference on, Spain, 2014-6-30, 461-470.
- 31.Yan Shengen, Li Chao, Zhang Yunquan, Zhou Huiyang. "yaSpMV: Yet Another SpMV Framework on GPUs" . PPoPP, Orlando, USA, 2014.
- 32.Peng He, Gaogang Xie, Kave Salamatianz, Laurent Mathy, IEEE International Conference on Network Protocols, USA, 2014-10-21.
- 33.Wei Wang, Yi Sun, Kai Zheng, Mohamed Ali Kaafar, Dan Li, Zhongcheng Li, Concise Paper: Freeway: Adaptively Isolating the Elephant and Mice Flows on Different Transmission Paths, The 22nd IEEE International Conference on Network Protocols, USA, 2014-10-21, 362-367.
- 34.Yi Sun, Seyed K. Fayazy, Yang Guo, Vyas Sekary, Yun Jin, Mohamed Ali Kaafar, Steve Uhlig, Trace-Driven Analysis of ICN Caching Algorithms on Video-on-Demand Workloads, The 10th International Conference on emerging Networking EXperiments and Technologies (CoNEXT), Australia, 2014-12-2.
- 35.Zhen Cui, Hong Chang, Shiguang Shan, and Xilin Chen, Generalized Unsupervised Manifold Alignment, Neural Information Processing Systems, 加拿大, Dec.8-11
- 36.Jie Zhang, Shiguang Shan, Meina Kan, Xilin Chen, Coarse-to-Fine Auto-encoder Networks (CFAN) for Real-time Face Alignment, European Conference on Computer Vision, 瑞士, Sep.6-12
- 37.Zhen Cui, Hong Chang, Shiguang Shan, Bineng Zhong, Xilin Chen, Deep Network Cascade for Image Super-resolution, European Conference on Computer Vision, 瑞士, Sep.6-12
- 38.Xing Yan, Hong Chang, Shiguang Shan, Xilin Chen, Modeling Video Dynamics with Deep dynencoder, European Conference on Computer Vision, 瑞士, Sep.6-12
- 39.Xin Jin, Fuzhen Zhuang, Hui Xiong, Changying Du, Ping Luo and Qing He, Multi-task Multi-view Learning for Heterogeneous Tasks, ACM Conference on Information and Knowledge Management(CIKM 2014), 中国, Nov.3-7.

40.Kai Shu, Ping Luo, Wan Li, Peifeng Yin, Linpeng Tang, Deal or Deceit: Detecting Cheating in Distribution Channels, ACM Conference on Information and Knowledge Management(CIKM 2014), 中国, Nov.3-7.

41.Hui Yu, Xiaofeng Wu, Jun Xie, Wenbin Jiang, Qun Liu and Shouxun Lin, RED: A Reference Dependency Based MT Evaluation Metric, the 25th International Conference on Computational Linguistics. Pages 2042-2051, 爱尔兰, Aug.23-29

42.Wenjuan Luo, Fuzhen Zhuang, Xiaohu Cheng, Qing He, Zhongzhi Shi, Ratable Aspects over Sentiments: Predicting Ratings for Unrated Reviews, IEEE International Conference on Data Mining (ICDM 2014)。中国, Dec.14-17

43.Lingyang Chu, Shuhui Wang, Yanyan Zhang, Shuqiang Jiang, Qingming Huang, Graph-Density-based visual word vocabulary for image retrieval. IEEE International Conference on Multimedia, International Conference on Multimedia and Expo (ICME), July 14-18, Chengdu, China, 2014

44.Shuhui Wang, Zhenjun Wang, Shuqiang Jiang, Qingming Huang, Cross media topic analytics based on synergetic content and user behavior modeling. International Conference on Multimedia and Expo (ICME), July 14-18, Chengdu, China, 2014

### CCFC 类:

45.N. Wang, Y. Su, J.L. Shi, Y. Zhou and G. Gui, Sparse Channel Estimation for OFDM Based Two-Way Relay Networks, IEEE ICC2014, Jun. 2014.

46.G. Zhai, L. Tian, Y. Zhou and J.L. Shi, Load Diversity Based Processing Resource Allocation for Super Base Stations in Large-scale Centralized Radio Access Networks, IEEE ICC2014, Jun. 2014.

47.Yi Sun, Junchen Jiangy, Vyas Sekary, Hui Zhangy, Fuyuan Lin, Nanshu Wang,Using Video-Based Measurements to Generate a Real-Time Network Traffic Map,The Thirteenth ACM Workshop on Hot Topics in Networks (HotNets-2014),USA, 2014-12-2.

### 其它:

48.Y. Zhou, L. Liu, H.Y. Du, L. Tian and J.L. Shi, An Overview on Intercell Interference Management in Mobile Cellular Networks: from 2G to 5G, accepted by IEEE ICCS2014.

49.Y. Wang, M. Qian, X. Han, Y. Zhou and J.L. Shi, Game-Theoretic Power Control for Interference Mitigation in Two-tier Small Cell Networks, VTC2014-Spring, 2014.



## 期刊论文：

### CCF A类：

1. Y. Zhou, H. Liu, Z. Pan, L. Tian, J.L. Shi and G. Yang, Two-Stage Cooperative Multicast Transmission with Optimized Power Consumption and Guaranteed Coverage, IEEE JSAC on SEED, Vol.32, Issue 2, pp. 274-284, Feb. 2014.
2. Zhuo Zhang, Zhibin Zhang, Patrick P. C. Lee, Yunjie Liu, and Gaogang Xie, Toward Unsupervised Protocol Feature Word Extraction, IEEE Journal on Selected Areas in Communications, 32, 10, 2014, 1894-1906.
3. Lin Wang, Fa Zhang, Jordi Arjona Aroca, Athanasios V. Vasilakos, Kai Zheng, Chenying Hou, Dan Li, and Zhiyong Liu, “GreenDCN: a General Framework for Achieving Network Energy Efficiency in Data Centers”, IEEE Journal on Selected Areas in Communications (JSAC), vol. 32(1), pp. 4-15, January 2014.
4. Meina Kan, Junting Wu, Shiguang Shan, Xilin Chen, Domain Adaptation for Face Recognition: Targetize Source Domain Bridged by Common Subspace, International Journal of Computer Vision, vol. 109, no.1-2, pp. 94-109, 2014
5. Lei Zhang, Yongdong Zhang, Xiaoguang Gu, Jinhui Tang, Q. Tian: Scalable Similarity Search With Topology Preserving Hashing. IEEE Transactions on Image Processing 23(7): 3025-3039, 2014.
6. Yongdong Zhang, Xiaopeng Yang, Tao Mei: Image Search Reranking With Query-Dependent Click-Based Relevance Feedback. IEEE Transactions on Image Processing 23(10): 4448-4459, 2014.
7. 李绍欣, 刘忻, 柴秀娟, 张海虹, 劳世竑, 山世光, Maximal Likelihood Correspondence Estimation for Face Recognition Across Pose, IEEE Trans.on Image Processing, Vol. 23, No. 10, pp 4587 – 4600, 2014
8. Wei Chen, etc. , Microtransition cascades to percolation, Physical Review Letters, 10.1103/PhysRevLett.112.155701.

## CCFB类：

9. Lei Liu, Yong Li, Zehan Cui, Yungang Bao, Mingyu Chen, Chengyong Wu, “BPM/BPM+: Software-based Memory Partitioning Mechanisms for Eliminating DRAM Bank-/Channel-level Interference in Multicore Systems”, ACM Transactions on Architecture and Code Optimization (TACO), 11(1): Article 5, February 2014
10. Yongbing Huang, Licheng Chen, Zehan Cui, Yuan Ruan, Yungang Bao, Mingyu Chen, Ninghui Sun, “HMTT: A Hybrid Hardware/Software Tracing System for Bridging the DRAM Access Trace’s Semantic Gap”, ACM Transactions on Architecture and Code Optimization (TACO), Volume 11, Issue 1, Article No. 7, February 2014
11. Shuangde Fang, Zidong Du, Yuntan Fang, Yuanjie Huang, Yang Chen, Lieven Eeckhout, Olivier Temam, Huawei Li, Yunji Chen, Chengyong Wu, “Performance Portability Across Heterogeneous SoCs Using a Generalized Library-Based Approach”, ACM Transactions on Architecture and Code Optimization (TACO), Vol. 11, No. 2, Article 21, June 2014
12. Xueliang Li, Guihai Yan, Yinhe Han, Xiaowei Li, “SmartCap: Using Machine Learning for Power Adaptation of Smartphone’s Application Processor”, ACM Transactions on Design Automation of Electronic Systems, Vol.20, No.1, Article 8, November 2014
13. Qi Guo, Tianshi Chen, Yunji Chen, Rui Wang, Huanhuan Chen, Weiwu Hu, Guoliang Chen, “Pre-Silicon Bug Forecast”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD), vol. 33, no.3, 2014, pp.451-463
14. Binzhang Fu, Yinhe Han, Huawei Li, Xiaowei Li, “ZoneDefense: A Fault-Tolerant Routing for 2-D Meshes Without Virtual Channels”, IEEE Transactions on VLSI Systems (TVLSI), Vol.22, No.1, 2014, pp.113-126
15. Keheng Huang, Yu Hu, Xiaowei Li, “A Reliability-Oriented Placement and Routing Algorithm for SRAM-based FPGAs”, IEEE Transactions on VLSI Systems (TVLSI), Vol.22, No.2, Feb 2014, pp.256-269
16. Jing Ye, Yu Hu, Xiaowei Li, Wu-Tung Cheng, Yu Huang, and Huaxing Tang, “Diagnose Failures Caused by Multiple Locations At-a-Time”, IEEE Transactions on VLSI Systems (TVLSI), Vol.22, No.4, April 2014, pp.824-837



- 17.Yuntan Fang, Huawei Li, Xiaowei Li, "Lifetime enhancement techniques for PCM-based image buffer in multimedia applications," IEEE Transactions on VLSI Systems (TVLSI), Vol.22, No.6, June 2014, pp. 1450-1455
- 18.Dawen Xu, Huawei Li, Amirali Ghofrani, K.-T. Cheng, Yinhe Han, Xiaowei Li, "Test-Quality Optimization for Variable n-Detections of Transition Faults" , IEEE Transactions on VLSI Systems (TVLSI), Vol.22, No.8, August 2014, pp. 1738-1749
- 19.Xing Hu, Guihai Yan, Yu Hu, Xiaowei Li, "Orchestrator: Guarding Against Voltage Emergencies in Multi-threaded Application" , IEEE Transactions on VLSI Systems (TVLSI), Vol. 22, No. 12, 2014, pp. 2476 - 2487
- 20.Ying Wang, Lei Zhang, Yinhe Han and Huawei Li, "Reinventing Memory System Design for Many-Accelerators Architecture" , Journal of Computing Science and Technology (JCST), Vol.29, No.2, 2014, pp.273-280.
- 21.Peng Chen, Lei Zhang, Yinhe Han, Yunji Chen, "A General-Purpose Many-Accelerator Architecture Based on DataflowGraph Clustering of Applications," Journal of Computing Science and Technology (JCST), Vol.29, No.2, 2014, pp. 239-246.
- 22.Liu Yiqun, Li Yan, Zhang Yunquan, Zhang Xianyi. "Memory Efficient Two-Pass 3D FFT Algorithm for Intel Xeon PhiTM Coprocessor" .Journal of Computer Science, Technology., 29, 6, 989-1002, 2014.
- 23.V. Garcia, Y. Zhou and J.L. Shi, Coordinated Multipoint Transmission in Dense Cellular Networks with User-Centric Adaptive Clustering, IEEE Trans. Wireless Comm., Vol. 13, No. 8, pp. 4297-4308, Aug. 2014. (Corresponding author)
- 24.Yongdong Zhang, Lei Zhang, Qi Tian: A Prior-Free Weighting Scheme for Binary Code Ranking. IEEE Transactions on Multimedia 16(4): 1127-1139, 2014 .
- 25.Wu Liu, Tao Mei, Yongdong Zhang, Instant Mobile Video Search With Layered Audio-Video Indexing and Progressive Transmission. IEEE Transactions on Multimedia, 16(8):2242-2255, 2014.
- 26.Yu Liu, Jizhong Duan, Qiang Tang, Yongdong Zhang: A Simple and Efficient Re-Scrambling Scheme for DTV Programs. IEEE Transactions on Multimedia 16(1): 137-146, 2014.
- 27.Hongtao Xie, Yongdong Zhang, Jianlong Tan, Li Guo, Jintao Li: Contextual Query Expansion for Image Retrieval. IEEE Transactions on Multimedia 16(4): 1104-1114, 2014 .

28.GuiPeng Li, Ming Li, YiWei Zhang, Dong Wang, Rong Li, Roger Guimer,Juntao Tony Gao and Michael Q. Zhang. ModuleRole:tool for modulization, role determination and visualization in protein-protein interaction networks. *PLoS One*. 2014 May 1;9(5):e94608. doi: 10.1371/journal.pone.0094608.

29.Yaqin Zhou, Xiang-Yang Li, Min Liu, Xufei Mao, Shaojie Tang and Zhongcheng Li, Throughput Optimizing Localized Link Scheduling for Multi-hop Wireless Networks under Physical Interference Model, *Parallel and Distributed Systems, IEEE Transaction on* (Volume:25, Issue:10), Volum25, 2014, 2708-2720.

### 其它：

30.Huawei Shen, etc. ,Collective credit allocation in science, *Proceedings of the National Academy of Sciences*, 08/2014; 111(34).

31.Junming Huang, Chao Li, Wenqiang Wang, Huawei Shen, Guojie Li, Xueqi Cheng. Temporal scaling in information propagation. *Scientific Reports* 4, 5334.

32.Xiaoqian Sun, Huawei Shen, Xueqi Cheng, Trading network predicts stock price, *Scientific Reports*, 01/2014; 4:3711.

33.Bingjie Sun, Huawei Shen and Xueqi Cheng , Detecting overlapping communities in massive networks, *Europhysics letters*, 2014 EPL 108 68001.

34.Qinghua Wu, Zhenyu Li, Jianer Zhou, Heng Jiang, Zhiyang Hu, Yunjie Liu, and Gaogang Xie, SOFIA: Toward Service-Oriented Information Centric Networking , *IEEE Network Magazine*, 28,3,2014,12-18 .

35.Tan Yan, Wensheng Zhang, Guiling Wang, and Yujun Zhang, Access Points Planning in Urban Area for Data Dissemination to Drivers, *IEEE Transactions On Vehicular Technology*, 63,1,2014, 390-402.

36.Y. Zhou, H. Liu, Z. Pan, L. Tian and J.L. Shi, Spectral and Energy Efficient Two-Stage Cooperative Multicast for LTE-A and Beyond, *IEEE Wireless Magazine*, pp. 34-41, Apr. 2014.

37.G.W. Zhai, L. Tian, Y. Zhou and J.L. Shi, Load Diversity Based Optimal Processing Resource Allocation for Super Base Stations in Centralized Radio Access Networks, *Science China Information Sciences*, vol. 57, issue 4, pp. 1-12, Apr. 2014. (Corresponding author)



- 38.V. Garcia, C.S. Chen, Y. Zhou and J.L. Shi, Gibbs Sampling Based Distributed OFDMA Resource Allocation, *Science China Information Sciences*, vol. 57, issue 4, pp. 1-12, Apr. 2014. (Corresponding author)
- 39.K. Yang, J. Yang, J.S. Wu, C.W. Xing and Y. Zhou, Performance Analysis of DF Cooperative Diversity System with OSTBC over spatially correlated Nakagami-m Fading Channels, *IEEE Trans. Vehicular Tech.*, Vol. 63, no. 3, pp. 1270-1281, Mar. 2014.
- 40.Chao Liu, Chun-Qing Song, Zuo-Fei Yuan, Yan Fu, Hao Chi, Le-Heng Wang, Sheng-Bo Fan, Kun Zhang, Wen-Feng Zeng, Si-Min He, Meng-Qiu Dong, and Rui-Xiang Sun, pQuant Improves Quantitation by Keeping out Interfering Signals and Evaluating the Accuracy of Calculated Ratios, *Anal Chem.* 2014 Jun 3;86(11):5286-94.
- 41.Ruoyu Miao, Haitao Luo, Huandi Zhou, Guangbing Li, Dechao Bu, Xiaobo Yang, Xue Zhao, Haohai Zhang, Song Liu, Ying Zhong, Zhen Zou, Yan Zhao, Kuntao Yu, Lian He, Xinting Sang, Shouxian Zhong, Jiefu Huang, Yan Wu, Rebecca A. Miksad, Simon C. Robson, Chengyu Jiang, Yi Zhao, Haitao Zhao. Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *Journal of Hepatology* 2014 Oct;61(4):840-9. doi: 10.1016/j.jhep.2014.05.025.
- 42.Xiaobo Yang, Ming Li, Qi Liu, Yabing Zhang, Junyan Qian, Xueshuai Wan, Anqiang Wang, Haohai Zhang, Chengpei Zhu, Xin Lu, Yilei Mao, Xinting Sang, Haitao Zhao, Yi Zhao, and Xiaoyan Zhang. Dr.VIS v2.0: an updated database of human disease-related viral integration sites in the era of high-throughput deep sequencing. *Nucleic Acids Research* Epub ahead of print 2014 Oct 29.
- 43.Chaoyong Xie, Jiao Yuan, Hui Li, Ming Li, Guoguang Zhao, Dechao Bu, Weimin Zhu, Wei Wu, Runsheng Chen and Yi Zhao. NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.* (1 January 2014) 42 (D1): D98-D103. doi: 10.1093/nar/gkt1222.
- 44.Renmin Han, Fa Zhang, Xiaohua Wan, et. al, “A marker-free automatic alignment method based on scale-invariant features” . *Journal of Structural Biology*, 2014, 196(1): 167-180. (SCI, IF is 3.497).
- 45.Yicheng Song, Yongdong Zhang, Juan Cao, Jinhui Tang, Xingyu Gao, Jintao Li: A Unified Geolocation Framework for Web Videos. *ACM TIST* 5(3): 49, 2014).

# 学位论文

2014 年计算所有 139 人取得硕士学位，有 86 人取得博士学位。2007 级硕博连读生阚美娜完成的题为《基于非完整标注数据的人脸识别研究》的博士学位论文，获 2014 年中科院优秀博士学位论文奖，获 2014 年 CCF 优秀博士学位论文奖。2014 年计算所博士毕业生 66% 进入学术界，其中 39% 毕业生进入科研单位工作。硕士毕业生进入企业工作的比例为 77%，其中去国内企业和外资企业工作的比例为 4.4: 1。

## 优博论文摘要



作者姓名：阚美娜

论文题目：基于非完整标注数据的人脸识别研究

作者简介：阚美娜，女，1984 年 12 月出生，2009 年 9 月师从中国科学院计算技术研究所陈熙霖研究员，于 2013 年 7 月获博士学位。

### 中文摘要

近年来，生物特征识别技术受到了广泛的的关注和应用，可用于有效识别的生物特征主要有指纹、虹膜、人脸、掌纹等。其中，人脸识别具有稳定、唯一、操作简便、成本低等优越的特性，

已被广泛应用于公共安全、监控、个人信息管理、娱乐生活等诸多领域，并且成为了计算机视觉以及模式识别领域的研究热点。过去四十年中，人脸识别技术取得了显著的进展。可控条件下的人脸识别技术已基本趋于成熟，然而在非可控条件下，由于受表情、姿态、光照、年龄、遮挡等因素



素的影响，人脸识别依然是一个很有挑战性的问题。此外，非可控条件下很难得获取到大量包含丰富变化并且标注完整的训练数据，如在对犯罪分子的抓捕中通常只能得到一幅犯罪分子的图像或根据目击者描述得到的素描图像，不完整的标注进一步加剧了非可控条件下人脸识别的难度。本文针对非可控条件下出现的数据标注不完整的问题进行了深入的分析，并提出了使用不同完整程度的标注数据进行模型学习的方法。本文主要研究工作包括：

(1) 针对单一样本下的人脸识别问题，提出了自适应判别分析方法，在单一样本人脸识别上取得了良好的性能。该方法利用一个每人多个样本的世界集来对单一样本的类内散度矩阵进行估计，解决了单一样本场景下没有类内变化的问题。该方法基于长相相似的人具有相似的类内变化的思想，通过组合单一样本的若干近邻的类内散度矩阵来实现估计该单一样本的类内散度矩阵。

(2) 针对边信息情况下的人脸识别问题，提出了基于边信息的判别分析方法。边信息只提供了一对样本是否属于同一个人的信息，没有提供每个样本的类别信息。该方法在边信息的场景下重新定义了类内散度矩阵和类间散度矩阵，从而将 Fisher 线性判别分析扩展到了边信息的场景下。此外，还从理论上证明了类别信息已知时，基于边信息的判别分析与 Fisher 线性判别分析的等价性。

(3) 针对半监督的人脸识别问题，提出了原型超平面学习方法，该方法旨在结合无标注数据和弱标注数据学习一组判别性的中层特征表示，取得显著优于底层特征的性能。该中层特征的每一维对应一个建立在无标注数据上的支持向量机（称为原型超平面）的输出，同时为了得到一组判别式的中层特征表示，通过最大化弱标注数据集上基于中层特征表示的类 Fisher (Fisher-like)

准则得到原型超平面的最优参数。

(4) 针对无监督的领域适应人脸识别问题，提出了目标域化源域样本的方法将源域的知识迁移到目标域上，从而有效地提升了目标域上的识别性能。在无监督领域适应场景下，标注的源域与无标注的目标域的分布是不同的，无法直接用于目标域上的模型学习。本文通过在原始样本空间将源域样本表示为目标域少数近邻样本的线性组合以将源域的样本变换到目标域上，从而得到一组有标注的目标域样本，实现监督知识从源域到目标域的迁移。为了保证目标域化的可靠性和稳定性，本文提出在源域与目标域的公共子空间中对用于目标域化的稀疏线性组合系数进行求解，从而解决了原始样本空间中标注数据与无标注数据分布不同，不能直接用于学习识别模型的问题。

(5) 针对多视图人脸识别问题，提出了多视判别分析的方法，显著提升了跨视图人脸识别的性能。在多视图识别的场景下，来自于多个视图的数据位于不同的空间，通常无法直接进行比对或识别。该方法通过学习多个视图特定的投影矩阵以将多个视图的数据投影到一个公共的子空间，从而去除视图间的差异。该公共子空间是通过最大化所有视图下的类间散度、最小化所有视图下的类内散度得到的判别式公共子空间，进一步提升了跨视图识别的性能。

综上所述，本文对人脸识别中出现的训练数据标注不完整的问题开展了广泛和深入的研究，提出了基于非完整标注数据进行模型学习的方法。大量实验结果表明，本文提出的方法可以有效利用非完整标注的数据学习判别式的模型，显著提升了非完整标注下人脸识别的性能。

**关键词：**人脸识别，单一样本，边信息，原型超平面学习，半监督人脸识别，无监督领域适应，多视判别分析，跨视图人脸识别





中国科学院计算技术研究所  
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

地址：北京市海淀区中关村  
科学院南路 6 号  
邮编：100190  
电话：(010) 62601116  
传真：(010) 62567724  
<http://www.ict.ac.cn>