

3-15-2024

## 3D Streamline Visualization Method Based on Clustering Fusion

Xuqiang Shao

*Department of Computer Science, North China Electric Power University, Baoding 071003, China;  
Engineering Research Center of Intelligent Computing for Complex Energy Systems, Baoding 071003,  
China, shaoxuqiang@163.com*

Ya Cheng

*Department of Computer Science, North China Electric Power University, Baoding 071003, China*

Yizhong Jin

*Department of Computer Science, North China Electric Power University, Baoding 071003, China*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the Artificial Intelligence and Robotics Commons, Computer Engineering Commons, Numerical Analysis and Scientific Computing Commons, Operations Research, Systems Engineering and Industrial Engineering Commons, and the Systems Science Commons

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact [xtfzxb@126.com](mailto:xtfzxb@126.com).

---

## 3D Streamline Visualization Method Based on Clustering Fusion

### Abstract

**Abstract:** In order to solve the problems of incomplete feature extraction, continuity destruction of flow field by visual results, and poor representation of streamline caused by unstable clustering division when the clustering method is used to realize 3D streamline visualization. A 3D streamline visualization method based on clustering fusion is proposed. It consists of a distance measurement method between features and a clustering fusion method, which takes the inter-feature distance and spatial distance as the similarity between streamlines for clustering and then performs weighted merging and subdivision of the obtained clustering result. The method has been tested on data sets with different features and compared qualitatively and quantitatively with the existing methods. The results show that compared with the existing methods, the proposed method can better balance the relationship between feature extraction and streamline distribution, and the stability of clustering division is improved by 2%~5%. The accuracy of vector field reconstruction is improved by 3%~5%.

### Keywords

flow field visualization, streamline visualization, clustering fusion, feature extraction, streamline selection

### Recommended Citation

Shao Xuqiang, Cheng Ya, Jin Yizhong. 3D Streamline Visualization Method Based on Clustering Fusion[J]. Journal of System Simulation, 2024, 36(3): 625-635.

# 基于聚类融合的三维流线可视化方法

邵绪强<sup>1,2</sup>, 程雅<sup>1</sup>, 金佚钟<sup>1</sup>

(1. 华北电力大学 计算机系, 河北 保定 071003; 2. 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003)

**摘要:** 为解决使用聚类方法实现三维流线可视化时, 存在特征提取不全面、可视结果破坏流场连续性、聚类簇划分不稳定导致流线代表性差等问题, 提出了基于聚类融合的三维流线可视化方法。该方法由特征间距离度量方法和聚类融合方法两部分组成, 将特征间距离和空间距离分别作为流线间的相似度进行聚类, 对得到的聚类结果进行加权合并后再划分。将该方法在具有多个不同特征的数据集上进行了实验, 并与现有方法进行了定性、定量比较。结果表明, 与现有方法相比, 该方法能够较好地平衡特征提取和流线分布之间的关系, 聚类簇划分的稳定性提高了2%~5%, 矢量场重构的精度提高了3%~5%。

**关键词:** 流场可视化; 流线可视化; 聚类融合; 特征提取; 流线选择

中图分类号: TP391.41 文献标志码: A 文章编号: 1004-731X(2024)03-0625-11

DOI: 10.16182/j.issn1004731x.joss.22-1257

**引用格式:** 邵绪强, 程雅, 金佚钟. 基于聚类融合的三维流线可视化方法[J]. 系统仿真学报, 2024, 36(3): 625-635.

**Reference format:** Shao Xuqiang, Cheng Ya, Jin Yizhong. 3D Streamline Visualization Method Based on Clustering Fusion[J]. Journal of System Simulation, 2024, 36(3): 625-635.

## 3D Streamline Visualization Method Based on Clustering Fusion

Shao Xuqiang<sup>1,2</sup>, Cheng Ya<sup>1</sup>, Jin Yizhong<sup>1</sup>

(1. Department of Computer Science, North China Electric Power University, Baoding 071003, China;

2. Engineering Research Center of Intelligent Computing for Complex Energy Systems, Baoding 071003, China)

**Abstract:** In order to solve the problems of incomplete feature extraction, continuity destruction of flow field by visual results, and poor representation of streamline caused by unstable clustering division when the clustering method is used to realize 3D streamline visualization. A 3D streamline visualization method based on clustering fusion is proposed. It consists of a distance measurement method between features and a clustering fusion method, which takes the inter-feature distance and spatial distance as the similarity between streamlines for clustering and then performs weighted merging and subdivision of the obtained clustering result. The method has been tested on data sets with different features and compared qualitatively and quantitatively with the existing methods. The results show that compared with the existing methods, the proposed method can better balance the relationship between feature extraction and streamline distribution, and the stability of clustering division is improved by 2%~5%. The accuracy of vector filed reconstruction is improved by 3%~5%.

**Keywords:** flow filed visualization; streamline visualization; clustering fusion; feature extraction; streamline selection

收稿日期: 2022-10-20 修回日期: 2022-12-29

基金项目: 国家自然科学基金(61502168); 河北省自然科学基金(F2020502014); 中央高校基本科研业务费专项(2021MS095)

第一作者: 邵绪强(1982-), 男, 副教授, 博士, 研究方向为计算机图形学、虚拟现实。E-mail: shaouxuqiang@163.com

## 0 引言

流场可视化是科学可视化的重要研究内容，被广泛地应用在医学、航空航天等领域<sup>[1]</sup>。流线因其几何结构简单、绘制方便已成为可视化流场的重要手段。对于大型复杂的3D流场，密集粒子轨迹或场线的存在容易导致视觉遮挡和渲染混乱。过滤流线以突出特征区域可减少遮挡和混乱，但会破坏流场连续性。因此，选择代表性流线实现视觉感知与信息传递之间的平衡在流场探索过程中至关重要。

基于聚类的三维流线可视化方法依赖于相似性度量的选择。目前，相似性度量主要分为两类<sup>[2]</sup>：一类是根据流线的空间位置，使用欧式距离、豪斯多夫距离<sup>[3]</sup>、最近点距离的平均值<sup>[4]</sup>等空间距离，作为相似性度量执行簇划分，该方法局限于将空间距离较近的流线作为相似流线聚类到同一簇中，空间距离很远但形状相似的流线不会划分到同一簇中<sup>[2]</sup>；另一类是将流线形状特征<sup>[2]</sup>，如曲率、扭转率、信息熵等特征间距离作为相似性度量执行簇划分，该类方法能将具有相似特征的流线划分到同一个簇中，但存在特征提取不全面、丢失流场连续性的问题。此外，不同聚类方法的簇划分策略不同，同一条流线在不同聚类方法划分中所属的簇可能不同，提取代表性流线可视化时，流线代表性差。

为解决上述问题，本文提出聚类融合方法实现三维流线可视化。首先研究了特征间距离计算方法，将特征矩阵间的Procrustes距离<sup>[5]</sup>作为流线特征间距离。在此基础上，提出聚类融合方法，分别将空间距离和特征间距离作为流线相似性度量方法，产生两个基聚类，加权融合基聚类结果，再执行Ncut(normalized cut)划分得到最终的聚类结果，在提取流场特征的同时保证了其连续性。最后，将本文方法应用于不同特征的数据集和城市风场，验证了本文方法的有效性。

## 1 相关工作

### 1.1 流线相似度计算方法相关研究

许多流线相似性度量的现有工作是流线间的逐点距离。文献[4]使用MCP(mean of closest point)作为相似性度量，通过分层流线捆绑可视化实现在不同细节级别的探索。文献[6]使用两阶段k-means算法对流线进行聚类：第一阶段用流线的起点、中点和终点计算它们的MCP距离进行聚类；第二阶段通过计算每个簇中流线的线性和角度熵进一步细分。文献[3]使用豪斯多夫距离作为流线之间的相似度距离进行聚类。

基于特征的相似度是将特征对之间的距离作为2条流线之间的距离。文献[7]使用特征向量来测量流线相似度，使用曲率、扭转、曲折度和速度方向熵等特性描述流线，特征向量间的加权曼哈顿距离作为流线间相似性。文献[5]将流线建模为字符串，根据缠绕角度对流线进行重新采样，捕获相邻采样点之间流线段相对简单模式，获得相似性度量。文献[8]对用于流可视化的几种曲线聚类方法进行了深入的比较研究，为用户提供了选择特定方法的系统指南。文献[9]介绍了FlowNet，该方法基于自编码器和卷积神经网络，在单个框架内学习流线和流表面的潜在特征，生成特征描述符。对特征描述符采用t-SNE降维，选择簇中的代表。文献[10]将脑纤维束用矢量场表示，用流线简化算法进行压缩，在流线归一化和规则多面体投影后，计算每个纤维束的高维特征，用t-SNE降维特征后将其输入自编码器中训练，但是该聚类算法受限于流线简化程度，且只适用于单个数据集，没有验证其他数据集。

### 1.2 聚类融合方法的相关应用

对于某个大规模流线数据集，使用不同的聚类方法或使用同一聚类方法结合不同的流线间距离计算方法，产生的聚类簇不完全相同，任何一个单一的算法不能达到普遍适用的效果。聚类融

合方法可以缓解该问题, 主要做法是将不同方法生成的流线聚类簇组合后按照一定的策略重新划分, 其目标是通过聚类生成的聚类簇进行融合得到鲁棒性更高、效果更好的聚类结果。据本文调研, 目前聚类融合算法的应用集中在文本数据, 尚未将其应用于流线选择方面的研究。

聚类融合方法近期的研究详见文献[11-12]。在聚类融合算法中, 生成基聚类后, 根据集成方法所依赖的信息矩阵的不同, 可以将信息矩阵的构建分成4种: 基于特征的方法、基于样本与类簇的方法、基于共联矩阵的方法、基于类簇相交的方法。但上述方法存在算法复杂度高、基聚类质量对最终聚类集成结果的影响等问题。

## 2 基于聚类融合的三维流线可视化方法概述

本文整体流程如图1所示。首先, 分别产生基于空间位置和基于形状特征两个基聚类。为全面考虑流场的空间分布和形状特征, 一方面, 使用文献[8]推荐的k-means+MCP算法产生基于空间位置的基聚类, 但由于传统的MCP算法需要比较流线中每个点与目标流线中每个点间的距离, 计算速度慢、效率低, 因此本文使用改进的MCP算法, 选取流线的起点、中点和终点作为流线间距离比较的对象<sup>[6]</sup>, 能保证计算流线间准确距离的同

时加快计算速度; 另一方面, 提出了一种特征间距离度量方法, 将流线划分为有限结构后, 对有限结构提取三维特征和二维特征, 将有限结构特征与点坐标一起组成特征矩阵。矩阵间的Procrustes距离为流线特征间距离, 以该距离作为相似度执行特征相关的聚类, 聚类方法选择排名前三的凝聚层次聚类(agglomerative hierarchical clustering, AHC)<sup>[8]</sup>。

然后, 对得到的2个基聚类结果进行加权融合再执行Ncut图划分, 得到最终的聚类结果。下面从特征间距离度量方法和基聚类融合算法2个过程对本文方法进行论述。

## 3 特征间距离度量方法

### 3.1 数据预处理

本文对文献[10]的流线简化方法进行了改进, 主要思想是保留流线的第一个和最后一个节点以保证流线的完整性, 除此之外, 只需保留能代表流线局部变化的关键节点, 忽略不重要的节点。

在流线中, 2个相邻流线段向量间的夹角反映流线的局部变化。当向量间夹角大于阈值 $\theta$ 时, 局部流线变化较为剧烈。将流线段间的公共节点作为关键节点进行保留。当夹角较小时, 局部流线变化缓慢, 此时可以忽略中间节点。

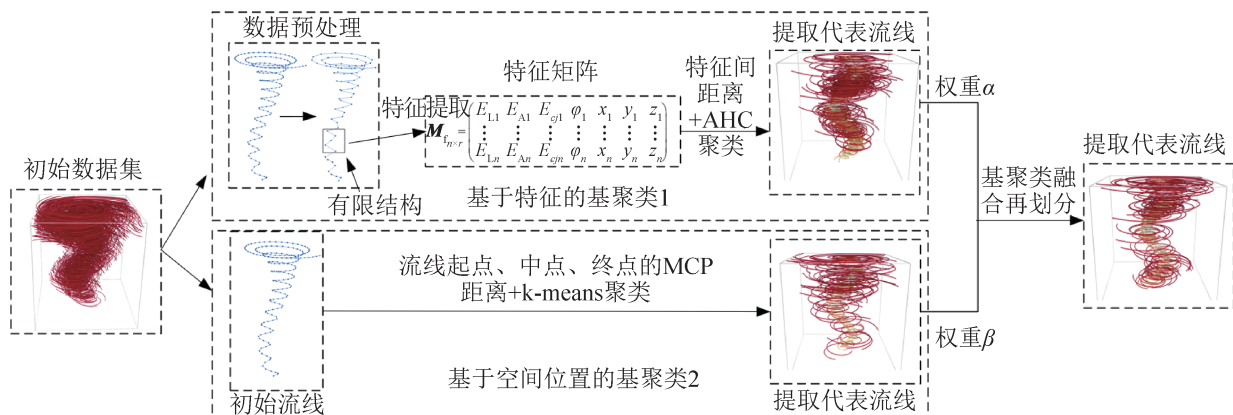


图1 本文方法整体流程

Fig. 1 Overall process of proposed method

为减少流线简化的误差，将每两个流线段向量之间的角度进行累加，作为累计值 $\omega$ ，当 $\omega < 2\theta$ 时继续执行当前操作，当累计值 $\omega \geq 2\theta$ 时，累计值清0，同时保存当前节点为关键节点；如果当前两个流线段之间的角度值符合阈值条件，则将流线段的中间节点放入关键点序列中，累计值清0。

图2(a)和(b)是tornado数据集单条流线简化前后的对比，图2(c)和(d)是tornado数据集简化前后的对比。通过观察图2(a)和(b)可以看到，在简化前单条流线中点的分布较为密集，点的数目也较多；简化后去掉了对流线特征影响较小的点，从外形看单条流线简化前后其流动特征基本没有改变。由图2(c)和(d)可以观察到，在简化前后，整个数据集的流动特征也基本没有改变，只是减少了每条流线中点的数目。因此，预处理方法可以有效保留数据集的流动特征，同时减少流线中点的数目。但由于不同的数据集特征不同，因此在阈值选取时也不同，本文所用到的数据集在进行简化时角度阈值 $\theta$ 设置如表1所示。

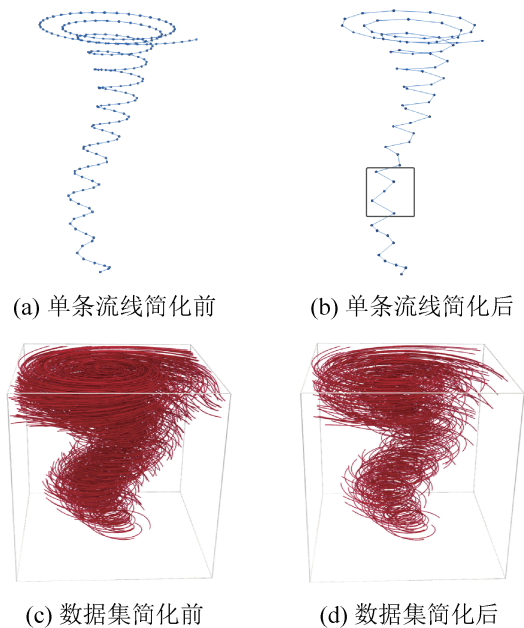


图2 tornado数据集简化前后流线对比  
Fig. 2 Comparison of streamlines before and after simplification of tornado data set

表1 不同数据集流线简化阈值设置

Table 1 Streamline simplified threshold settings for different data sets

数据集	vortex_ring	tornado	crayfish	Bernard
$\theta$	5	18	25	30

### 3.2 有限结构提取

简化后的流线，保留了流线中剧烈变化的部分，但简化后的流线中2个点之间的距离相较于初始流线更长，特征变化也更大，因此，对简化后流线的特征提取应尽量详尽。受文献[5, 13]的启发，本文将流线划分为多个有限结构，提取有限结构的二维和三维特征。

有限结构提取的主要方法是，在简化后的流线中随机选取 $n$ 个点作为中心点，分别向前向后各选取2个点，提取由5个特征点组成的 $n$ 个有限结构，选取5个点作为有限结构中特征点的数目是因为：首先，有限结构中的流线段应该足够简单；其次，最少需要3个点来确定局部流动行为，所以在确定中心点之后，将该中心点看作是2个局部流动行为间的公共点，分别向前和向后选取2个点构成有限结构，有限结构中的特征变化规律代表着流线的局部流动趋势。如选取图2(b)标记区域中的5个点作为一个有限结构。一方面，该有限结构能捕获流线的局部流动行为；另一方面，该局部范围内的特征既不太复杂也不太简单，便于后续提取该有限结构的二维和三维特征。

### 3.3 特征提取

#### 3.3.1 三维特征提取

对于流线的有限结构，其三维形状特征包括线性熵 $E_L$ 和角度熵 $E_A$ 。

线性熵量化了沿流线的流动幅度中的信息量，计算公式为

$$E_L = -\frac{1}{\ln(m+1)} \sum_{w=0}^m \frac{L_w}{L_s} \ln \frac{D_w}{L_s} \quad (1)$$

式中： $m$ 为流线的段数； $L_w$ 是第 $w$ 段的长度； $L_s$ 为有限结构中流线段的总长度。

角度熵量化了沿着流线的角度变化量, 计算公式为

$$E_A = -\frac{1}{\ln u} \sum_{w=0}^{u-1} \frac{A_w}{A_L} \ln \frac{A_w}{A_L} \quad (2)$$

式中:  $u$  为流线段数目;  $A_w$  为第  $w$  个流线段之间角度的绝对值;  $A_L$  为沿着流线的总角度变化(有限结构中角度的绝对值之和)。

### 3.3.2 二维特征提取

受文献[10]等二维特征提取的启发, 本文将流线分别投影到  $XOY$ 、 $YOZ$ 、 $ZOX$  平面, 然后分别提取有限结构在这三个平面上的线性率和变化角度。

#### (1) 流线的线性率

流线的线性率用于表征流线的单调性, 将投影后的二维子结构的每个点与原点相连, 点到坐标原点之间的距离记为  $d_p$ , 如图 3(a) 所示。计算流线的对数增益率  $I_p$ , 将流线转换到一个 0 和 1 组成的序列(0,1)中, 具体转换方法为<sup>[14]</sup>

$$I_p = \begin{cases} 1, & \ln(d_p - d_{p-1}) > 0 \\ 0, & \ln(d_p - d_{p-1}) \leq 0 \end{cases} \quad (3)$$

根据流线的(0,1)序列计算流线的线性率  $r_L$ , 计算方法为

$$r_L = \frac{\sum_{p=1}^{l-1} I_p I_{p+1} + (1 - I_p)(1 - I_{p+1})}{l - \sum_{p=1}^{l-1} I_p I_{p+1} + (1 - I_p)(1 - I_{p+1})} \quad (4)$$

式中:  $p \in [0, l]$ ,  $l$  代表序列中元素的个数, 若  $I_i > 0$  则表示流线在第  $p$  个节点处单调递增, 反之, 则流线单调递减。

当流线是明显的单调趋势时,  $r_L$  远离 1, 若流线是非单调的, 则  $r_L$  接近 1。

#### (2) 有限结构的峰值计算

有限结构的峰值  $\varphi$  显示了流线的缠绕状态。将有限结构中的第一个点  $n_b$  和最后一个点  $n_c$  连起来作直线  $l$ , 找到与该直线  $l$  距离最大的节点  $n_m$  作为峰值节点, 距离  $h$  表示流线的弯曲变化范围, 峰值  $\varphi$  是由第一个节点  $n_b$ 、峰值节点  $n_m$  和最后一

个点  $n_c$  组成的三角形的顶点角度, 如图 3(b) 所示。

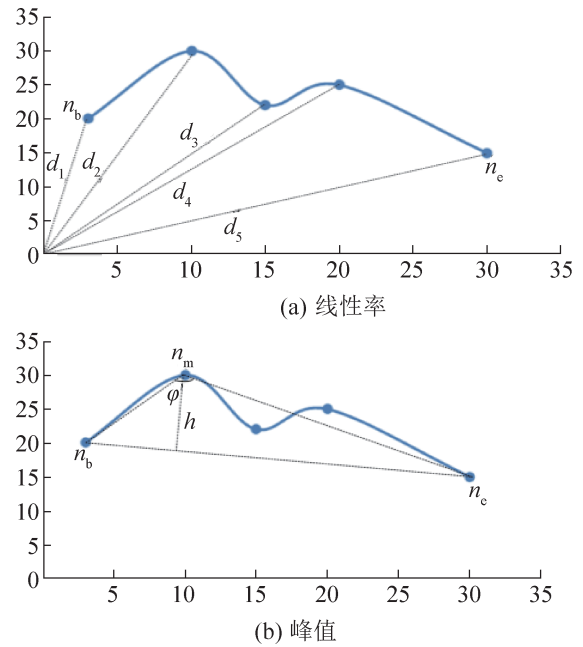


图 3 子结构的二维特征  
Fig. 3 2D features of substructure

在有限结构特征提取完成之后, 将每个有限结构的所有特征与子结构中的点坐标放入流线的特征矩阵中。其中特征矩阵的行数目代表每条流线划分的有限结构个数, 列包括子结构的三维特征、二维特征线以及子结构中的点坐标, 组成流线的特征矩阵  $M_{n \times r}$ , 表示为

$$M_{n \times r} = \begin{bmatrix} E_{L_1} & E_{A_1} & N_{c_j-1} & \varphi_1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ E_{L_n} & E_{A_n} & N_{c_j-n} & \varphi_n & x_n & y_n & z_n \end{bmatrix} \quad (5)$$

式中:  $n$  为子结构的数目;  $r$  为流线中提取的特征及个数。

将特征矩阵之间的 Procrustes 距离<sup>[5]</sup>作为流线间距离, 结合 AHC 聚类方法产生基于形状属性的基聚类。一条流线的特征矩阵到另一条流线特征矩阵的最佳平移  $T$ 、旋转  $R$  和缩放  $s$ , 即特征矩阵之间的最小距离  $d$ , 表示为

$$d = \min \left\{ \sum_{i=1}^r |M_{b_i} - M'_{a_i}|^2, \sum_{i=1}^r |M_{a_i} - M'_{b_i}|^2 \right\} \quad (6)$$

式中:  $M'_{a_i} = sRM_{a_i} + T$ ,  $M_{a_i} \in M_a$  中的列向量;

$M_{b_j}$ 同理； $d$ 为流线特征矩阵 $M_a$ 和流线特征矩阵 $M_b$ 之间的Procrustes距离。

## 4 基聚类融合算法

### 4.1 构建加权的共联矩阵

设初始数据集为 $S=\{s_1, s_2, \dots, s_n\}$ ，其中， $s_i$ 表示的是数据集中的第 $i$ 条流线。根据基聚类的结果 $\Pi=\{\Pi^1, \Pi^2\}$ ，将基聚类的结果转换为相似度矩阵 $M_k$ ， $1 \leq k \leq 2$ 。根据流线是否在同一个聚类簇中确定该聚类的相似度矩阵 $M_k$ 。计算为

$$\text{Con}(s_i, s_j) = \begin{cases} 1, & C(s_i) = C(s_j) \\ 0, & \text{else} \end{cases} \quad (7)$$

式中： $C(s_i)$ 为 $s_i$ 所在的簇，若 $s_i$ 与 $s_j$ 位于同一个簇，则相似度矩阵 $M_k$ 中对应的第 $i$ 行第 $j$ 列的值为 $\text{Con}(s_i, s_j)=1$ ，否则为0。通过此方法确定2个基聚类的相似度矩阵 $M_{k1}$ 与 $M_{k2}$ 。2次基聚类产生的相似度矩阵进行加权得到最终的共联矩阵 $M_{n \times n} = \alpha M_{k1} + \beta M_{k2}$ ，其中 $\alpha, \beta$ 为不同的权重。

### 4.2 Ncut划分

将加权后的共联矩阵 $M_{n \times n}$ 看作一个 $n$ 个点的无向图，矩阵上的值就是无向图边的权值。使用归一化割Ncut方法将该图划分 $k$ 个分区 $(A_1, A_2, \dots, A_k)$ ，得到最终的划分结果。

首先，将共联矩阵 $M_{n \times n}$ 作为邻接矩阵 $A_{n \times n}$ ，邻接矩阵每行的和作为度矩阵 $D_{n \times n}$ 的对角线元素；通过邻接矩阵和度矩阵计算得到拉普拉斯矩阵 $L$ ，最后对 $L$ 进行特征分解并进行后处理，将特征值按照从小到大的顺序排列，取前 $k$ 个特征值，并计算其对应的特征向量 $u_1, u_2, \dots, u_k$ ，将上述的 $k$ 个列向量组成矩阵 $U=\{u_1, u_2, \dots, u_k\}$ ， $U \in R^{n \times k}$ ，随后使用传统的AHC聚类方法以相对应的加权距离作为相似度量聚类，得到最终的聚类结果 $C_1, C_2, \dots, C_k$ 。

这里使用AHC方法作为后处理方法的主要原

因是：①传统的层次聚类算法能确定最终的聚类簇数；②层次聚类方法相对于Ncut常用的k-means聚类，稳定性和健壮性更好。

## 5 实验结果分析与讨论

本文方法允许用户根据自身需求，交互调整相似度矩阵的权重。图4列举了不同权重下聚类融合的结果，从图4(a)~4(e)特征和距离基聚类的相似度矩阵权重取值分别为(1,0)、(0.8,0.2)、(0.5,0.5)、(0.2,0.8)、(0,1)，其中，4(a)是使用基于流线特征聚类提取到的代表性流线；4(e)是基于空间位置聚类提取到的代表性流线；4(b)~4(d)是不同权重下本文聚类融合方法提取到的代表性流线。根据速度大小对数据集进行着色，蓝、黄、红对应速度分别从低到高。聚类簇数设置如下：tornado<sup>[8]</sup>、vortex\_ring<sup>[8]</sup>设聚类簇数为60，crayfish<sup>[8]</sup>设置为150，并选择每一簇的聚类中心作为代表性流线；Bernard<sup>[8]</sup>设为100，选择聚类中心和距离聚类中心最远的流线作为代表性流线，数据集从上到下依次是：vortex\_ring、tornado、crayfish、Bernard。

由图4可知，当基于特征聚类的权重较高时，此时的可视结果对特征区域有较好的呈现，如图4 vortex\_ring中的旋涡结构和crayfish数据集中标记处；当基于空间距离聚类的权重较高时，此时的可视结果更多地体现了流线的空间分布。

但本文方法也有一定的缺点，若想得到较为满意的可视结果，需要不断调整权重进行多次聚类对比，耗时较长。

### 5.1 定性对比

下面将本文方法与其他方法进行对比，对比方法为文献[9]、从文献[8]中选取的3个具有代表性的聚类方法+相似性度量方法的组合：AHC+dP、AHC+dG、k-means+MCP；dP为普氏距离，dG为几何距离。对比结果如图5所示。



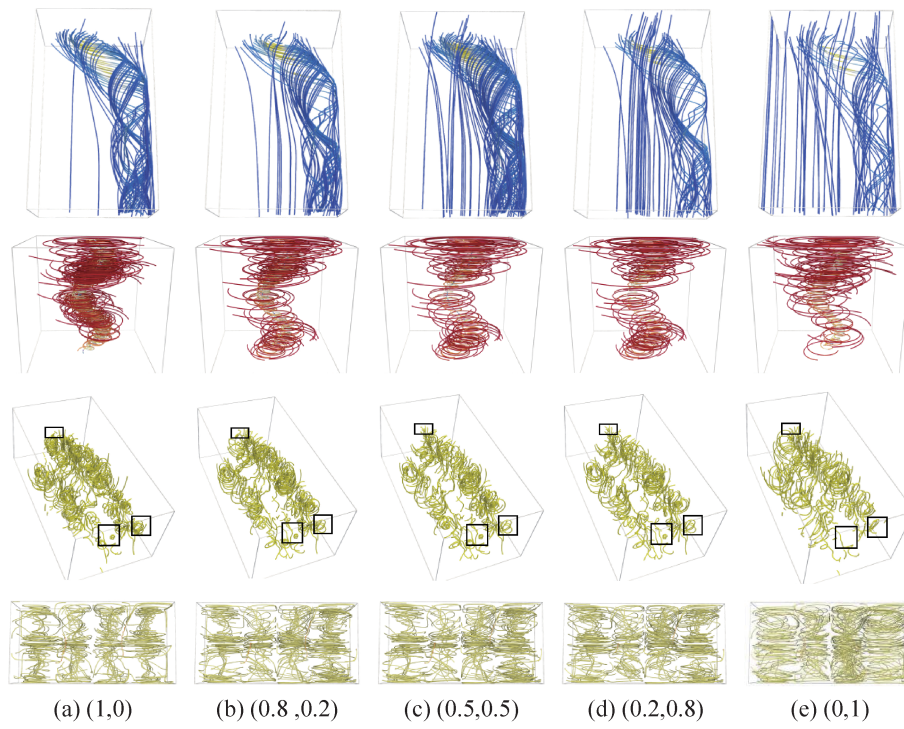


图 4 不同权重下的可视结果  
Fig. 4 Visual results under different weights

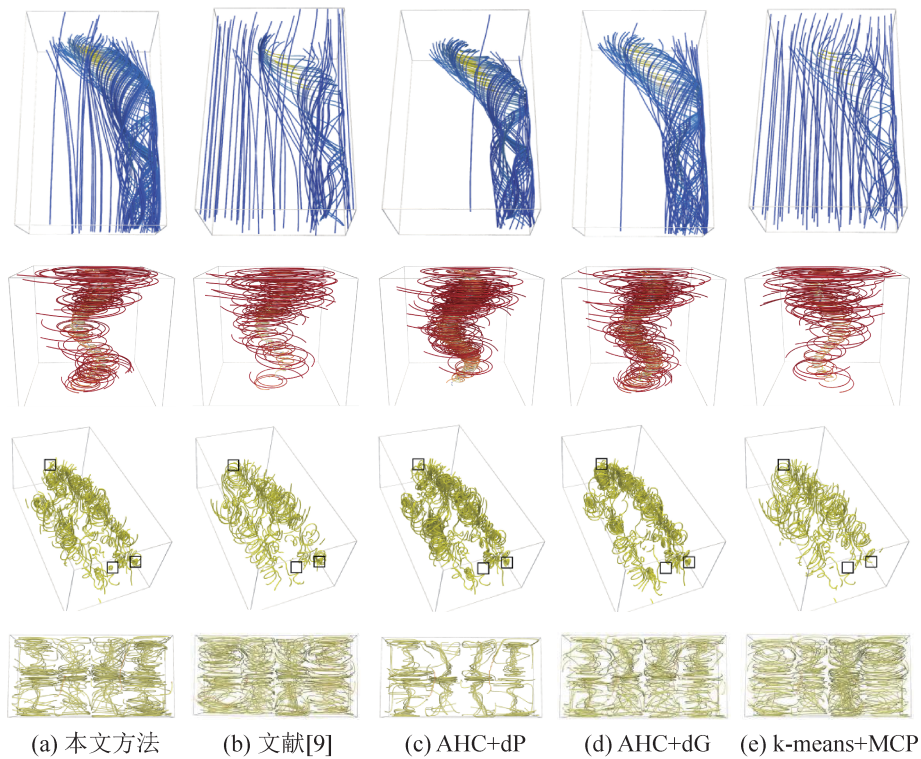


图 5 不同方法对比  
Fig. 5 Comparison of different methods

对于 vortex\_ring 数据集, 设置权重为(0.5, 0.5), 由可视结果可知, 本文方法提取到的流线主要分布在涡环区域及涡环周围的次流动区域, 相比于 AHC+dP 和 AHC+dG, 本文方法保留了特征周围的流动信息, 可视化结果具有较好的连续性; 与文献[9]的方法相比, 本文方法减少了上下文流线对特征区域的遮挡。与 k-means 方法相比, 本文方法提取的特征明显, 流线分布均匀。

对于 tornado 数据集, 设置权重为(0.8, 0.2), 由可视结果可知, 本文方法能够在流线空间分布和特征呈现两方面达到较好的平衡。与 AHC+dP 和 AHC+dG 相比, 本文方法很大程度上减少了流场中心漩涡区域的遮挡现象, 可以轻松地观察到流场内部的流动结构, 如漩涡最内部的蓝色部分; 与文献[9]和 k-means 方法相比, 本文方法提取的特征更完全, 有助于用户掌握整体流动行为。

对于 crayfish 数据集, 设置权重为(0.8, 0.2), 相对于 AHC+dP 和 AHC+dG, 本文方法减少了流线间遮挡。与文献[9]和 k-means 相比, 本文方法提取到更多特征, 如图 5 中标记处。

对于 Bernard 数据集, 设置权重为(0.5, 0.5), 本文方法提取到涡环内部区域, 减少了流线间的遮挡。文献[9]和 k-means 方法提取的流线分布较均匀, 但未能提取到涡环内部的流动结构。AHC+dP 方法虽然提取到较多涡环内部流动结构, 但其

丢失了流场的整体流动趋势。AHC+dG 方法和本文方法一样, 能提取涡环内部的流线且流线分布均匀。

由上述定性的可视化结果分析可知。本文方法在某些权重下, 对流场的特征提取和流线分布能达到较好的平衡状态。

## 5.2 聚类融合方法的定量评价

为了定量评价聚类融合结果的有效性, 计算本文方法产生聚类簇的簇间距离和簇内距离, 聚类评价结果能反映聚类的有效性。本文将 Silhouette<sup>[15]</sup> 作为聚类簇划分的评价指标。Silhouette 是聚类的非线性组合度量, 当簇密度较高且分离较大时, Silhouette 越大, 其计算方法为

$$S(i) = \frac{d_o(i) - d_1(i)}{\max\{d_o(i), d_1(i)\}} \quad (8)$$

式中:  $i$  为第  $i$  条流线;  $d_1(i)$  为簇中的第  $i$  条流线与同簇中其他流线距离的平均值;  $d_o(i)$  为簇中的第  $i$  条流线与距离最近的聚类簇中所有样本点的平均距离。当簇密度较高且分离较大时, 聚类的轮廓系数也越大。

为验证本文方法聚类簇划分的鲁棒性和健壮性, 计算多个数据集上不同聚类簇数的 Silhouette, 如表 2 所示。文献[9]的工作是定性的, 本文仅对其进行定性评价, 在定量方面不做评价<sup>[2]</sup>。

表 2 Silhouette 评价指标  
Table 2 Silhouette evaluation metrics

数据集	流线条数	聚类簇数	本文方法	AHC+dP	AHC+dG	k-means + MCP
vortex_ring	1 000	60	<b>0.752 351</b>	0.716 551	0.366 684	0.326 068
		45	<b>0.738 484</b>	0.706 709	0.452 228	0.333 323
		100	<b>0.741 136</b>	0.715 703	0.358 113	0.300 165
tornado	600	60	<b>0.747 280</b>	0.736 079	0.365 365	0.229 683
		50	<b>0.799 293</b>	0.760 512	0.330 180	0.220 296
		70	<b>0.781 450</b>	0.763 441	0.354 938	0.216 340
crayfish	1 000	100	<b>0.725 225</b>	0.702 365	0.237 904	0.202 035
		150	<b>0.693 944</b>	0.671 816	0.202 287	0.254 255
		180	<b>0.741 987</b>	0.666 802	0.270 784	0.196 223
Bernard	1 000	100	<b>0.363 009</b>	0.297 490	0.223 755	0.213 607
		80	<b>0.450 095</b>	0.373 900	0.163 908	0.219 926
		70	<b>0.562 473</b>	0.388 772	0.161 330	0.190 137

<http://www.china-simulation.com>

由表2的统计结果可知, 在不同数据集及同一数据集的不同簇数下, 本文方法的Silhouette指标优于其他方法, 表明本文方法结果在簇稳定性和健壮性方面均优于单一聚类算法。

为了客观评价本文方法提取流线的代表性, 使用矢量场重构前后的均方误差(mean square error, MSE)来衡量。表3是将图5中不同方法提取的代表性流线进行重采样体素化, 使用三角剖分技术产生重构矢量场 $V_a - V_c$ , 并比较它们与原矢量场 $V_0$ 之间的差距, 其中数据集1, 2, 3, 4分别代表: vortex\_ring、tornado、crayfish、Bernard。

表3 矢量场重构MSE指标

数据集	图5(a)	图5(b)	图5(c)	图5(d)	图5(e)
1	<b>0.004 7</b>	0.004 8	0.005 2	0.005 0	0.014 1
2	<b>0.004 4</b>	0.004 6	0.005 2	0.005 1	0.014 5
3	<b>0.003 4</b>	0.003 6	0.004 5	0.004 3	0.013 7
4	<b>0.004 5</b>	0.0048	0.005 3	0.004 6	0.012 9

重构后的矢量场与原始矢量场间的MSE表示提取流线的代表性。MSE越小代表重构后矢量场与原始矢量场差距越小, 提取的流线代表性越好。由表3可知, 本文方法在不同数据中取得了较为优秀的效果, 说明本文方法提取的流线代表性好。

### 5.3 本文方法在城市风场环境中的应用

可视化城市环境中的风场对研究城市环境通风<sup>[16]</sup>、城市环境街道设计<sup>[17]</sup>具有一定的现实意义。

本文使用的风场数据是香港荃湾真实城市地区的风流。主要做法是从GIS(geographic information system)数据中得到荃湾城市拓扑结构, 用OpenFoam的pimpleFoam求解器在非结构化网格上模拟得到风场数据, 对模拟结果进行规则化重采样, 将其转换为具有 $200 \times 180 \times 40$ 个网格点的常规网格体积。图6显示了用于仿真的城市环境, 在城市环境的左侧设置了一个沿x轴方向的风, 如蓝色箭头所示。在该场景内使用均匀播种方法结合四阶龙格库塔积分生成初始流线, 将该流线作为聚类的初始数据集, 整个城市环境风场中共

播种1137条流线, 选取代表性流线120条。



图6 荃湾城市环境

Fig. 6 Urban environment of Tsuen Wan

城市环境中, 在不同建筑配置的影响下, 建筑物周围气流模式的变化, 导致建筑物前后风速的变化<sup>[18]</sup>。当风被建筑物阻挡时, 建筑物背面会出现文丘里效应<sup>[18]</sup>。文丘里效应驱动回流产生螺旋上升气流, 这也是城市环境风场中的大部分重要流动区域。图7是在2 m、10 m、20 m和30 m的情况下对荃湾的城市风场的二维截面进行流线可视化。流线按照速度大小着色, 其中, 红、黄、蓝分别对应速度由高到低, 标记区域是文丘里效应的体现。对于图7(a)和(b)中的标记区域5和6, 在高度为20 m时, 该漩涡周围的流线速度明显低于周围流线, 见图7(c)中的相应标记区域; 图7(a)和(b)中的标记区域1-4, 则在高度为30 m时能观察到该区域的速度值小于周围流线速度, 见图7(d)中的相应标记区域。以上二维截面结果证明了荃湾城市环境存在具有文丘里效应的漩涡结构, 这些结构是城市环境流场可视化的重要区域。

图8是不同方法在城市风场环境中的可视化结果。本文方法与图8(b)相比提取到更多具有文丘里效应的漩涡区域, 如标记区域1、2、3; 与图8(c)相比本文方法流线分布更均匀, 如图8(c)在标记区域1附近几乎没有流线, 不能覆盖整个流场, 标记区域2只有少量流线分布; 与图8(d)相比能提取到更多特征区域, 如标记区域1、2、6, 在标记区域5和7本文方法提取的特征更完整; 与图8(e)相比, 本文方法在2、4、5、6、7区域能提取到具有文丘里效应的漩涡结构, 流线分布也相对均匀。

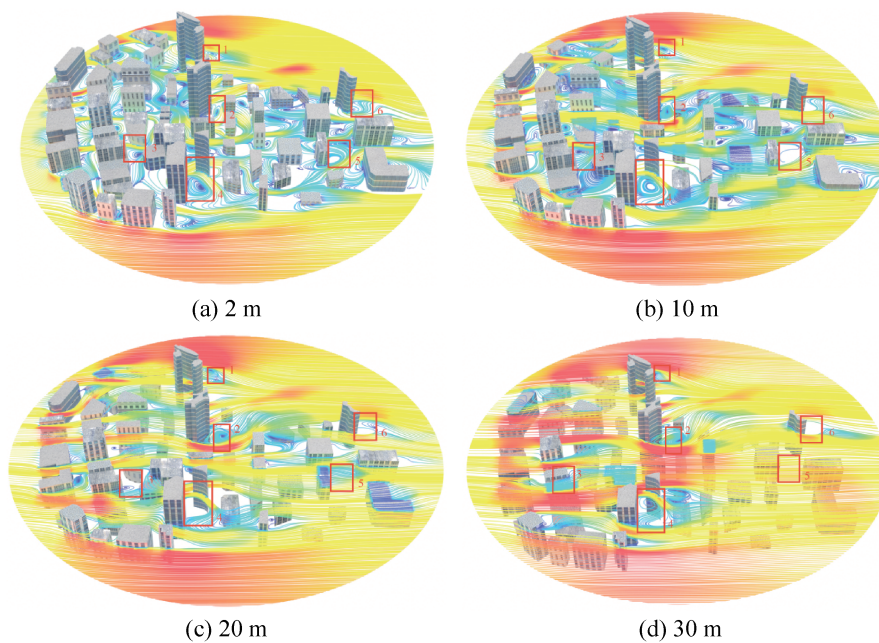


图 7 不同高度下城市环境风场流线截面图  
Fig. 7 Streamline section of urban environmental wind field at different heights

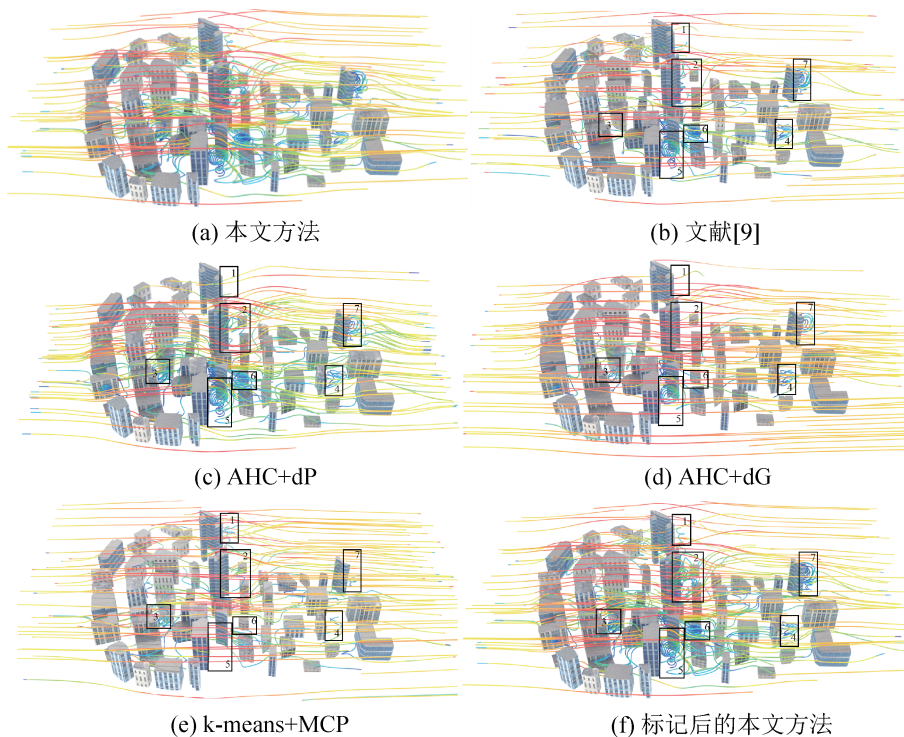


图 8 城市风场不同流线选择方法的可视结果  
Fig. 8 Visual results of different streamline selection methods for urban wind fields

## 6 结论

本文提出了一种基于聚类融合的三维流线可视化方法，能从大量流线中选择代表性流线，在

展现流场特征的同时保证连续性。此外，提出了特征间距离计算方法，将其与 AHC 结合得到的聚类结果和 k-means+MCP 得到的聚类结果，加权融

合再划分得到最终的聚类结果。将本文方法与其他方法在具有不同特征的数据集上进行了定性、定量的比较, 结果均表现良好, 证明本文方法在流场探索方面具有一定的价值。此外, 将本文方法应用于真实城市风场环境, 也获得了良好的可视效果, 表明本文方法在研究城市环境可视化方面有一定的现实意义和参考价值。

### 参考文献:

- [1] 宋汉戈, 刘世光. 三维流场可视化综述[J]. 系统仿真学报, 2016, 28(9): 1929-1936.  
Song Hange, Liu Shiguang. Review of 3D Flow Visualization[J]. Journal of System Simulation, 2016, 28(9): 1929-1936.
- [2] Sane S, Bujack R, Christoph Garth, et al. A Survey of Seed Placement and Streamline Selection Techniques[J]. Computer Graphics Forum, 2020, 39(3): 785-809.
- [3] Christian Rössl, Holger Theisel. Streamline Embedding for 3D Vector Field Exploration[J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(3): 407-420.
- [4] Yu Hongfeng, Wang Chaoli, Shene C K, et al. Hierarchical Streamline Bundles[J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(8): 1353-1367.
- [5] Tao Jun, Wang Chaoli, Shene C K, et al. A Vocabulary Approach to Partial Streamline Matching and Exploratory Flow Visualization[J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(5): 1503-1516.
- [6] Chen Chengkai, Yan Shi, Yu Hongfeng, et al. An Illustrative Visualization Framework for 3D Vector Fields [J]. Computer Graphics Forum, 2011, 30(7): 1941-1951.
- [7] Li Yifei, Wang Chaoli, Shene C K. Streamline Similarity Analysis Using Bag-of-features[C]//Proceedings Volume 9017, Visualization and Data Analysis 2014. Bellingham, WA, USA: SPIE, 2014: 90170N.
- [8] Shi Lieyu, Laramée R S, Chen Guoning. Integral Curve Clustering and Simplification for Flow Visualization: A Comparative Evaluation[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 27(3): 1967-1985.
- [9] Han Jun, Tao Jun, Wang Chaoli. FlowNet: A Deep Learning Framework for Clustering and Selection of Streamlines and Stream Surfaces[J]. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(4): 1732-1744.
- [10] Xu Chaoqing, Sun Guodao, Liang Ronghua, et al. Vector Field Streamline Clustering Framework for Brain Fiber Tract Segmentation[J]. IEEE Transactions on Cognitive and Developmental Systems, 2022, 14(3): 1066-1081.
- [11] Zhou Peng, Du Liang, Liu Xinwang, et al. Self-paced Clustering Ensemble[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(4): 1497-1511.
- [12] Huang Dong, Wang Changdong, Lai Jianhuang. Locally Weighted Ensemble Clustering[J]. IEEE Transactions on Cybernetics, 2018, 48(5): 1460-1473.
- [13] Liu Shiguang, Song Hange. Streamline Querying Based on Finite Substructures[J]. Journal of Visualization, 2019, 22(3): 571-585.
- [14] Chen Gang. Deep Learning with Nonparametric Clustering[EB/OL]. (2015-01-13) [2022-09-11]. <https://arxiv.org/abs/1501.03084>.
- [15] Steffen Oeltze, Dirk J Lehmann, Alexander Kuhn, et al. Blood Flow Clustering and Applications in Virtual Stenting of Intracranial Aneurysms[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(5): 686-701.
- [16] Qin Hao, Lin Pingying, Stephen Siu Yu Lau, et al. Influence of Site and Tower Types on Urban Natural Ventilation Performance in High-rise High-density Urban Environment[J]. Building and Environment, 2020, 179: 106960.
- [17] Guo Fei, Zhu Peisheng, Wang Shiyuan, et al. Improving Natural Ventilation Performance in a High-density Urban District: A Building Morphology Method[J]. Procedia Engineering, 2017, 205: 952-958.
- [18] Zhang Shaowei, Kwok K C S, Liu Haihong, et al. A CFD Study of Wind Assessment in Urban Topology with Complex Wind Flow[J]. Sustainable Cities and Society, 2021, 71: 103006.