

# 从科学计算可视化到信息可视化<sup>\*</sup>

From Visualization in Scientific Computing to Information Visualization

杨 峰

(广东商学院 广东省电子商务市场应用技术重点实验室 广州 510320)

**摘 要** 信息可视化技术是情报学研究的新领域。它是从科学计算可视化发展来的,是可视化众多分支的一个。分析了从科学计算可视化到信息可视化的若干阶段;对比了它们之间各自的特点、相同点和区别;最后从数据来源、处理技术和处理过程上进行了总结。

**关键词** 科学计算可视化 数据可视化 信息可视化

## 1 引 言

信息可视化(Information visualization, InfoVis)是在计算机、网络通讯技术支持下,以认知为目的,对非空间的、非数值型的和高维信息进行交互式视觉表现的理论、技术与方法<sup>[1]</sup>。Info-Vis 是情报学领域一个较新的研究热点。国外,信息管理与信息系统专业、图书情报学专业对这一领域的研究非常活跃,一些大学的信息管理类专业开设了这方面的课程。如美国加州大学伯克利分校信息管理与信息系统学院开设了“信息可视化与表示”(Information visualization and presentation)课程<sup>[2]</sup>。

常言说“百闻不如一见”,视觉是人类获取信息的最重要的途径。人类 80% 以上的信息是通过视觉系统获得的。经过漫长的进化,人类视觉信息处理具有高速、大容量、并行工作的特点。这些特点早已为祖先们所认识和应用,古长城上的烽火台,可将重要的信息迅速大范围传递。千百年来“河图洛书”的传说,显示出“图”在我们文明的发端及以后的发展中所起的作用。

从朴素的使用视觉信号到可视化技术有着一个过程,完成这个重要转变的是计算机技术的使用。今天,可视化技术的大家庭中包括了科学计算可视化(Visualization in Scientific Computing, ViSC)<sup>[3]</sup>、数据可视化(Data Visualization)<sup>[4]</sup>、信息可视化等一系列的分支。一些新的热点,如知识可视化(Knowledge visualization)<sup>[5]</sup>、知识域可视化(Knowledge domain visualization)<sup>[6]</sup>等正在形成。在这些具体分支中科学计算可视化和信息可视化是两个重要代表,前者是可视化技术的第一次出现,后者将可视化技术扩大到非数值、非空间和高维领域。本文总结了从科学计算可视化到信息可视化的发展过程,并对其中各个过程之间的关系进行了比较。

## 2 数值型数据的可视化

2.1 科学计算可视化 ViSC (中文翻译中也简称科学可视化,甚至是可视化),是可视化领域中最早的应用,也看成是计算机图形学的一个重要研究方向,是图形科学的新领域。

“Visualization”一词,来自英文的“Visual”,原意是视觉的、形象的,国内多译成“可视化”,也有译成“视觉化”。事实上,将任何抽象的事务、过程变成图形图像的表示都可以称为可视化。

人们用可视化符号展现事物的方法可以追溯到远古时代,但作为学科术语,“可视化”一词正式出现于 20 世纪 80 年代。1987 年 2 月在美国国家科学基金会(National Science Foundation, NSF)召开的图形图像专题研讨会上,专题讨论组会后发表的正式报告给出了 ViSC 的定义、覆盖的领域,并对可视化的需求、近期目标、远景规划和应用前景方面作了相应的阐述<sup>[7]</sup>。这标志着“科学计算可视化”作为一个学科在国际范围内的确立。ViSC 是用计算机图形学和图形处理技术,将科学计算过程中产生的数据及计算结果转换为图形或图像在屏幕上显示出来,并进行交互处理的理论、方法和技术。它涉及计算机图形学、图像处理、计算机视觉、计算机辅助设计及图形用户界面等多个研究领域。

在计算机用于科学计算的早期,由于计算机硬件条件的限制,科学计算只能以批处理方式进行。由于没有交互功能,科学家不能对计算过程进行干预和引导,只有被动地等待计算结果的输出,且大量的输出结果也是用人工方式处理。最好的辅助方法是用绘图仪以二维的方式展示结果。这样以来,大量信息被处理方法丢失,科学家得不到计算结果的直观、形象的整体印象,处理结果数据的任务繁杂,花费时间往往是计算时间

基金项目:广东商学院博士启动项目资助。

作者简介:杨 峰,1968 年生,副教授,博士,研究方向为信息可视化、网络信息组织、信息管理与电子商务。

的十几倍到几十倍。随着计算机处理图形、图像功能的不断提高,科学计算可视化成为了可能。

科学计算可视化的目的是理解自然的本质。要达到这个目的,科学家把科学数据,包括测量获得的数值或是计算中涉及、产生的数字信息变为直观的、以图形图像形式表示的、随时间和空间变化的物理现象或物理量呈现在研究者面前,使他们能够观察、模拟和计算。这个过程可细化为以下四个步骤: a. 过滤,预处理原始数据、转换数据形式、滤掉噪声、抽取感兴趣的数据等; b. 映射,将过滤后的数据映射为几何元素,常见有点、线、面图元、三维体图元等; c. 绘制,几何元素绘制得到结果图像; d. 反馈,显示图像,并分析得到的可视结果。VisC 技术的意义重大,它大大加快了数据的处理速度,使每日每时都在产生的庞大数据得到有效的利用;实现人与人、人与机器之间的图像通讯,增强了人们观察事物规律的能力;使科学家在得到计算结果的同时,知道在计算过程中发生了什么现象,并可改变参数,观察其影响,对计算过程实现引导和控制。

2.2 空间场可视化 可视化技术应用领域不断扩大,人们对其需求越来越复杂,可视化概念扩展到工程数据和测量数据时,学术界把这种空间数据场的可视化称为体可视化(Volume visualization)技术。一般不将它看成是可视化的一个分支,而作为科学计算可视化的一个方向处理。

体可视化技术研究如何表示、维护和绘制体数据集,从而观察数据内部结构和理解事物复杂特性。它是一种从体数据集中抽取内在的、本质的信息,并以图形图像方式交互地表现出来的技术。体数据集由三维笛卡尔网格体元(网格点上的数据值)构成,将网格体元存储在三维数组(体缓冲区)中。体数据集可来源于许多领域,这决定了体可视化的应用范围。如:工程建筑的空间场,气象卫星测量的空间场,航空航天实验、核爆炸模拟等对应的速度场和温度场,地震预报的力场,工业产品超声波探测的零件密度场,核磁共振产生的人体器官密度场等。

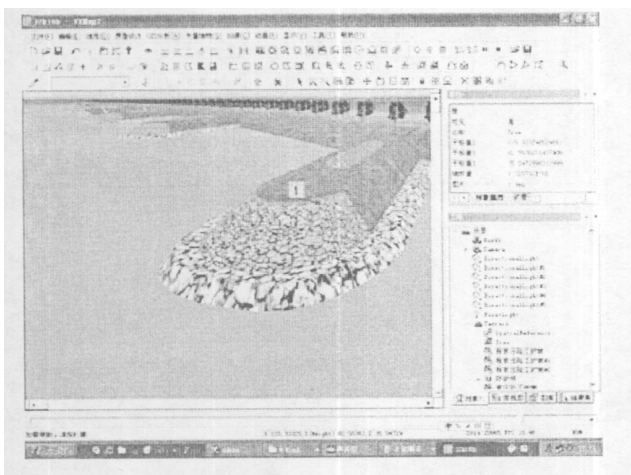


图 1 模拟研究黄河下游河道规律的虚拟试验平台<sup>[9]</sup>

“黄河下游交互式三维视景系统”是一个典型的工程建筑的空间场的可视化系统。范围为黄河小浪底到黄河入海口 900 多公里的河道,系统对小浪底水利枢纽、黄河下游临黄大堤、险工、控导工程、各类涵闸、桥梁、工程管理机构、防汛仓库、村庄

等工程地物采用三维实体模型进行了逼真表现,是“数字黄河下游”的一个基础设施平台(如图 1 所示)<sup>[8]</sup>。

前些年,由于计算机的速度、显示器的分辨率等原因,使得体可视化在数据量即使较小的情况下也要花费很长的时间。解决这个问题来源于两个方面的努力:一是新型算法的研制成功,如表面拟合法(Surface fitting)<sup>[10]</sup>和体绘制(Volume rendering)<sup>[11]</sup>。表面拟合法从数据场中提取等值面后对其进行展示,其显示速度快。缺点是由于只显示指定等值面非常容易丢失数据体中的关键信息。体绘制算法无需几何元素(点、线、面等)作中介表示,而是直接将整个数据场投影到屏幕空间。由于每次显示均要涉及整个数据集,因此它的显示速度比表面拟合法要慢,但它的显示质量很高且不丢失任何信息。二是硬件技术和并行计算技术的发展。并行体可视化既提供快速绘制的可能,又能够处理大的分布式数据。一个体可视化系统一般具有下列功能: a. 高质量的数据表示方法; b. 处理大规模的数据集合,且分辨率较高; c. 快速的图形、图像绘制算法; d. 支持交互式操作,使用户参与控制。

2.3 流场可视化 流场可视化是科学计算可视化的另一个分支方向,同时是流体力学的重要部分。图 2 是一个流场可视化的典型例子。正是流场可视化的研究使得计算流体力学(CFD)快速发展。计算流体力学中的数据类型主要有向量场和标量场。流场可视化用箭头、流线和粒子跟踪技术研究二维流场。对三维流场,人们还没有找到一个普遍接受的、直观的、符合视觉习惯的可视化方法。主要通过以下几种方法进行研究: a. 改造体可视化技术,将其拓展,使其可以用于流场可视化; b. 流面技术,用流面加梯度线索的方法表示流场; c. 图标技术; d. 基于特征的可视化,这是从较高层次进行的流场可视化。

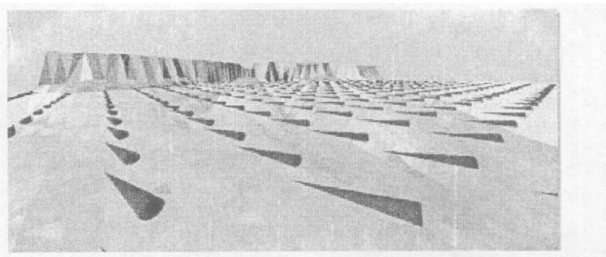


图 2 二维流场的虚拟场景<sup>[12]</sup>

### 3 数据可视化和信息可视化

随着数据仓库技术、网络技术、电子商务技术等新技术的发展,可视化技术扩展到更广泛的内容,提出了数据可视化的概念,即对大型数据库或数据仓库中数据的可视化,它是可视化技术在非空间数据领域的应用,使人们不再局限于通过关系数据表来观察和分析数据信息,还能以更直观的方式看到数据及其结构关系。数据可视化技术的基本思想是将数据库中每一个数据项作为单个图元元素表示,大量的数据集构成数据图像,同时将数据的各个属性值以多维数据的形式表示,可以从不同的维度观察数据,从而对数据进行更深入的观察和分析。信息可视化是在计算机、网络通讯技术支持下,以认知为目的,对非空间的、非数值型的高维信息进行交互式视觉表现的理论、技术与方法。人类任何一个新思想、新技术都是一个逐渐

发展的结果,人们使用可视化的思想由来已久,例如,古代的烽火台就是可用可视方式表达信息。考察这种思想的历史过程,我们可以发现在目的与手段上古今是一致的,即人们想方设法用可视手段表示问题,其目的是为了更清晰地表达问题;工具总是在改进的,人们总是用当时最先进的工具完成这一任务;可视化的对象是不断变化的,什么是信息,不同时代与不同领域有不同理解,这导致可视化的不同层次和分支;而交互式这一特点是直到使用了计算机作为可视化工具时才具有的特性。

因此,如果我们取消工具的限制,增加信息对象的研究范围,则可以将信息可视化看成是以认知为目的的,对信息进行视觉表现的科学。这样的定义存在以下特点:

在当今信息爆炸的时代,没有计算机支撑的可视化系统难以处理足够量的信息,因此一般也不会有太大的实用价值。同时也失去了交互能力,这也是现代信息管理工作所不能接受的,但人们可以借鉴这样一些可视化现象中的思想。不限制工具可以开拓研究的视野,帮助研究人员从这样的系统中获得新的灵感,并通过计算机实现,增加信息可视化的技术形式。事实上,Windows 中资源管理器左边的“文件夹”树就来源于手工的“组织机构图”。信息的非空间性、非数值性和高维性是信息可视化区别于其他可视化分支的标志,因此增加信息对象的研究范围可能会弱化研究重点。但从可视化的总体发展来看各分支最终会汇合到一点上,因此在对信息可视化研究的同时也要关注其他的可视化研究领域。

4 各种可视化技术比较

4.1 数据可视化与科学计算可视化 数据可视化与科学计算可视化的区别是被可视化的对象是物理空间数据还是非物理空间数据。一般来说,如果是物理空间场或工程建筑的空间结构数据,无论是可以看见的还是看不见的,其数据在物理空间上都有一个对应位置。如多块磁铁相互靠近时产生的复杂磁场。为了直观地观察磁场,人们过去将一些细小的铁屑撒在其中,铁屑在磁力的作用下排列出一定的规律,大量的铁屑展示了磁力线的方向,进而帮助人们理解弯曲的磁场(如图 3 所示)。今天,人们根据物理学计算出的磁场数据,在显示器上绘制足够的点,并可形成从任意角度观察所对应的可视化效果,达到物理手段获得不到的效果。这是典型的科学计算可视化的例子。

而数据可视化的数据一般来源于经济、商业、金融等领域。这些数据有具体大小,但它们不对应一个物理空间意义,即不存在一个物理空间场,在某种意义下的数值刚好是该数值。但为了发现其中的规律,数据可视化的方法就是将它们对应到 2 维或 3 维空间中,通过在空间场中对大量数据的展示,帮助人们管理、利用、认知这些数据及其规律。常见的股票走势 K 线图是典型的数据可视化例子(如图 4 所示)。

4.2 信息可视化与数据可视化 信息可视化与数据可视化的区别在于数据是否非数值型。例如有两个表:表 1 是一张数值表,对它进行可视化的时候,无论是在 2 维还是在 3 维空间上布局,其数值是我们可视化的依据。而表 2 中没有数值型数

据,对它进行可视化必须在某种意义下定义相应的数值关系,转换成能够表达论文特征及其关系特征的数值,这个工作是数据可视化所没有的。

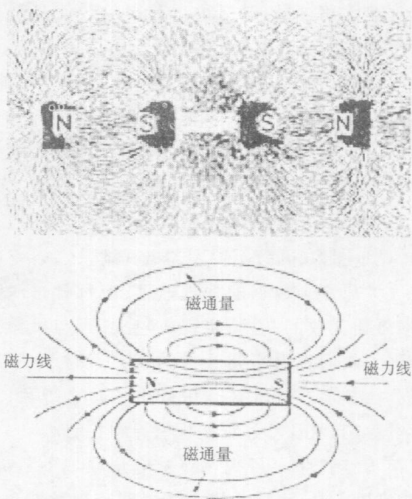


图 3 磁场的可视化表示

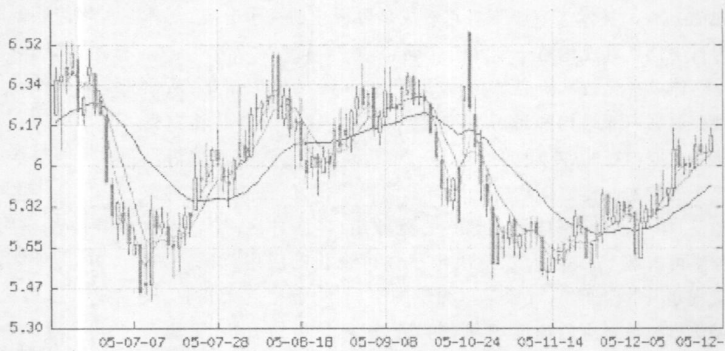


图 4 股票走势 K 线图表示

表 1 数据表(学生成绩表)

学生编号	语文	数学	外语
001	78	97	86
002	82	76	81
⋮			
059	64	75	76

表 2 论文关键词

论文序号	关键词 1	关键词 2	关键词 3
001	信息管理	信息检索	信息可视化
002	模型	UML	可视化
⋮			
678	信息管理	信息经济	科技法

目前,对低维、小规模的数据型数据的可视化方法与技术比较成熟,有较多的应用系统。当数据可视化面对大规模、高维数据时,以前的数据可视化方法渐不实用,其处理方法越来越接近信息可视化。因此,有相当多的研究专题在面对这些情况时不再区分两者的界限。

(下转第 24 页)

函数依赖的概念下改写成  $EA_1$ 、 $EA_2$ 、 $EA_3$  即可,另外增加了下面两个规则

覆盖律( $EA_4$ ) 设  $U$  是类  $C \in CL_S$  的所有属性集 则  $OID \rightarrow oU$ 。

等价律( $EA_5$ ) 设  $U$  是类  $C \in CL_S$  的所有属性集,  $X \subseteq U$ , 若  $X$  是关键字 则  $X \rightarrow oOID$ 。

定义 13 设  $U$  是类  $C \in CL_S$  的所有属性的集合,  $X \subseteq U$ , 属性集  $X$  关于对象函数依赖集  $OF$  的闭包记为  $X^+$ ,  $X^+ = \{Y \mid Y \in U \text{ 且 } X \rightarrow oY\}$  可由扩展的 Armstrong 公理系统导出。

定理 3 扩展的 Armstrong 公理系统是有效的、完备的。  
证明 有效性的证明类似于关系数据库中 Armstrong 公理的证明,这里只就完备性进行证明。设  $X \rightarrow oY$  不能根据扩展的 Armstrong 公理系统从  $OF$  导出,其中  $X \subseteq U$ ,要证它必不为  $OF$  所蕴涵,分两种情况讨论:

- a. 若  $OID \notin X^+$ , 用类似于 Armstrong 公理的完备性证明方法可证得;
- b. 若  $OID \in X^+$ , 由定义 13 知,  $X \rightarrow oOID$  可由扩展的 Armstrong 公理系统导出。根据覆盖律  $EA_4$  得  $OID \rightarrow oU$ , 再由传递律  $EA_3$  得,  $X \rightarrow oU$ 。此时,若  $Y \subseteq U$ ,  $X \rightarrow oY$  成立,若  $Y$  不被  $U$  包含,  $X \rightarrow oY$  必不为  $OF$  所蕴涵。

因此,  $OF$  所蕴涵的对象函数依赖必然可根据扩展的 Armstrong 公理系统从  $OF$  导出,定理得证。

定义 14 设有一个类  $C \in CL_S$ , 若  $C$  中定义了关键字, 则称  $C$  是满足对象第一范式的类, 记作  $C \in 1ONF$ 。

定义 15 设有一个类  $C \in CL_S$ , 若  $C$  中每个非主属性完全对象函数依赖于关键字, 则称  $C$  是满足对象第二范式的类, 记作  $C \in 2ONF$ 。

定义 16 设  $U$  是类  $C \in CL_S$  的所有属性集,  $X \cup Y \subseteq U$ ,

对任一非平凡的依赖  $X \rightarrow oY$ , 若  $X$  含有关键字或  $Y$  是主属性, 则称  $C$  是满足对象第三范式的类, 记作  $C \in 3ONF$ 。

定义 17 设有一个类  $C \in CL_S$ , 若  $C$  中所有非平凡的对象函数依赖中, 对象决定子都含有关键字, 则称  $C$  是满足对象第四范式的类, 记作  $C \in 4ONF$ 。

这一部分只给出了对象  $k$  范式的有关概念, 关于它的求解算法类似于关系数据库中  $k$  范式的求解算法, 此处我们并不给出。

5 结束语

面向对象中的封装、继承、对象标识等概念备受人们的重视, 用对象可以自然、直观地表达工程领域的复杂结构对象, 用操作封装来增强数据处理能力<sup>[5~9]</sup>。这样, 人们开始尝试以面向对象概念为基本出发点来研究和建立数据库系统, 并保持关系数据库管理系统的安全性、完整性、可靠性和 SQL 功能, 将面向对象技术与关系数据库理论相结合, 建立功能更强大的对象—关系数据库。本文先回顾了关系数据库的有关规范化理论, 而后遵循 ODMG—93 的标准, 初步探讨了 ORDB 的范式。

参 考 文 献

1 Rumbaugh J. et al. Object—Oriented Modeling and Design. Prentice Hall, Englewood Cliffs, NJ, 1991  
2 陈伟鹤, 鞠时光. 对象—关系型 DBMS 的研究与开发. 计算机工程与应用, 2000  
3 Marc Gyssens, Dan Suciu, Dirk Van Gucht. Equivalence and Normal Forms for the Restricted and Bounded Fix Point in the Nested Algebra. Dept. WNI, University of Limburg, 1998  
4 马 垣. 关系数据库理论. 北京: 清华大学出版社, 1999  
5 陆洪波. 论面向对象数据库的发展. 科技情报开发与经济, 2000  
6 吴洪森. 面向对象数据库的发展与研究. 计算机工程与应用, 1999

(责编: 梅王京)

(上接第 20 页)

表 3 科学计算可视化、数据可视化、信息可视化的比较

	科学计算可视化	数据可视化	信息可视化
数据来源	空间场、物理场	非空间数值数据集	非空间、非数值大数据集
处理技术	较少, 如体绘制、表面拟合等	很多, 如平行坐标法、散点图、多维堆放等	非常多, 如轮廓图、锥形树、双曲线等
处理过程	预处理、映射、绘制与反馈	预处理、转换、显示、分析	建模、数据预处理、映射、交互绘制、分析评价

它们三者之间的区别可以从表 3 可以看出, 从科学计算可视化到信息可视化的过程中, 被处理的数据的规模不断扩大, 抽象性不断加强; 但人们研究出来的技术却逐渐增多, 用来针对这些越来越有个性的信息集合。同时, 由于信息的非数值性的增强, 处理过程也变得复杂起来, 表现在“绘制”步骤的前后增加了更多的转换过程。

参 考 文 献

1 杨 峰. 信息可视化系统框架与关键技术研究. 武汉大学博士学位论文, 2006

2 Information Visualization and Presentation. <http://www.ischool.berkeley.edu/program/s/courses/247>, 2006.09.16  
3 唐卫清, 刘慎权, 余盛明等. 科学计算可视化. 软件世界, 1996  
4 刘 勤, 周晓峥, 周洞汝. 数据可视化的研究和发展. 计算机工程, 2002; (8)  
5 Eppel M J, RA Burkard. Knowledge Visualization: Towards a New Discipline and its Fields of Application. ICAWorking Paper # 2/2004, University of Lugano, 2004  
6 Chen C M, R J Paul. Visualizing a Knowledge Domain's Intellectual Structure. IEEE Computer, 34(3)  
7 唐伏良, 张向明, 茅及愚等. 科学计算可视化的研究现状和发展趋势. 计算机应用, 1997; (3)  
8 黄河设计公司. 黄河下游交互式三维视景系统. <http://www.yrc-design.com.cn/NumProj/hhxywshj-xtgk.htm>, 2006.02.26  
9 <http://www.yrc-design.com.cn/NumProj/hhxywshj-yyqj.htm>, 2006.02.26  
10 Levoy M, P Hanrahan. Light Field Rendering. In Computer Graphics Proceeding, Annual Conference Series. ACM SIGGRAPH'96, 1996  
11 Rademacher P, G Bishop. Multiple Center of Projection Images. In Computer Graphics Proceeding, Annual Conference Series. ACM SIGGRAPH, 1998  
12 马劲松, 朱大奎. 海岸海洋潮流模拟可视化与虚拟现实建模. <http://www.86vr.com/files/2005-07/6086/Majingsong.doc>, 2006.02.26

(责编: 梅)