

基于特征的时间序列数据场可视化方法

梁训东 刘慎权

(中国科学院计算技术研究所 CAD 开放实验室 北京 100080)

摘要 本文给出了一种在与时间有关的多时间片序列数据场中抽取特征并进行可视化的方法。首先讨论了特征可视化的一般概念,给出了数据场中特征的定义及特征可视化的基本方法,提出了一个在二维数据场中进行特征边界跟踪的方法,采用2个边界算子抽取边界,用树形结构表示特征边界之间的关系,用 Fourier 描述器描述特征边界并重构。最后,讨论了算法的实现及有关问题。

关键词 特征,科学可视化,数据场,算子,计算机图形学。

科学计算可视化是将大批量科学数据映射为可视图象或直观图形符号的方法。这种映射大大方便了对原始数据的分析和解释过程,有利于揭示隐藏在数据中不易发现的复杂物理现象。在许多与时间有关的问题中,人们试图研究物理现象随时间动态变化的过程,但是,求解或测量过程所得到的数据量往往是非常巨大的,例如计算流体力学(CFD)问题有成百上千甚至上万个时间步,而每个时刻都可能有 256^3 或更多的网格数据点,如此巨大的时间序列数据是不可能都存储在局部存储器中的,也不可能交互地存取这些数据。一般地,研究者们感兴趣的往往是那些高层次的物理现象及其随时间的演变规律,一种可行的方法是把着眼点集中于这些“特征”,即将特征从时间序列数据场中抽取出来并跟踪他们的变化规律,抛弃那些不感兴趣的冗余数据,将这些特征存储起来,从而减少数据存储量。这样的可视化处理使交互式的动态显示数据场中的特征量成为可能。

特征是隐含在数据集合中的任何有助于解释相应物理现象的信息。^[1]基于特征的代表方法是一种更简洁、更清晰、更具含义的代替原始数据的高级表示方法,所抽取的特征用于进一步分析及可视化过程。特征可以是满足用户定义的某种条件的数据子集,用这种条件过滤原始数据从而得到某个感兴趣的数据子集。另一类特征是空间域或时间域的一个子域。第3类特征由一些特征模式组成,如临界点及其局部性质,或是一些有意义的物理客体,如涡流或激波。

本文给出了一种在时间序列数据场中抽取特征并使之可视化的方法。在第1节中给出

* 本文研究得到国家自然科学基金、中国科学院院长基金资助,并得到科学与工程计算国家重点实验室支持。作者梁训东,1961年生,博士生,主要研究领域为计算机图形学,科学计算可视化。刘慎权,1930年生,研究员,博士生导师,主要研究领域为计算机图形学,科学计算可视化,动画等。

本文通讯联系人:梁训东,北京 100080,中国科学院计算技术研究所 CAD 开放实验室

本文 1996-03-04 收到修改稿

特征可视化的一般概念,数据场中特征的定义及有关特征可视化方法研究的现状与进展.第2节提出一种二维数据场特征边界跟踪与表示的方法,采用2个边界算子抽取边界,用树形结构表示特征边界之间的关系,用Fourier描述器描述特征边界并重构.最后,讨论了算法的实现与今后的工作.

1 特征可视化的概念及问题

在图象处理与机器人视觉技术中,已有许多自动检测与识别二维图象特征的方法,包括边缘检测、分割、运动图象分析^[2~4]等.理论上,这些方法可以应用到可视化技术中,但在科学计算与数值模拟研究中,由于计算空间的划分方法不同,其数值分布规律不同于二维图象数据,尤其对三维空间或随时间变化的序列数据场,特征的定义与抽取仍是富有挑战性的工作,许多问题仍没有得到很好的解决.^[1]

为了明确表示特征,必须首先定义特征.特征的定义决定了所抽取特征的性质及特征抽取过程中所施加的条件.给定一个数据场(标量场或向量场),一般的简单特征是指数据场中具有特定性质的一个区域.设数据场 \mathcal{V} ,场值为 f ,值域为 \mathcal{F} ,即 $f \in \mathcal{F}$,2种典型的特征描述为:①一个区域 $\mathcal{R} \in \mathcal{V}$, \mathcal{R} 中的所有场值 f 都在指定的范围 \mathcal{S} 内.②一个区域 $\mathcal{R} \in \mathcal{V}$, \mathcal{R} 中的场值 f 都明显不同于其邻近区域中的场值.

第1种情况中,特征直接由其场值来定义.若 \mathcal{S} 不超出 \mathcal{F} 的范围,特征是存在的.第2种情况中,特征不是直接由其场值来严格定义,而是由其区域边缘来确定.区域内部具有不同于场中其它区域中的值,此区域被认为具有特征,且具有明显的边缘.这从2个方面来考虑,把场值有显著变化的区域作为特征区域;把比内部区域更容易检出的边缘作为特征.除了场值之外,其它的属性,如大小、形状、同一空间中共存的各个数据场的相关性等都可作为特征.如何定义感兴趣的特征依赖于特定的应用问题.

从数据场中抽取特征区域后,可以计算出每个特征区域的属性.属性是特征区域的量化参数,可作为特征跟踪时的匹配参数.一些常用的属性有质量、质心、体积、矩、形状参数等.

现有的边缘探测算子如Sobel算子、Laplaceian算子、Marr算子、Haralick离散正交多项式最佳拟合算子都是基于二维图象处理,不适用于结构化网格或非规则网格上的数据场.文献[5]中提出了一个三维边缘算子,用于抽取墨西哥海湾中的涡流并跟踪其随时间的运动,但这种边缘算子只适用于三维网格中,其中二维是均匀分布的情况.

2 二维特征的抽取与表示

为了研究时间序列数据场中的物理量的变化规律,对于不同时刻 t ,抽取感兴趣的特征,问题是如何来表示特征、存储特征、将特征分类以便跟踪它们而实现特征的交互连续显示.^[6~8]这里我们给出了一个完整的特征区域边界跟踪算法,用双向边界算子跟踪特征的边界,用Fourier系数来描述边界的形状,用树形结构表示特征边界之间的相互关系,用Fourier系数来重构特征边界实现特征的连续交互式动态显示.

设二维数据场 \mathcal{V} ,其计算空间中的结构化网格为 $G = \{(x_i, y_j) | i = 1, \dots, M; j = 1, \dots, N\}$.这里 (x_i, y_j) 是网格点 (i, j) 的物理坐标.场值 $f(t, V)$ 在 G 上取值.

2.1 特征边界跟踪

设所抽取的感兴趣的特征区域为 \mathcal{R} , 为讨论方便, 不妨在特征区域中的网格点置标记

1, 非特征区域置标记 0, 即
$$d(i, j) = \begin{cases} 1, & \text{if } (i, j) \in \mathcal{R} \\ 0, & \text{if } (i, j) \notin \mathcal{R} \end{cases}$$

对于每一网格点 (i, j) , $1 \leq i \leq M, 1 \leq j \leq N$, 定义 2 个边界算子, I 向边界算子 $I(i, j)$ 和

J 向边界算子 $J(i, j)$:
$$I(i, j) = d(i+1, j) - d(i, j) \quad J(i, j) = d(i, j+1) - d(i, j)$$

现在来考察 I 向边界算子 $I(i, j)$, $1 < i < M$ 的相邻 2 个网格点 (i, j) 和 $(i+1, j)$ 的值.

4 种可能的取值情况如图 1 所示 (空圈表示 $d(i, j) = 0$; 实圈表示 $d(i, j) = 1$).

对于 (a), (b) 2 种情况, $I(i, j) = 0$, 表示 2 个网格点之间不存在特征边界, (c) 中 $I(i, j) = -1$ 表示存在边界, 我们用向左的向量来表示 $I(i, j)$ 的前进方向, 并称网格点 (i, j) 为它的所属网格点.

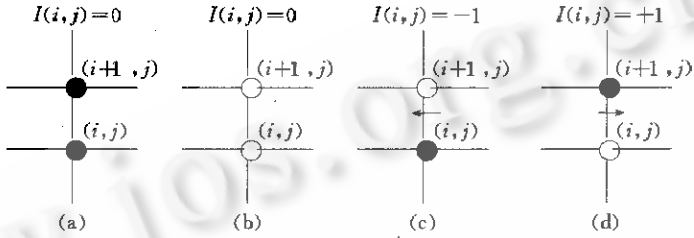


图1 I 向边界算子的图示表示

(d) 中 $I(i, j) = 1$ 表示存在边界, 我们用向右的向量来表示, 并称 $(i+1, j)$ 为它的所属网格点.

对于 J 向边界算子 $J(i, j)$ 也可类似地表示出来 (如图 2 所示).

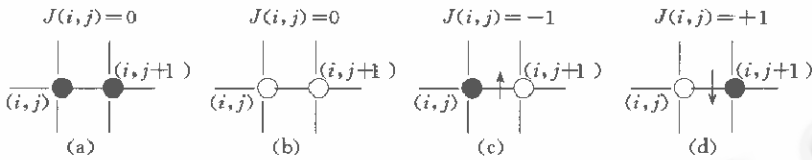


图2 J 向边界算子的图示表示

现在来分析 $I(i, j) = 1$ 时, 沿其前进方向 (箭头方向) 上相邻 2 个网格点的值的情况, 总共有 4 种可能 (如图 3 所示).

在 (a), (b), (c) 中, 除 $I(i, j)$ 外, 都有另外一个非零边界算子, 分别为 $J(i, j), J(i+1, j), I(i, j+1)$, 称它们为 $I(i, j) = 1$ 的非零向前边界算子. 在 (d) 中, 除 $I(i, j)$ 外有另外 3 个非零边界算子 $I(i, j+1) = -1, J(i, j) = +1, J(i+1, j) = -1$. 沿 $I(i, j)$ 的前进方向, 有 2 种选择, 选 $J(i+1, j)$ 为 $I(i, j)$ 之后继边界算子, 这相当于选择了四连通; 选 $J(i, j) = 1$ 为 $I(i, j)$ 之后继边界算子, 这相当于选择了八连通.

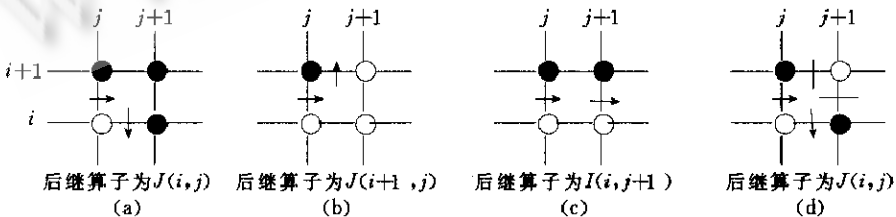


图3 I 向边界算子的后继算子

类似地, 我们可以求出 $I(i, j) = -1, J(i, j) = 1, J(i, j) = -1$ 的非零后继边界算子.

由上面的定义可知, 对任何非零边界算子, 有且仅有一个非零后继边界算子. 设非零边界算子 b 为非零边界算子 a 的后继算子, 则称 a 为算子 b 的前趋边界算子. 由此我们得到以

下结论:①对任何非零边界算子,有且仅有一个前趋边界算子一个后继边界算子.②从任一非零边界算子出发,沿其后继边界算子的方向跟踪,必然回到其本身,从而形成一条以非零边界算子构成的回路区域围线.不难证明以下定理:

定理 1. 所有的以边界算子构成的特征边界线之间,或者一条包含另一条,或者两者互不包含,不可能发生两者相交的情况.

定理 2. 对于任一特征区域,所有的外围边界线都是逆时针方向跟踪的而所有的内围线都是顺时针方向跟踪的.

于是我们得到基于边界算子的特征边界线跟踪算法:任取一未标记过的非零边界算子,按后继算子进行跟踪,得到特征边界线.同时将此边界线上的边界算子都标记为“用到过的”,重复上述过程直到没有未标记过的非零边界算子为止.如果将非零边界算子组成的边界线中的非零边界算子换以它所属的网格点,则可得到以网格点组成的特征边界线.

数据场内可能存在若干互不包含于其它特征边界线之内的特征区域边界线,且由定理 1 可知,它们之间是互不相交的,称它们为第 1 层特征边界线.其它的特征边界线围成的特征区域必包含于某 1 条第 1 层边界线之中.由此,假想有 1 条包含所有边界线的边界线,并以包含关系为父子关系,则整个场中的特征边界线构成以上述假想边界线为根结点,以其它边界线以及包含关系为父子关系的树.

若要找出场中所有的特征边界线,对场按 $i=1,2,\dots,M; j=1,2,\dots,N$ 的顺序扫描边界算子,执行前面的算法,并建立特征边界树,容易得到以下结论:①若边界 C_j 被边界 C_i 所包围,则 C_j 一定在 C_i 之后被检出.②边界线树中,偶数层上(第 1 层为根)的边界一定为特征区域的外边界;奇数层上边界线一定为特征区域的内边界(除根外).

2.2 特征边界的 Fourier 描述与生成

2.2.1 Fourier 描述器

假设特征边界曲线 γ 的参数表示形式是 $(x(l), y(l))=z(l)$, 式中 l 表示边界线的弧长, $0 \leq l \leq L$, L 是特征边界线的全长. 设 $\theta(l)$ 是边界线为 l 的点的切线方向, $\delta_0 = \theta(0)$ 是初始点 $z(0)$ 的切向, $\varphi(l)$ 表示角积累函数,是沿边界线从初始点切线方向到弧长为 l 的点的切线的旋转角度. $\varphi(l)$ 的变化规律描述了特征边界线 γ 的形状.^[9] 令 $t = \frac{Lt}{2\pi}$, 注意到 $l \in [0, L]$, 则 t 的取值范围为: $t \in [0, 2\pi]$, 定义

$$\varphi^*(t) = \varphi\left(\frac{Lt}{2\pi}\right) + t, t \in [0, 2\pi] \tag{1}$$

$\varphi^*(t)$ 是 $[0, 2\pi]$ 上的周期函数, 且 $\varphi^*(0) = \varphi^*(2\pi) = 0$

将 $\varphi^*(t)$ 在 $[0, 2\pi]$ 上展开成 Fourier 级数

$$\varphi^*(t) = \mu_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt) \tag{2}$$

其中 $\mu_0 = \frac{1}{2\pi} \int_0^{2\pi} \varphi^*(t) dt, a_n = \frac{1}{\pi} \int_0^{2\pi} \varphi^*(t) \cos ntdt, b_n = \frac{1}{\pi} \int_0^{2\pi} \varphi^*(t) \sin ntdt$. 其相位振幅形式为

$$\varphi^*(t) = \mu_0 + \sum_{n=1}^{\infty} (A_n \cos nt - \alpha_n) \tag{3}$$

式中 $A_n = (a_n^2 + b_n^2)^{1/2}, \alpha_n = \text{tg}^{-1}\left(\frac{b_n}{a_n}\right)$. $\{A_n, \alpha_n\}_1^{\infty}$ 是 $\{a_n, b_n\}_1^{\infty}$ 的极坐标, A_n 与 α_n 称为边界线 γ 的描述器. 不难证明, 特征边界线可根据下式重构.

$$z(l) = z(0) + \frac{L}{2\pi} \int_0^{\frac{2\pi l}{L}} e^{[-t + \delta_0 + \mu_0 + \sum_{k=1}^N A_k \cos(kt - \alpha_k)]} dt \quad (4)$$

式中 δ_0 是初始点 $z(0)$ 对应的初始角.

2.2.2 Fourier 描述器的系数计算

假设我们得到的边界线由 m 个点 v_0, v_1, \dots, v_{m-1} 组成, 边 (v_{i-1}, v_i) 的长度是 Δl_i , 在点 v_i 上的变化角度是 $\Delta \varphi_i$, 且 $L = \sum_{i=1}^m \Delta l_i$, 由 2.1 节中得到的边界点列很容易地计算出 $\Delta l_i, \Delta \varphi_i$.

对于 $\sum_{i=1}^k \Delta l_i \leq l \leq \sum_{i=1}^{k+1} \Delta l_i$, 有 $\varphi(l) = \sum_{i=1}^k \Delta \varphi_i$ 且对 $0 \leq l \leq \Delta l_1$, 有 $\varphi(l) = 0$.

由公式(2), 可以推出
$$\mu_0 = -\pi - \frac{1}{L} \sum_{k=1}^m l_k \Delta \varphi_k \quad (5)$$

$$a_n = \frac{-1}{n\pi} \sum_{k=1}^m \Delta \varphi_k \sin \frac{2\pi n l_k}{L} \quad (6)$$

$$b_n = \frac{1}{n\pi} \sum_{k=1}^m \Delta \varphi_k \cos \frac{2\pi n l_k}{L} \quad (7)$$

至此, 我们可以用 Fourier 系数 $\{a_n, b_n\}$ 或 $\{A_n, \alpha_n\}$ 来表示、存储 2.1 节中得到的特征边界线, 用重构公式(4)来显示特征区域.

特征边界的 Fourier 描述不仅提供了单个时间片上数据场内特征的表示方法与存储方法, 而且为时间序列数据场的特征跟踪提供了特征的匹配参数(跟踪方法另文给出), 因为 $\{A_n, \alpha_n\}$ 具有位移、旋转、大小、起点等不变的性质, 对于相邻 2 个时间片上的 2 个对应的特征边界线 γ, γ' , 它们的 Fourier 描述器 A_n 也相似.

3 实现

本文提出的方法已在 SGI Indigo² 工作站上实现, 图 4 是一个驱动正方腔内流动非稳定性的数值模拟结果, 对 140 个时间片对应的序列数据场进行了可视化处理, 图中只给出了 6 个时刻 ($t=1, 20, 35, 60, 75, 110$) 的涡度的变化, 分别对应于图 4 中的 (a), (b), (c), (d), (e), (f). 采用 2.1 节的边界跟踪方法得到各特征区域边界的点列, 用 2.2.2 节中的方法计算 Fourier 系数, 本例中, 取 $N=12$. 一般地, 当 $N=7$ 时, 所生成的边界就能基本正确地反映特征区域边界的形状. N 越大, 边界越精确. 用公式(4)生成边界, 根据定理 2 确定边界所围成的特征区域并着色.

在特征抽取并跟踪之后, 只需存储特征参数及其属性, 动态显示过程中不需存取原始数据场中的数据, 动态显示的速度可以自动计算或由交互界面进行控制.

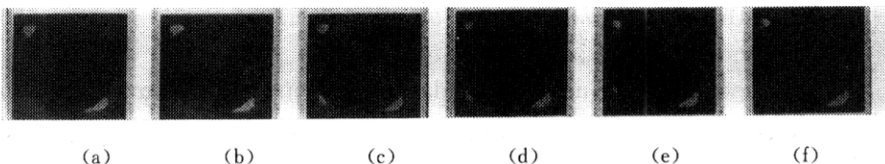


图 4 驱动正方腔内流动非稳定性的数值模拟结果

4 结 语

对与时间有关的多时间片序列数据场,特征可视化方法是一种有效的方法,它只抽取感兴趣的特征数据并进行可视化,减少了数据存储量,使大批量数据的交互式动态显示成为可能.但还有许多需要解决的问题.数据场对应的领域不同,就有不同的特征与特征属性,很难有统一的定义方法.第1.1节中的Fourier描述器 $\{A_n, \alpha_n\}_N^N$ 当 N 很小时,用重构公式(4)显示的特征区域存在误差,但这对研究特征的变化规律并不十分重要.

本文提出的方法作为我们开发的数据场可视化系统的一部分.通过系统提供的友好的用户界面,可以清楚地显示出序列数据场中特征的变化过程,使大批量数据的交互式动态显示变为现实.

参 考 文 献

- 1 Hesselink L, Post F H. Research issues in vector and tensor field visualization. IEEE Computer Graphics and Application, March 1994. 76~79.
- 2 Gonzalez R C, Woods R E. Digital image processing. New York: Addison-Wesley, 1992.
- 3 钟玉琢等. 机器人视觉技术. 北京:国防工业出版社,1994.
- 4 徐建华. 图象处理与分析. 北京:科学出版社,1992.
- 5 Moorhead R J, Zhu Z. Feature extraction for oceanographic data using a 3D edge operator. IEEE Visualization'93 Proceedings, 1993. 402~405.
- 6 Carlbom I, Chakravarty I, Hsu W. Integrating computer graphics, computer vision, and image processing in scientific applications. Computer Graphics, 1992, 26(1):9~16.
- 7 Villasenor J, Alain Vincent. An algorithm for space recognition and time tracking of vorticity tubes in turbulence. CVGIP: Image Understanding; 1992, 55(1):27~35.
- 8 Samtaney R *et al.* Visualizing features and tracking their evolution. Computer, July 1994. 20~27.
- 9 Zahn C T, Roskies R Z. Fourier descriptors for plane closed curves. IEEE Transactions on Computers, 1972, c-21(3):269~281.

A FEATURE VISUALIZATION METHOD FOR TIME-DEPENDENT DATA FIELD

Liang Xundong Liu Shenquan

(CAD Laboratory Institute of Computing Technology The Chinese Academy of Sciences Beijing 100080)

Abstract This paper presents a feature extraction and visualization method for time-dependent data field. Feature definition and attributes are discussed. Then it gives a border tracking method using two edge operators to track the border of feature area. The relationship of the feature is represented by Fourier descriptors. The implementation and problems of this method are also discussed.

Key words Feature, scientific visualization, data field, operator, computer graphics.