

文章编号: 1673-6338(2020)06-0622-06

# 顾及运动特征的滑动窗口轨迹数据压缩改进算法

张奇, 成毅, 徐立, 葛文

(信息工程大学, 河南 郑州 450001)

**摘要:** 滑动窗口轨迹数据压缩算法使用垂直欧式距离作为轨迹点能否被压缩的判断依据, 且能够针对实时数据进行在线压缩, 在轨迹数据处理中应用广泛。但是, 同目前大多数轨迹压缩算法一样, 其将轨迹数据的几何形态特征作为判断特征点是否应当被压缩的唯一依据。除几何特征外, 轨迹数据中还包含了速度、时间间隔等运动特征信息, 且轨迹主体的运动状态对轨迹数据几何形态有很大影响。针对该问题, 在滑动窗口压缩算法的基础上, 提出一种顾及时间同步欧式距离和运动速度的轨迹数据压缩算法, 通过运动特征的加入更加准确地捕捉轨迹特征点, 实现更高精度的数据压缩。实验结果表明, 较现有的滑动窗口轨迹数据压缩算法, 本文算法能够有效减小压缩误差, 提高压缩准确率。

**关键词:** 轨迹数据压缩; 滑动窗口算法; 垂直欧式距离; 同步欧式距离; 运动速度

中图分类号: P208

文献标识码: A

DOI 编码: 10.3969/j.issn.1673-6338.2020.06.012

## Improved Sliding Window Trajectory Compression Algorithm Considering Motion Characteristics

ZHANG Qi, CHENG Yi, XU Li, GE Wen

(Information Engineering University, Zhengzhou 450001, China)

**Abstract:** In the sliding window track data compression algorithm, the vertical Euclidean distance is used as the basis to judge whether the track points can be compressed and the real-time data can be compressed online, so it is widely used in trajectory data processing. However, like most current trajectory compression algorithms, the geometric features of trajectory data are taken as the only basis to judge whether feature points should be compressed. In addition to the geometric features, the trajectory data also contains motion characteristics such as velocity and time interval, and the motion state of the main body of the trajectory has a great impact on the geometry of the trajectory data. In order to solve this problem, based on the sliding window compression algorithm, a track data compression algorithm considering the synchronous Euclidean distance and motion speed is proposed in the paper. By adding the motion features, the track feature points can be captured more accurately, and the data compression with higher accuracy can be realized. Experimental results show that, compared with the existing sliding window track data compression algorithm, the improved algorithm can effectively reduce the compression error and improve the compression accuracy.

**Key words:** trajectory data compression; sliding window algorithm; vertical Euclidean distance; Synchronous Euclidean Distance (SED); velocity

海量、大比例尺的全球导航卫星系统 GNSS (Global Navigation Satellite System) 轨迹能够揭示城市动态以及和人类行为相关的隐含事实<sup>[1]</sup>。但轨迹数据具有来源广、种类多、复杂度高、数据量大的特点, 数据的高量级是轨迹数据研究和应用中需要面对和解决的首要难题。为了在尽可能保证压缩后轨迹与原始轨迹的相似度的条件下, 尽可能减少轨迹数据的数据量<sup>[2]</sup>, 就需要对轨迹

进行压缩处理。

当前, 轨迹数据压缩算法可分为离线压缩技术和在线压缩技术两类<sup>[3]</sup>。离线压缩技术是指线下对一条完整的、首尾轨迹点已经确定的轨迹进行压缩, 删除轨迹中冗余和重复的轨迹点。由于离线压缩技术充分考虑到轨迹的全局特征, 因此能保证压缩后轨迹相对于原始轨迹的准确性。但由于只有在明确轨迹的起始轨迹点和终止

收稿日期: 2020-09-18; 修回日期: 2020-11-20。

基金项目: 国家自然科学基金项目(41901397)。

作者简介: 张奇(1996-), 男, 湖南岳阳人, 硕士生, 主要研究方向为地理信息服务。E-mail: xgdxzq@163.com

轨迹点后才能进行压缩,因此该方法只适用于明确已知的历史轨迹数据,适合线下分析历史轨迹数据的情形。最具有代表性的离线压缩技术是 Douglas-Peucker 算法<sup>[4-5]</sup>。在此算法的基础上,许多学者还提出了各种改进算法。文献[6]提出了一种动态 Douglas-Peucker 算法,在算法基础上引入时间维度,通过时空三维空间来对动态数据进行压缩。文献[7]考虑到了轨迹点含有的速度信息,并在此基础上构造伪图,通过求解最短路径的方法来简化轨迹。文献[8]提出了快速 Douglas-Peucker 算法,将考虑时间维的轨迹压缩算法对应到向量空间中,利用向量的内积和外积的几何性质,对算法压缩效率和效果进行了改善。在线压缩技术具有能够压缩处理实时轨迹数据的特点,从轨迹数据的起始轨迹点出发,轨迹数据的终止点可进行动态增加。同离线压缩技术相比,其不需要完整的事先明确起终点的轨迹数据,在进行压缩时仅需要轨迹数据的起始轨迹点即可,因此更加灵活,适用于动态实时的轨迹压缩情景。由于在线压缩技术是从局部出发,因而能够做到局部占优;但由于不能兼顾轨迹数据整体的趋势,因此全局优势性相对较弱。在线压缩技术方面,也有学者提出了许多经典的算法。文献[9]提出了滑动窗口 (Sliding Window) 算法和开放窗口 (Open Window) 算法,利用窗口内轨迹点的累计误差来进行轨迹点的选取。文献[10]提出了 Threshold-Guided Sampling 算法和 STTrac 算法,通过速度和方向特征来选取轨迹点。文献[11]通过对轨迹点进行分级并赋值权重来选取轨迹点,并调整压缩比和压缩误差,得到了更好的压缩效果。无论是离线压缩技术还是在线压缩技术,都是以轨迹中间点与首尾节点之间的垂直距离作为依据进行取舍判断,其中涉及到大量的距离运算。为了提高算法运算效率,文献[12]提出了改进的滑动窗口轨迹数据压缩算法,将最大偏移距离作为待压缩轨迹点能否被压缩的判据,降低了运算次数,提高了运算效率,但本质上并没有改变以垂直距离作为判断依据的基本原理。

轨迹数据不是普通的几何线段,是对轨迹主体运动过程的客观记录,其中包含了运动速度、时间采样间隔等运动特征信息<sup>[13-15]</sup>。本文认为,运动特征信息对于准确描述轨迹形态具有重要的参考价值,因此在滑动窗口压缩算法的基础上,采用同步欧式距离来代替传统的垂直欧式距离,并利用速度差来实现动态阈值设定的算法思路。

## 1 相关概念

### 1.1 同步欧式距离

定义 1 同步欧式距离。指原始轨迹数据中的轨迹点与其在近似轨迹段上按其时间比例所求得的对应位置之间的欧氏距离<sup>[16]</sup>。如图 1 所示,  $P_{cur}P_{per}$  表示轨迹点  $P_{cur}$  对应的垂直欧式距离,  $P_{cur}P_{syn}$  表示轨迹点  $P_{cur}$  对应的同步欧式距离,其意义是轨迹点  $P_{cur}$  在其相应的近似轨迹段上按时间比例的对应位置。

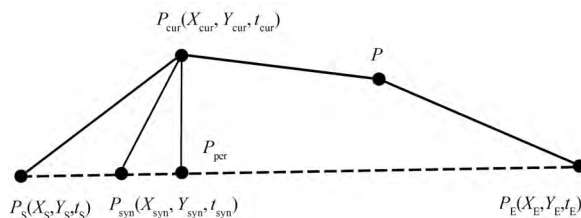


图 1 同步欧式距离

因此,同步欧式距离的计算过程如下:首先,

计算点  $P_{per}$  对应的时间比例  $\frac{t_{syn}-t_S}{t_E-t_S}$ , 其中,  $P_{per}$  点对应的时间为  $t_{cur}$ ,  $t_{syn}=t_{cur}$ ,  $P_S$  与  $P_E$  为该段轨迹的始末位置点,  $t_S$  与  $t_E$  分别为其对应的时间;然后,按照时间比例计算出  $P_{per}$  在相应的近似轨迹段上的对应位置  $P_{syn}$  点;最后采用空间距离的计算公式,利用两点的位置坐标,计算出同步欧氏距离。

$P_{syn}(X_{syn}, Y_{syn}, t_{syn})$  的坐标计算公式为:

$$X_{syn} = X_S + \frac{t_{syn}-t_S}{t_E-t_S} (X_E - X_S), (t_S < t_{syn} < t_E); \quad (1)$$

$$Y_{syn} = Y_S + \frac{t_{syn}-t_S}{t_E-t_S} (Y_E - Y_S), (t_S < t_{syn} < t_E). \quad (2)$$

因此,同步欧式距离可通过下式计算:

$$D(P_{cur}, P_{syn}) = \sqrt{(X_{cur} - X_{syn})^2 + (Y_{cur} - Y_{syn})^2}. \quad (3)$$

### 1.2 算法性能度量指标

对于轨迹数据压缩算法而言,衡量一种算法性能的指标主要有压缩时间、压缩率和压缩误差<sup>[17]</sup>。现如今,计算速度和计算时间已经不再是制约轨迹压缩算法研究的核心问题。压缩的精度主要由压缩率和压缩误差这两种指标来表征,以此来衡量算法的性能。

定义 2 压缩时间  $T$ 。指完成压缩所需要的时间。设压缩前的时间为  $T_1$ , 压缩后的时间为  $T_2$ , 则

$$T = T_2 - T_1. \quad (4)$$

定义 3 压缩率 (Compression Ratio)。指压缩后数据较原始数据所节省的存储空间大小与原始数据所占存储空间大小的比值<sup>[18]</sup>。设压缩前

的轨迹点数为  $S_1$ , 压缩后的轨迹点数为  $S_2$ , 则压缩率计算公式为

$$R = 1 - \frac{S_2}{S_1}. \quad (5)$$

在保证压缩精度的前提下, 压缩率越高, 说明压缩后得到的轨迹点数越少, 压缩效果也就越好。

**定义4** 平均同步欧氏距离误差。指原始轨迹中被舍弃的轨迹点与还原轨迹中对应位置的平均误差, 即非特征轨迹点到它相邻的轨迹特征点之间连线的同步欧式距离, 反映了还原轨迹和原始轨迹之间的差距<sup>[19]</sup>。本文使用同步欧式距离来代替垂直欧式距离, 因此使用平均同步欧氏距离误差  $\bar{D}$  来代替传统的压缩误差。计算公式为

$$\bar{D} = \frac{1}{S_1 - S_2} \sum_{i=1}^{S_1 - S_2} D(P_i). \quad (6)$$

式中,  $D(P_i)$  是第  $i$  个非特征轨迹点的同步欧式距离。

## 2 算法描述

### 2.1 经典滑动窗口轨迹压缩算法

经典滑动窗口轨迹压缩算法采用分段逐步压缩的思想。确立初始窗口后, 根据轨迹点的特征与性质, 对同步传输的数据进行实时的压缩。同离线算法相比, 这种算法不需要明确轨迹段的终点, 可以有效提高处理的效率。如设定的初始窗口为  $(P_1, P_2, P_3)$ , 该起始窗口中  $P_1$  为起点,  $P_2$  为待压缩点,  $P_3$  为终点。计算  $P_2$  到  $P_1, P_3$  之间连线的垂直欧式距离, 并与设定的距离阈值作比较, 若小于阈值, 则将  $P_2$  舍去, 终点更新为  $P_4$ , 待压缩的点更新为  $P_3$ , 窗口更新为  $(P_1, P_3, P_4)$ 。此时再计算  $P_3$  到  $P_1, P_4$  之间连线的垂直欧式距离, 若大于阈值, 则  $P_3$  保留, 窗口向后滑动, 并以  $P_3$  作为滑动窗口的起点, 此时窗口更新为  $(P_3, P_4, P_5)$ 。继续对后续轨迹点作如上判断, 直到轨迹段更新的最终点。图2为滑动窗口轨迹压缩算法原理。

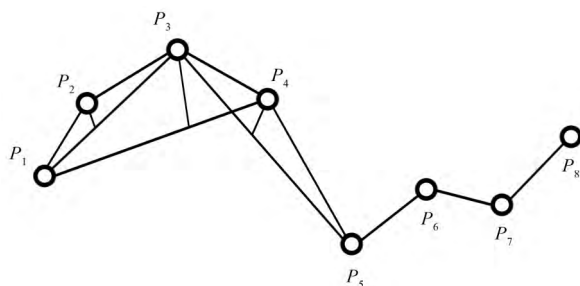


图2 滑动窗口轨迹压缩算法原理

### 2.2 算法改进思路

经典的轨迹压缩算法, 往往都只考虑了轨迹数据所包含的空间几何特征, 即轨迹点的经纬度信息, 而忽略了轨迹中的时间信息、速度、方向以及转角等运动参数信息, 这样使得压缩后的轨迹数据出现压缩误差较大、轨迹特征信息缺失等问题。在整段轨迹中, 运动主体的速度是在发生变化的, 各轨迹子段对应的平均速度会产生差异, 因此导致了在某一平均速度下衡量偏移量的差异。如一辆运动的汽车, 10 m 的距离偏移量在 10 m/s 和 100 m/s 的不同速度尺度下, 衡量其偏移量偏移程度的标准是不相同的。如果在这两种不同的速度尺度下使用同一距离阈值来进行压缩点的判断, 则忽略了因速度差异而带来的尺度偏差。另外, 同样长度的一段轨迹, 不同的时间采样间隔下, 对于轨迹点的选取也是不同的。如在 2 s 和 20 s 的不同时间采样间隔下, 轨迹段所含有的信息量是不同的, 其距离阈值的设定也应该相应改变。针对这些问题, 采用同步欧式距离 SED 来代替传统的垂直欧式距离, 综合考虑到了轨迹点中的时间信息, 并利用速度差来实现动态阈值设定, 找出前后平均速度差小于所设定的速度差阈值的轨迹子段, 使组内各轨迹段之间的速度差值在一定范围内, 这样就保证了在该组内平均速度下的尺度一致性, 从而可以在该速度尺度下设置相同的距离阈值。同时, 由于前后速度差异较大的轨迹点已自动作为不同距离阈值的分界点, 因此也作为轨迹段中的轨迹特征点得到了保留, 轨迹段的特征信息得到了充分体现。

### 2.3 算法步骤

图3为该改进算法的流程图。其中,  $V_s$  为设定的速度差阈值;  $D_s$  为通过平均速度来设置对应的同步欧式距离阈值;  $D$  为  $P_{cur}$  到  $P_{start}, P_{end}$  之间连线的同步欧式距离。

## 3 实验与分析

### 3.1 实验数据

本文采用的数据为微软亚洲研究院 GeoLife 项目组收集的数据, 原始数据为 2017 年 4 月—2018 年 8 月收集的 182 个用户的轨迹数据, 共包含了 17 621 个轨迹, 总距离约为 120 000 km, 总时间约为 48 000 h。本次实验截取了一位用户在 2007-10-14—2007-10-15 的一段 GPS 数据。由于数据集所含的信息量巨大, 因此在实验前对原始轨迹数据集进行了数据预处理、关键信息提取等工作。

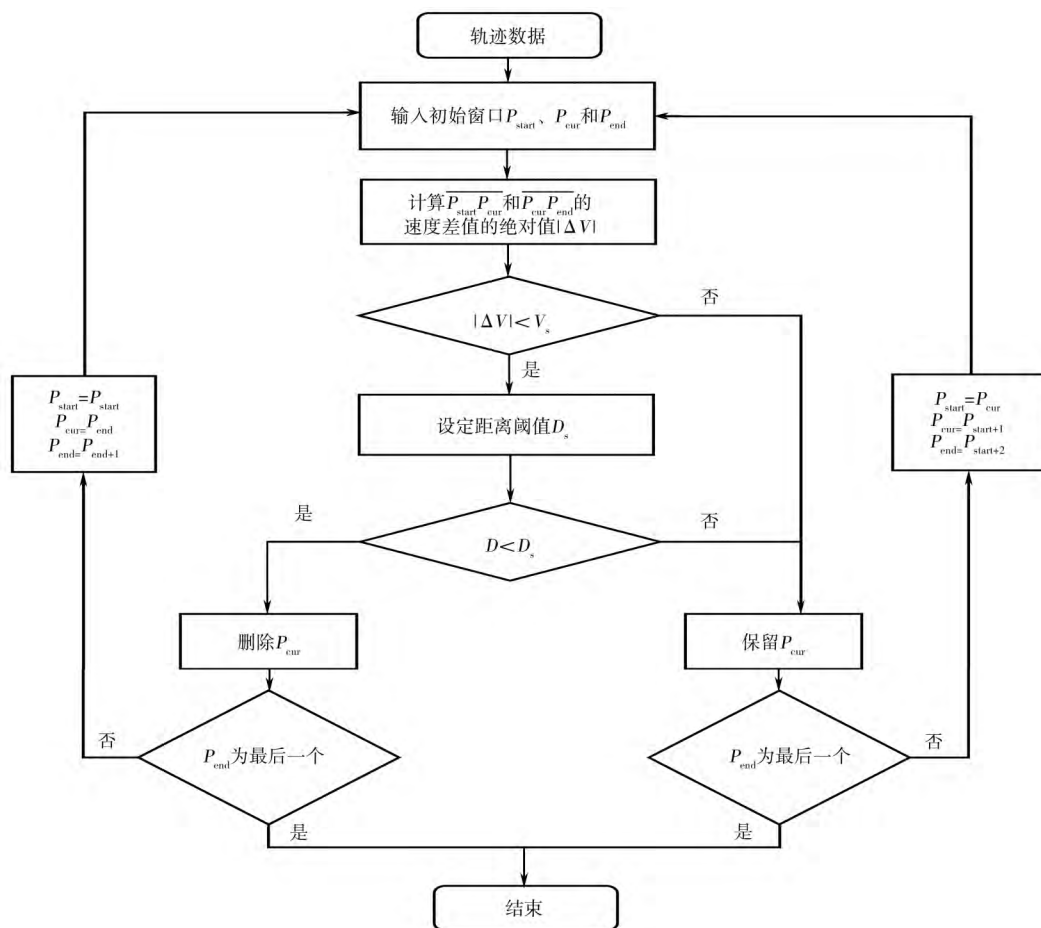
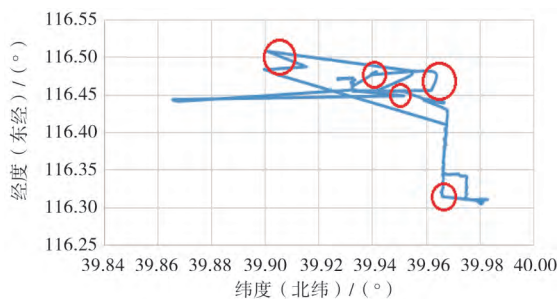


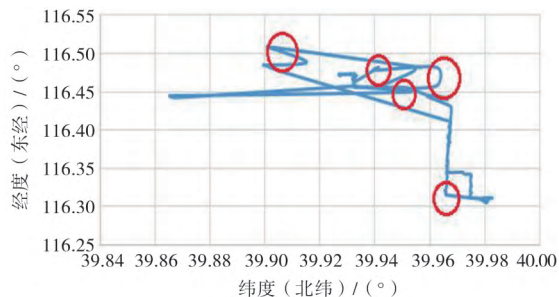
图 3 改进的压缩算法流程

### 3.2 结果分析

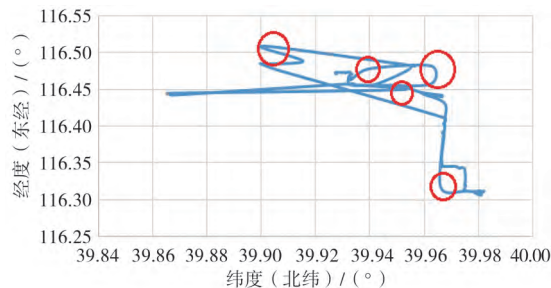
为说明改进算法对于实时数据的压缩效果,采用按照时间序列进行动态读取的方法,模拟实时在线的情况,进行实时轨迹数据的接收。改进算法的压缩效果如图 4 所示。



(a) 原始轨迹



(b) 滑动窗口算法压缩轨迹



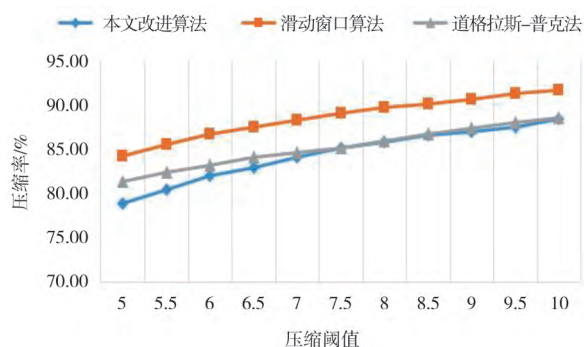
(c) 改进算法压缩轨迹

图 4 压缩结果对比

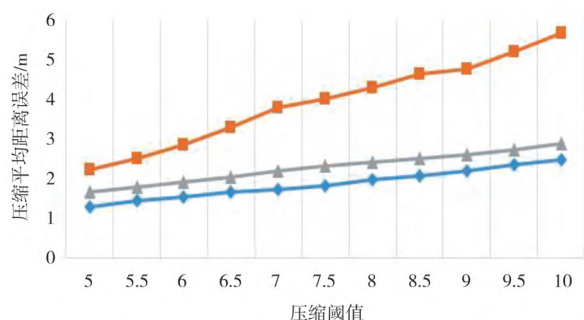
图 4 中,红色圆圈圈出的为各轨迹图像的对比部分。可以看出,与图 4(a) 相比,尽管图 4(c) 在部分轨迹图像上产生了小部分误差,但从整体上看,其压缩后的轨迹图像与原始轨迹图像仍能够大体拟合。同图 4(b) 相比,图 4(c) 能够更好地反映原始轨迹图像的形状特征。改进后算法的有效性从直观上得到了验证。

为了准确评价算法的性能,采用压缩率和平均距离误差这两种指标来进行算法的性能描述,并将本文算法同传统的道格拉斯-普克压缩法,以及经典的滑动窗口轨迹压缩算法进行实验对比。由交通规则及城市内汽车行驶速度限制,以及日常生活汽车行驶的经验,将速度差阈值设置为

10 m/s,对比实验通过设置不同的距离阈值来进行比较。其实验结果如图5所示。



(a) 压缩率对比



(b) 压缩算法的平均误差对比

图5 对比结果

从图5可以看出,在设置相同的固定距离阈值的情况下,由于改进算法参考了轨迹点的速度特征,保留的轨迹特征点增多,因此压缩率较低。但也因为考虑了时间信息与速度特征,其平均误差跟其他两种算法相比也是最小的。从图5(a)压缩率的变化图像上看,设定的距离阈值越大,其压缩率就越高,这是因为更大的距离阈值使得保留到轨迹数据集中的点相应减少,但轨迹数据所含的信息量也受到了削减,信息丢失的风险增大。从图5(b)所反映的算法平均距离误差变化情况来看,随着设定的距离阈值增大,非特征点到相邻两特征点之间的距离之和增大,3种算法的平均距离误差也增大。同时,从图5中可以看出,随距离阈值的增加,本文算法距离误差增长的趋势是最稳定的。

为进一步说明本文算法在压缩精度上的优越性,调整算法的距离阈值,使3种算法达到相同的压缩率,对3种算法在相同压缩率下的压缩误差进行对比,实验结果如图6所示。可以看出,在相同的压缩率下,本文算法的平均距离误差是最小的,其压缩精度在3种算法中为最优,有效提高了压缩的准确率。

由于本文算法所采用的为基于速度的动态距离阈值,下面在不同速度差下设置不同的距

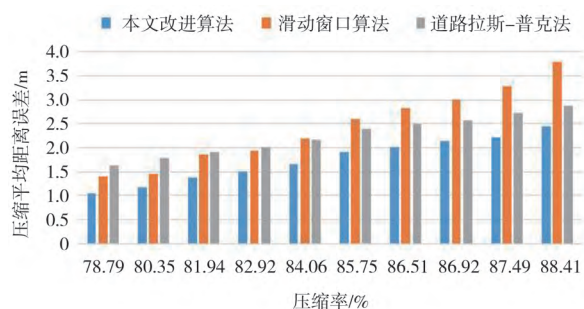
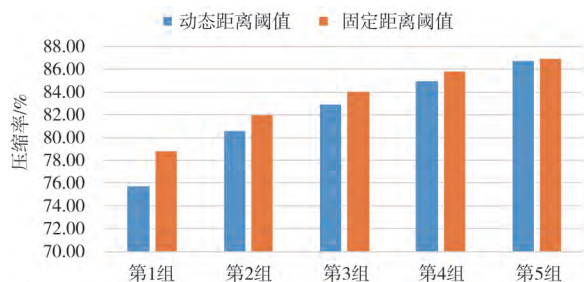
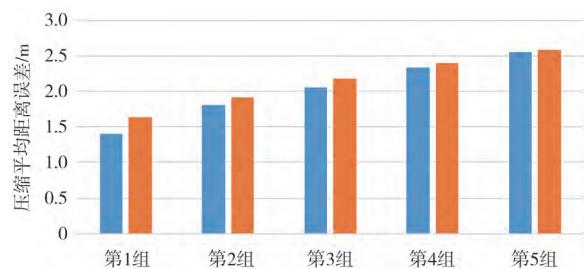


图6 相同压缩率下的压缩误差比较

离阈值,并对实验结果进行比较。将速度差分为0~2、2~4、4~6、6~8、8~10 m/s这5段,对每个不同的速度差赋予不同的距离阈值。第1组距离阈值分别设置为3、4、5、6、7 m;第2组距离阈值分别设置为4、5、6、7、8 m;第3组分别为5、6、7、8、9 m;第4组分别为6、7、8、9、10 m;第5组分别为7、8、9、10、11 m;各组距离阈值的平均值分别为5、6、7、8、9 m。实验结果见图7。



(a) 压缩率对比



(b) 平均压缩误差对比

图7 动态阈值和固定阈值下的对比结果

从图7(a)中可以看出,采用基于速度差的动态距离阈值的压缩方法,其压缩率较固定阈值低。图7(b)反映了其压缩误差得到了进一步的减小,这是因为动态距离阈值考虑到了因速度差异而造成的衡量尺度的不同,进一步提高了压缩的效果和精度。

顾及运动特征的滑动窗口轨迹数据压缩改进算法同其他两种算法相比,压缩误差有明显降低,其压缩精度在几种对比的算法中为最优,并且还保留了更多轨迹数据自身的特征信息,更好地保证了轨迹信息的完整度,压缩后得到的轨迹与原始的轨迹也更加接近。



#### 4 结论

针对经典滑动窗口轨迹压缩算法忽略轨迹点运动特征信息的问题,提出了一种顾及时间同步欧式距离和运动速度的轨迹数据压缩算法,采用同步欧式距离来代替传统的垂直欧式距离,并利用速度差来实现动态阈值的设定。实验结果表明:1) 顾及运动特征的滑动窗口轨迹数据压缩改进算法考虑到了轨迹数据所含有时间信息以及速度特征信息,实现了更高精度的数据压缩,并且能显著减小压缩误差,有效提高压缩准确率。2) 尽管算法的复杂度有所增加,耗费的时间也相应增多,但在计算能力与计算速度不断增强的今天,算法对于压缩精度的提高是十分有价值的。

下一步的工作中,将针对速度差阈值与动态距离阈值的设定,以及两者对压缩率与压缩精度的影响进行进一步研究,以设计出更加有效、精度更优的改进算法。

#### 参 考 文 献:

- [1] 郑宇.城市计算概述[J].武汉大学学报(信息科学版), 2015,40(1):1-13.  
ZHENG Y. Introduction to urban computing[J]. Geomatics & Information Science of Wuhan University, 2015,40(1):1-13.
- [2] CHEN M, XU M, FRANTI P. Compression of GPS trajectories [C]// Proceedings of the 2012 Data Compression Conference. Piscataway, USA, 2012:62-71.
- [3] 江俊文,王晓玲.轨迹数据压缩综述[J].华东师范大学学报(自然科学版),2015(5):61-76.  
JIANG J W, WANG X L. Review on trajectory data compression[J]. Journal of East China Normal University(Natural Sciences), 2015(5):61-76.
- [4] HERSHBERGER J, SNOEYINK J. Speeding up the Douglas-Peucker line-simplification algorithm [C]// International Symposium on Spatial Data Handling. Berlin, 1992:134-143.
- [5] DODGE M, PEUCKER T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature [J]. Canadian Cartographer, 2006,10(2):112-122.
- [6] 李名,胡勤友,孟良.基于 AIS 的船舶运动轨迹压缩技术研究[J].航海技术,2010(1):11-13.  
LI M, HU Q Y, MENG L. Research on ship trajectory compression technology based on AIS [J]. Marine Technology, 2010(1):11-13.
- [7] YING J C, SU J H. On Velocity-Preserving Trajectory Simplification [C]// Proceedings of the 8th Asian Conference on Intelligent Information and Database Systems. Berlin, 2016:241-250.
- [8] 徐凯,邱家瑜,李燕.一种加入时间维的船舶轨迹高效离线压缩算法研究[J].计算机科学,2017,44(S2):498-502.  
XU K, QIU J Y, LI Y. Offline efficient compression algorithm for AIS data retains time elapsing dimension [J]. Computer Science, 2017,44(S2):498-502.
- [9] KEOGH E, CHU S, HART D, et al. An online algorithm for segmenting time series [C]// Proceedings of the IEEE International Conference on Data Mining. San Jose, 2001:289-296.
- [10] POTAMIAS M, PATROUMPAS K, SELLIS T. Sampling Trajectory Streams with Spatiotemporal Criteria [C]// Proceedings of the 18th IEEE International Conference on Scientific and Statistical Database Management. Washington, 2006:275-284.
- [11] MUCKELL J, HWANG J H, PATIL V, et al. SQUISH: An online approach for GPS trajectory compression [C]// Proceedings of the 2nd International Conference on Computing for Geospatial Research & Application. New York, 2011:1-8.
- [12] 吴家皋,刘敏,韦光,等.一种改进的滑动窗口轨迹数据压缩算法[J].计算机技术与发展,2015,25(12):47-51.  
WU J G, LIU M, WEI G, et al. An improved trajectory data compression algorithm of sliding window [J]. Computer Technology and Development, 2015,25(12):47-51.
- [13] 刘磊军,房晨,张磊,等.基于运动状态改变的在线全球定位系统轨迹数据压缩[J].计算机应用,2016,36(1):122-127.  
LIU L J, FANG C, ZHANG L, et al. Online compression of global positioning system trajectory data based on motion state change [J]. Journal of Computer Applications, 2016,36(1):122-127.
- [14] 张达夫,张昕明.基于时空特性的 GPS 轨迹数据压缩算法[J].交通信息与安全,2013,31(3):6-9.  
ZHANG D F, ZHANG X M. A spatiotemporal compression algorithm for GPS trajectory data [J]. Journal of Transport Information and Safety, 2013,31(3):6-9.
- [15] 吴青华.基于相对同步欧氏距离筛选的在线 GPS 轨迹数据压缩算法[J].计算机应用与软件,2018,35(3):282-288.  
WU Q H. Online GPS trajectory data compression algorithm based on relative synchronous Euclidean distance filtering [J]. Computer Applications and Software, 2018,35(3):282-288.
- [16] MERATNIA N, ROLF A. Spatiotemporal compression techniques for moving point objects [M]. Berlin: Springer, 2004:765-782.
- [17] 吴佩莉.移动对象轨迹数据管理关键技术研究[D].北京:北京理工大学,2015:78.  
WU P L. Research on key technologies of trajectory data management for moving objects [D]. Beijing: Beijing Institute of Technology, 2015:78.
- [18] 冯神柱.路网轨迹数据的压缩存储技术研究[D].杭州:杭州电子科技大学,2014:16.  
FENG S Z. The research on technologies of compression storage for trajectory on road network [D]. Hangzhou: Hangzhou Dianzi University, 2014:16.
- [19] MUCKELL J, OLSEN P W, HWANG J, et al. Compression of trajectory data: A comprehensive evaluation and new approach [J]. Geoinformatica, 2014,18(3):435-460.

责任编辑 王 净