

对可视化研究的几点思考

陈为

chenwei@cad.zju.edu.cn
<http://www.cad.zju.edu.cn/~chenwei>
浙江大学CAD&CG国家重点实验室

内容

- 如何结合国家重大需求?
 - 活体肝脏移植的可视化
 - 实测断层数据的可视分析
 - 超大规模计算模拟数据场的可视化
- 对于可视化研究的几点看法
 - 传输函数设计在三维医学图像分析中的角色
 - 到底什么是可视分析?
 - 可视化与交互界面
- 如何促进国内学术界的合作?

2

数据分析的挑战

- 将数据转换为对人们有用的信息
- 可视化数据使人们可以更好地理解
- 感知的最高带宽
 - ~100 MB/s
 - 并行
 - 很强的模式识别能力
 - 大部分是下意识的，即不经过思考的

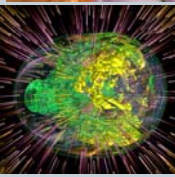
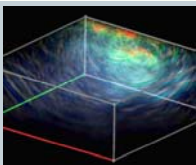
可视化

- 通过使用电脑，对数据进行交互的可视表达以增强认知 [Card, Mackinlay, Shneiderman '98]
- 通常被认为是一个生成图形图像的过程
- 实为认知过程
 - 形成某个物体的感知图像
 - 强化认知理解
- 可视化的目的是洞悉，而非图像
 - 洞察：发现，决策，解释，分析，探索，学习

科学计算可视化

- 探索空间的物理或化学属性
- 与计算紧密结合，多核计算和并行!
- 数据的压缩、存储、传输和快速绘制

UltraVis
SciDAC Institute for Ultrascale Visualization

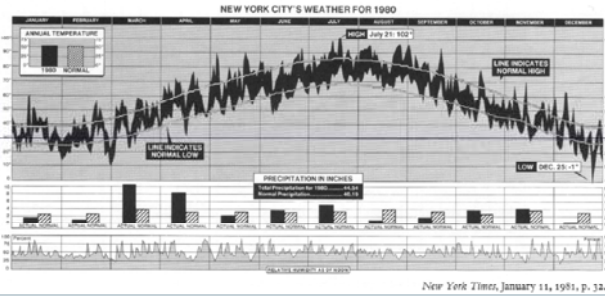


可视化的应用范围

- 数据分析技术
 - 数据挖掘，数据库查询，机器学习
- 可视化尤擅于探索性数据的分析
 - 不知道里面有什么
 - 没有一个预先的探索假设
 - 想知道存在那些疑问
 - I see what you're saying 我明白你的意思
 - Seeing is believing 眼见为实
 - A picture is worth a thousand words 一图胜千言

5

纽约的天气预报



New York Times, January 11, 1961, p. 32.

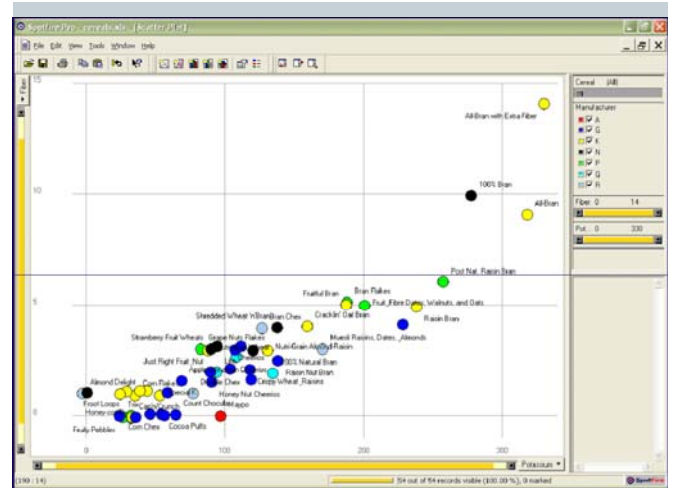
- E. Tufte, Visual Display of Quant Info

伦敦地铁



- 那种作物含钾量最高?
- 纤维和含钾量直接有没有关系?
- 若如此, 有例外吗?
- 那个制造商生产的作物最健康?

	A	B	C	D						
1	Crunch	Manufacturer	7	200	28	Honeycomb	P	0	25	
2	100% Bran	K	10	200	29	Just Right Fruit & Nut	K	2	25	
3	100% Natural Bran	G	2	135	30	Life	Q	4	55	
4	All-Bran	K	5	320	31	Lucky Charms	G	0	55	
5	All-Bran with Extra Fiber	K	14	330	32	Maysn	A	0	95	
6	Almond Delight	R	1	0	33	Mixed Raisins, Dates, &	P	3	170	
7	Apple Cinnamon Cheerik	G	15	70	34	Multi-Grain Cheerios	G	2	90	
8	Bran Chex	R	4	125	35	Nuts-Gran Almond-Rais	K	3	130	
9	Bran Flakes	P	5	180	36	Nutgran Wheat	K	3	90	
10	Cap'n Crunch	G	0	35	37	Original Raisin Crisp	G	15	100	
11	Cheerios	G	2	100	38	Pixel Raisin Bran	P	8	200	
12	Cocoa Puffs	G	0	65	39	Pixel Raisin Bran	K	1	45	
13	Corn Chex	R	0	25	40	Quaker Oatmeal	D	2	7	
14	Corn Flakes	K	1	35	41	Raisin Bran	K	5	240	
15	Count Chocula	G	0	85	42	Raisin Nut Bran	G	2.5	140	
16	Cracker Cut Bran	K	4	160	43	Rice Krispies	K	0	35	
17	Cream of Wheat (Dark)	N	1	0	44	Shredded Wheat	N	3	35	
18	Crisp Wheat & Raisins	G	2	120	45	Shredded Wheat 'n' Bran	N	4	140	
19	Quaker Chew	R	1	80	46	Shredded Wheat Spoon	N	3	120	
20	Frost Loops	K	1	30	47	Strawberries	K	1	40	
21	Frosted Flakes	K	1	25	48	Special K	K	1	55	
22	Fruit & Fibre Dates, Wal	P	5	200	49	Special Fruit Wheaties	N	3	90	
23	Fruited Bran	K	5	190	50	Total Corn Flakes	G	0	25	
24	Fruity Pebbles	P	0	25	51	Total Raisin Bran	G	4	230	
25	Golden Grahams	G	0	45	52	Total Whole Grain	G	3	110	
26	Grape-Nuts Flakes	P	3	85	53	Twee	G	0	25	
27	Honey Nut Cheerios	G	1.5	90	54	Wheaties	G	3	110	
					55	Wheaties Honey Gold	G	1	80	



4个数据集

11

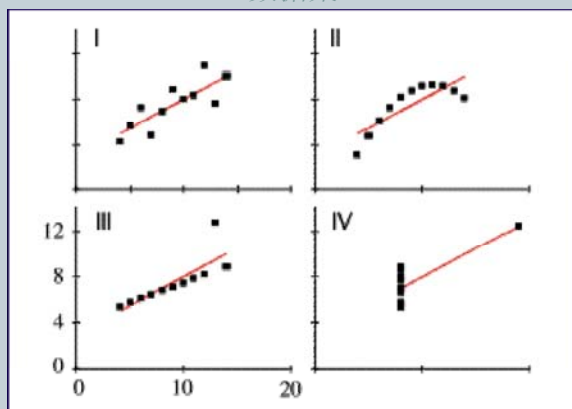
- X均值 = 9.0
- Y均值 = 7.5
- 最小二乘法回归线方程 $y = 3 + 0.5x$
- 误差的平方和 = 110.0
- X值的方差的回归和 = 27.5
- 均方误差的误差和 (相对于回归线) = 13.75
- 相关系数 = 0.82

数值表示

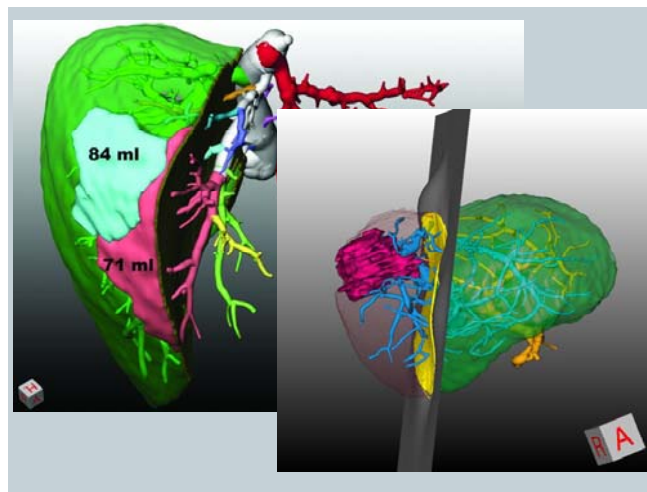
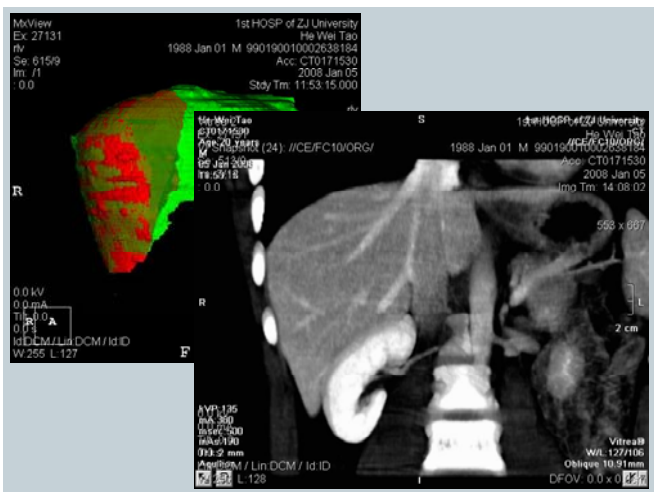
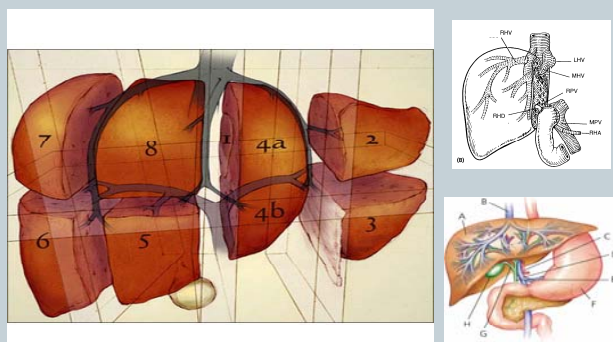
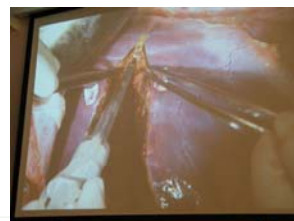
12

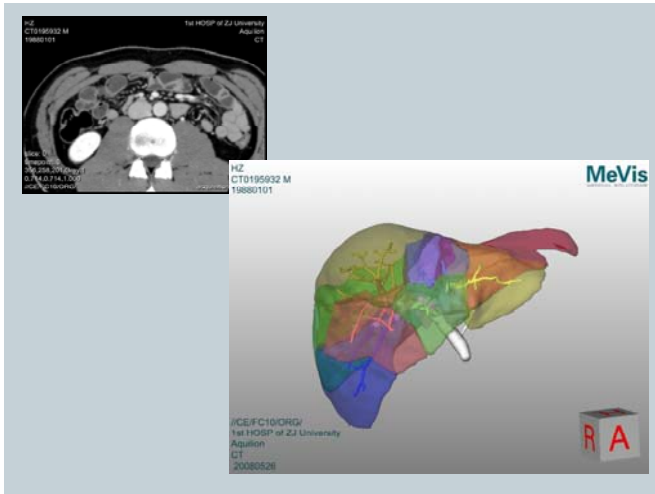
1	2	3	4
10.0, 8.04	10.0, 9.14	10.0, 7.46	8.0, 6.58
8.0, 6.95	8.0, 8.14	8.0, 6.77	8.0, 5.76
13.0, 7.58	13.0, 8.74	13.0, 12.74	8.0, 7.71
9.0, 8.81	9.0, 8.77	9.0, 7.11	8.0, 8.84
11.0, 8.33	11.0, 9.26	11.0, 7.81	8.0, 8.47
14.0, 9.96	14.0, 8.10	14.0, 8.84	8.0, 7.04
6.0, 7.24	6.0, 6.13	6.0, 6.08	8.0, 5.25
4.0, 4.26	4.0, 3.10	4.0, 5.39	19.0, 12.50
12.0, 10.84	12.0, 9.13	12.0, 8.15	8.0, 5.56
7.0, 4.82	7.0, 7.26	7.0, 6.42	8.0, 7.91
5.0, 5.68	5.0, 4.74	5.0, 5.73	8.0, 6.89

数据集



活体肝脏移植的可视化





地质数据的可视分析

目标
提供高质量、高精度、高效的三维地震波图像，用于地层探索

两大部分
第一部分(与数据本身有关)
第二部分(与领域有关)

地质数据的可视分析

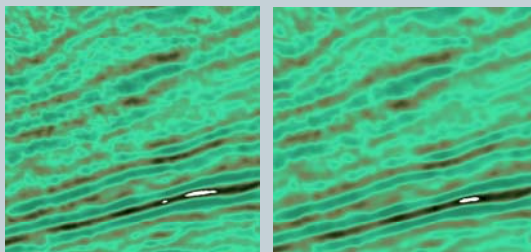
- **高质量可视化 + 结构增强**
 - 地震波数据预处理
 - 地震波数据的高质量直接体绘制
 - 与结构相关的滤波
 - 体内结构增强
- 属性驱动的滤波算法
- 多值体可视化
- 多频谱体可视化与应用

数据预处理

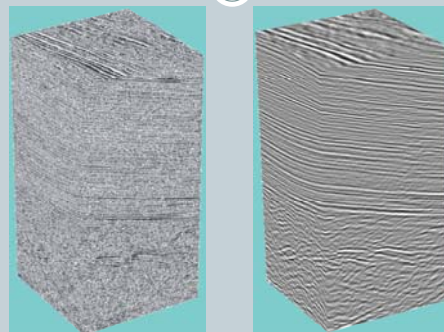
高质量体绘制

体切片观察

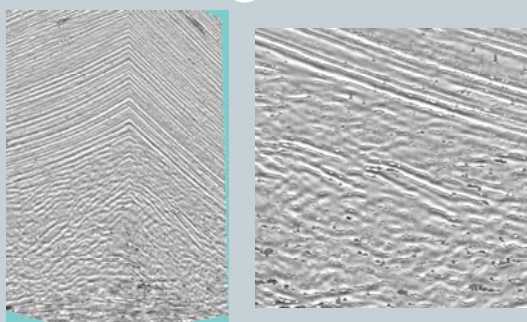
与结构相关的滤波



结构增强



结构增强



超大规模数据可视化

中科院计算机网络信息中心
超级计算中心

28

百万亿次计算机

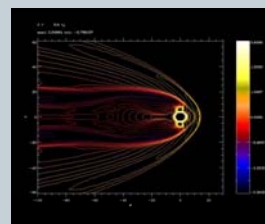
- 深腾7000高效能计算机系统具备每秒106.5万亿次浮点运算能力，2008年11月份第32期TOP500排名第19位（深腾7000排名）。该系统采用了创新的高性能异构体系架构，突破了传统高性能计算平台的应用限制，通过包含多种不同架构服务器节点。

- 空间科学
- 天文学
- 大气科学
- 环境科学
- ...



空间科学：静态磁层大尺度结构的数值模拟

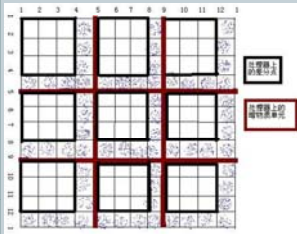
静态磁层大尺度结构的数值模拟，计算结果较好地再现了船艏激波、磁鞘、磁层顶、尾流区、电离层电势分布和整体磁场位形等典型结构，为进一步研究行星际如何影响地球空间打下基础。



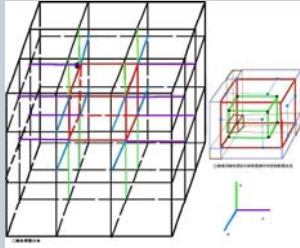
静态磁层大尺度结构模拟

天文领域：多体/流体耦合模型Wigeon（星系演化）

二维流场与暗物质

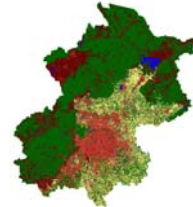


三维示意图

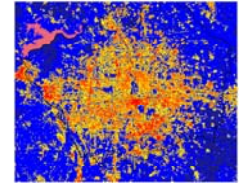


大规模科学计算生态环境研究，指导生态环境可持续发展

中科院生态环境研究中心、中国气象科学研究院运用大规模科学计算进行生态环境研究，在研究城市化和城市发展的生态环境影响方面作了开创性的探索，取得了具有理论意义和应用价值的初步成果。其中，有关北京城市格局对城市环境的模拟研究已运用于北京生态功能区划和生态环境保护规划中。

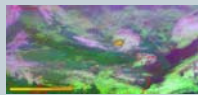


北京市地表覆盖现状

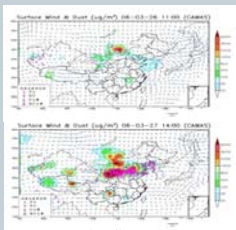


北京市城市热岛分布图

沙尘暴数值预报-- CUACE-Dust计算结果与实际对比



FY2C反演的沙尘暴位置及强度



数值预报的位置与强度

- 原来60小时预报需要10多小时，现在只需20多分钟（使用128个处理器），达到业务预报要求，提高了该系统预报服务能力
- 准确预报了2006年春季大多数沙尘暴过程，通过了中国气象局预测减灾司组织的业务化评估，成为中国气象局沙尘暴业务数值预报系统

目前完成的超大规模应用测试

- 自主并行化的8192核的天文星系风模拟计算
- 1024核的生物蛋白质inspect计算
- 1024核的空间天气MHD计算
- 开源软件Lammps的8192核计算

存在的可视化问题

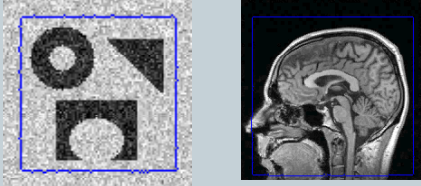
- 超大规模数据
- 复杂网络
- 多变量时变数据分析
- 特征检测，跟踪
- 专家知识
- ...

传输函数在医学图像分析中的作用

分割 vs 分类

37

- 分割：将用户感兴趣的区域分离成多个部分
 - 通常需要用户给定初值，选定作用范围
 - 根据图像内容和初始形状，定义力驱使区域的分割
 - Level set/Graph Cut/Random Walker/Region growing



GBUTem

分割 vs 分类

38

- 分类：根据数据属性（相似性）对一个数据集分为不同的部分
- 传输函数设计：大体上是一个体分类的过程
 - 密度
 - 梯度和高阶微分
 - 本质上是根据某些属性进行归类的方法



GBUTem

体绘制的作用

39

- 作为初始的观察手段，辅助用户理解结构之间的关系
 - 可视化：当我们不知道数据里有什么的时候
- 辅助用户快速定位感兴趣区域，并以此为基础做分割
- 医学可视计算的核心问题是重建、分割和配准
- 如何让体分类更接近分割的效果，定位感兴趣区域？
- 如何设计更有效的用户界面，方便用户实现分割和配准？

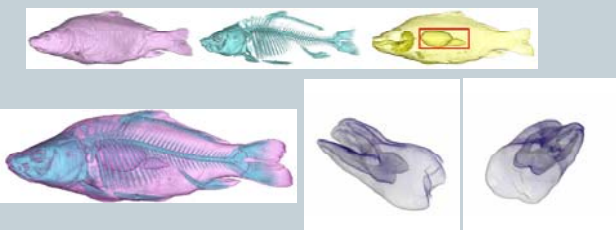
可视化与用户界面

40

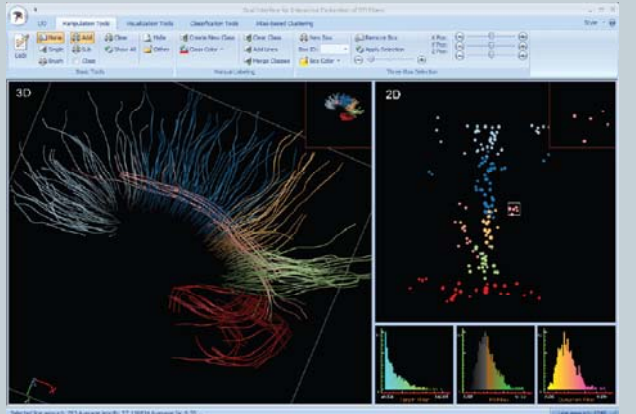
Visualization is an interface!

41

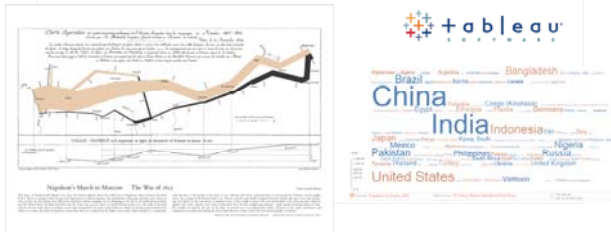
- 将图形和图像领域的一些交互方式引入到体可视化
- 用户的心理和感知是提高效率的要素



交互的DTI线选择



到底什么是可视分析



43

Tabelau

44

- **“Why is there all this computer graphics technology but the only real applications are video games and movies? We wanted to be the ones to create the breakthrough business application that helped everyday people see and understand their data like never before.”**

-- Christian Chabot, CEO and co-founder

信息可视化

45

- 通过交互的计算机可视化来表达、理解抽象的数据，通常指统计数据、数据库、文件和软件系统
- 1990开始兴起的领域，最新的挑战是：
 - × 海量信息可视化
 - × 如何评价
 - × 交互很关键
- 科学计算可视化的对象通常指采集或模拟的空间属性数据
 - × 医学测量数据
 - × 宏观空间信息
 - × 微观空间信息
 - × ...

可视分析的定义

46

- 非正式的说法：采用视觉表达来辅助决策
- 正式的说法：一门基于交互可视界面的分析推理的科学
- InfoVis++ ??
 - 它是信息可视化走向实践、走向领域的加深
- 2005年以来的新兴领域
 - 相较于科学计算可视化，它可考察非空间信息
- 如果我能说清楚它是什么，就不是新兴的

可视分析十大研究问题

47

- 以探索发现为目标的信息推理分析—基于评判式思考 (critical thinking) 的认知方面的新的交互方式
- 能处理多尺度、多类型、动态流的时变数据的新的视觉策略
- 数据、信息和知识的表达
- 协同的、预言性的、主动性的可视分析
- 可视分析方法的刻画与重用

可视分析十大研究问题 (续)

48

- 有效的演示与交流
- 可视的时变分析
- 基于公开数据集的评估与验证
- 保持全局视图情况下的短期效果发布
- 互用性界面与标准：多套可视分析系统的协同

如何促进国内学术界的合作？

算法研究与系统开发
如何和领域专家互动？
如何做出有显示度的成果？

算法研究与系统开发

50

- ▶ 可视化必须面向领域专家
 - 每一项可视化研究是一个与领域结合的过程
 - 同时也是改变合作领域的工作方式的机会
- ▶ 可视化的成果=论文+系统
 - SIGGRAPH和IEEE Visualization开始提倡上传视频、可执行程序甚至代码

如何和领域专家互动？

51

- ▶ 找合适的领域专家合作
 - 生物化学领域的看重知识产权
 - 医学领域的医生太忙
 - 空间计算领域的专家容易沟通，但研究问题不容易提炼
- ▶ 以什么方式合作？
 - 任务分工式远程合作？
 - 物理上聚集在一起？
 - 怎么达到共同兴趣、利益共享？

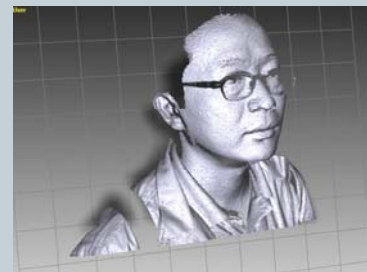
如何做出有显示度的成果？

52

- 可视分析
 - 美欧有不少成功案例
 - 数字城市
 - 国家自然科学基金重点项目？

致谢

- CAD&VR专委会-彭群生教授 -peng@cad.zju.edu.cn
- 活体肝脏移植-郑树森院士
- 浙江大学计算机学院姜晓红 博士
- 地层可视分析-林海教授-lin@cad.zju.edu.cn
- 超大规模可视化-迟学斌教授-chi@sccas.cn
- 浙江大学附属第一医院、浙江大学地理系等



chenwei@cad.zju.edu.cn
<http://www.cad.zju.edu.cn/~chenwei>