

文章编号:1671-8860(2012)05-0550-05

文献标志码:A

# 利用偏最小二乘方法修复高光谱影像等距映射降维中遗失点的坐标

孙伟伟<sup>1</sup> 刘春<sup>1,2</sup> 施蓓琦<sup>3</sup> 李巍岳<sup>1</sup>

(1 同济大学测量与国土信息工程系,上海市四平路1239号,200092)

(2 现代工程测量国家测绘地理信息局重点实验室,上海市四平路1239号,200092)

(3 上海师范大学城市信息研究中心,上海市桂林路100号,200234)

**摘要:**将Isomap流形学习方法应用于高光谱影像非线性降维时,在构建最短路径过程中,其边界点往往被忽略而没有低维流形坐标。对此,引入偏最小二乘方法来模拟修复遗失点的流形坐标,并从两个方面进行了综合评价。实验结果表明,模拟流形坐标与实际坐标吻合很好。

**关键词:**等距映射;偏最小二乘;流形坐标修复;最短路径

**中图法分类号:**P237.3

近年来,一些学者在模式识别领域提出了流形学习降维方法,其假设高维数据集均匀采样于低维流形上,通过发现其潜在的低维流形,实现降维或低维可视化<sup>[1]</sup>。尤其是等距映射(isometric mapping, Isomap)方法作为流形学习中全局保持方法的典型代表<sup>[2]</sup>,能够保持降维前后数据点间的测地线等距。将Isomap方法引入高光谱影像降维中,像元个数为高维空间的样本点大小,波段数为高维空间的维数。Isomap降维首先是基于邻域构建高光谱影像的最短路径图,图中邻域内部点的测地距离采用欧氏距离,邻域中心点与邻域外点的测地距离用Dijkstra算法计算最短路径来逼近,然后采用多维尺度变换降维最短路径矩阵。考虑到图谱的稳定性,最短路径图中具有较大连接数的点将被嵌入到低维空间中,而边界点被忽略而没有低维流形坐标<sup>[2]</sup>。通常,高光谱影像Isomap降维研究都采用训练集<sup>[3-7]</sup>,这些研究都将高光谱影像看作常规高维数据集,最短路径边界点的遗失对其研究的影响不大。然而,Isomap方法广泛应用于高光谱影像来实现降维,影像数据集并非是训练集,因此,最短路径边界点的问题必须重新审视。另一方面,高光谱影像具有图谱合一的特性,降维结果通过图像形式地表达,

因此不能将高光谱影像看作常规数据集而忽略最短路径边界点的问题。

针对Isomap降维中最短路径边界点的问题,有学者采用最短路径图谱自相容方法,通过在最短路径图中寻找边界点的最邻近非连接点,建立两者连接,实现其重新连接到最短路径图中<sup>[8]</sup>。虽然该方法可获得最短路径边界点的流形坐标,但无法证明该流形坐标与实际流形坐标吻合。本文采用数学模拟方法,基于偏最小二乘算法建立了最短路径非边界点的高维光谱向量和低维流形坐标间的拟合关系,来模拟修复边界点的流形坐标。

## 1 算法原理

Isomap的基本思想是:当数据集的分布具有低维流形结构时,可通过等距映射获得高维数据集在低维空间的表示<sup>[2]</sup>。高光谱影像已被证明存在低维非线性流形特性<sup>[9]</sup>,因此,可以用Isomap方法实现降维得到流形坐标。高光谱影像Isomap降维的算法步骤参见文献[2]。

偏最小二乘方法(partial least square, PLS)<sup>[9]</sup>常用于数据“软”建模,建立反映变量关于

收稿日期:2012-03-15。

项目来源:上海市教育委员会科研创新基金资助项目(10ZZ25,11YZ290);地理空间信息工程国家测绘局重点实验室经费资助项目(200914);现代工程测量国家测绘局重点实验室开放基金资助项目(TJES1010)。

解释变量的回归方程。常用的方法是非线性迭代偏最小二乘法 (nonlinear iterative partial least squares, NIMPLS), 它采用交叉核实法作为迭代停止准则<sup>[9]</sup>。

高光谱影像流形坐标向量间的相关性虽然大大减少, 但却依然存在, 而且高维光谱向量的维数远大于低维流形的维数, 这符合 PLS 建模的数据特点要求。因此, 采用 PLS 建立最短路径非边界点的高维光谱向量与流形坐标的回归关系, 模拟遗失点的流形坐标。遗失点流形坐标修复的流程如图 1 所示。

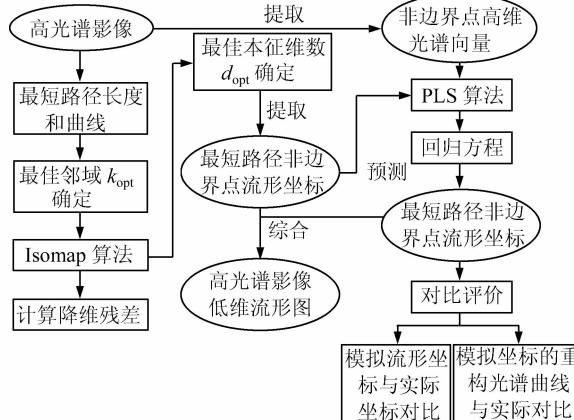


图 1 高光谱影像遗失点流形坐标修复流程

Fig. 1 Process of Repairing Manifold Coordinates of Lost Points in Hyperspectral Image

步骤如下:① 确定邻域大小  $k$ , 构建邻域图。本文采用最短路径长度和曲线“拐点”方法确定最佳邻域  $k_{\text{opt}}$ <sup>[10]</sup>。② 确定高光谱影像 Isomap 降维的本征维数  $d$ 。本文采用降维残差方法确定最佳本征维数  $d_{\text{opt}}$ <sup>[2]</sup>。③ 通过 Isomap 降维, 提取影像非边界点的  $d_{\text{opt}}$  维流形坐标。④ 采用 PLS 方法拟合非边界点的光谱向量与流形坐标间的线性关系, 并预测遗失点的 Isomap 模拟流形坐标; 然后综合边界点的模拟坐标及非边界点的低维流形坐标, 得到完整的影像低维流形图。⑤ 通过模拟坐标与相邻同类地物的坐标和模拟坐标的重构光谱曲线与实际光谱曲线的对比分析, 评价模拟流形坐标的准确性。

## 2 实验与分析

### 2.1 实验数据

选取不同区域的两种高光谱数据来验证 PLS 方法修复流形坐标的可靠性。数据 1 来自美国普渡大学遥感应用实验室网站, 是华盛顿特区库尔兹桥附近的 HYDICE 高光谱影像, 大小为

100 像素  $\times$  100 像素 (见图 2(a))。数据采集于 1995 年 8 月 23 日, 共 210 个波段, 波长范围为 0.4~2.4  $\mu\text{m}$ , 覆盖可见光至近红外区域。0.9~1.4  $\mu\text{m}$  间的波段由于大气窗口的原因而删去, 剩余 191 个波段。实验区内主要有树木、草地、道路和水体 4 种地物, 各地物间的光谱曲线差异较大 (见图 3(a))。

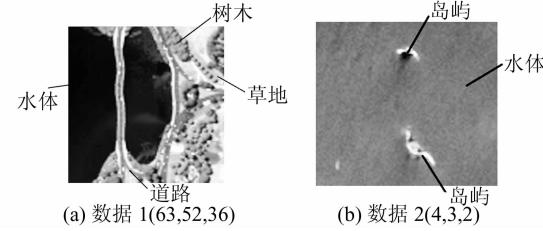


图 2 数据 1 和数据 2 的伪彩色图

Fig. 2 Pseudo Color Image of Data 1 and Data 2

数据 2 来自美国德克萨斯大学奥斯丁分校数据网站, 是福罗里达州肯尼迪空间中心附近的 AVIRIS 高光谱数据, 大小为 115 像素  $\times$  116 像素 (见图 2(b))。数据采集于 1996 年 3 月 23 日, 共 224 个波段, 光谱分辨率为 10 nm, 波长范围为 0.4~2.5  $\mu\text{m}$ , 空间分辨率为 18 m。除去水汽吸收和低信噪比波段, 剩余 176 个波段。实验区内有岛屿和水体两种地物, 两者的光谱曲线差异很大 (见图 3(b)), 其中光谱曲线跳变是由去除水汽吸收和低信噪比的波段引起的。

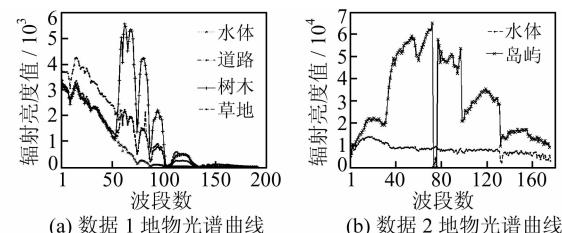


图 3 数据 1 和数据 2 的地物光谱曲线

Fig. 3 Spectral Curves of Main Ground Objects in Data 1 and Data 2

### 2.2 实验过程

设邻域  $k$  的取值区间都为 [2, 30], 对候选  $k$  值, 搜索各像元的邻域并构建最短路径图, 得到最短路径长度和曲线 (见图 4)。选取曲线“拐点”, 确定数据 1 和数据 2 的最佳邻域大小分别为 7 和 11; 选取残差曲线“拐点”, 确定数据 1 和数据 2 的最佳本征维数分别为 4 和 3。Isomap 作为数据驱动方法, 两种数据的残差曲线不同是由影像的自身差异导致的。由于高维数据的实际维数不超过 12<sup>[11]</sup>, 选取高光谱影像的本征维数  $d$  的选值区间为 [2, 12]。对每个候选  $d$ , 利用 MDS 降维最短路

径矩阵,得到非边界点的  $d$  维流形坐标及残差曲线(见图 5)。

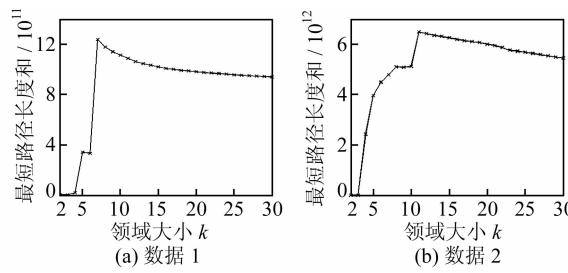


图 4 最短路径长度和与邻域大小  $k$  的关系图

Fig. 4 Relationship Between the Sum of the Shortest Path Graph and  $k$

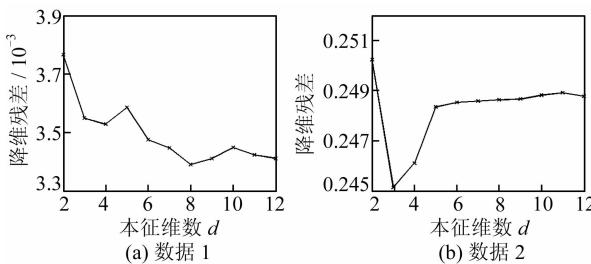


图 5 降维残差与本征维数  $d$  的关系图

Fig. 5 Relationship Between Residual Errors of Dimension Reduction and  $d$

考虑图谱的稳定性,最短路径图忽略了较小数目连接边的边界点,因此,根据  $d_{\text{opt}}$  选取的降维结果是非边界点的流形坐标,而边界点的流形坐标遗失,以致流形图出现许多空值点(图 6 标签所示为 0 值)。将非边界点的流形坐标看作反映变量 RV 的观测矩阵,相应的光谱向量看作解释变量 EV 的观测矩阵,提取 EV 和 RV,采用 PLS 算

法建立非边界点的流形坐标与光谱向量间的回归关系式。流形坐标的维数对应回归方程的个数,所以数据 1 和数据 2 的回归方程个数分别为 4 和 3。其中,PLS 算法中交叉有效性的阈值为 0.0975,EV 选取前 3 维成分  $t_{(1)}$ 、 $t_{(2)}$  和  $t_{(3)}$ 。同时,比较非边界点模拟的流形坐标与真实的流形坐标,得到回归方程相对误差。将边界点的光谱向量代入回归方程,得到模拟的流形坐标,整合得到低维流形图(见图 7)。相比图 6,流形图中的空值点已全部消除。

## 2.3 实验结果分析

### 2.3.1 遗失点的模拟坐标与真实流形坐标对比

高光谱影像中的同类地物光谱曲线相似,所以搜索邻域而得到的同类地物流形坐标应非常相近。因此,可以将其相邻同类地物的流形坐标作为遗失点的坐标真值。

数据 1 和数据 2 中,各地物的遗失点的模拟坐标与相邻同类地物的坐标对比如图 8 所示。图 8(a)中,4 种地物的模拟坐标与实际坐标都较为吻合;其中水体的最接近,其次是道路,然后是草地,最后是树木。图 8(b)中,相比岛屿,水体模拟的坐标与实际坐标更为接近。对数据 1,水体、道路、树木、草地模拟的流形坐标与实际坐标的光谱角分别为  $1.61^\circ$ 、 $3.50^\circ$ 、 $4.22^\circ$ 、 $2.81^\circ$ ;对数据 2,岛屿、水体的光谱角分别为  $3.64^\circ$ 、 $1.79^\circ$ ,都较小,与前面的结论是一致的,说明模拟坐标与实际坐标吻合较好。进一步观察发现,不同地物的模拟坐标与实际坐标间的差异不同是由 PLS 算法决定的,由于数据 1 和数据 2 中水体分布的面积最大,观测值最多,所以回归方程针对水体的效果最好。

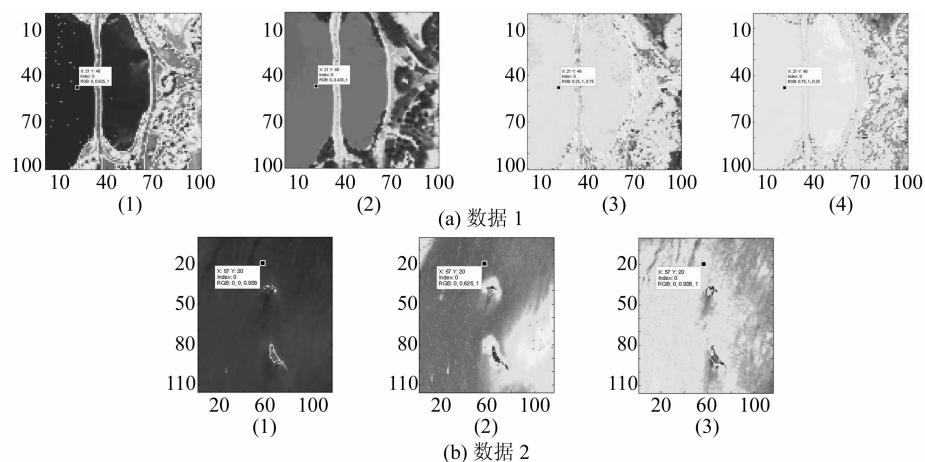


图 6 最短路径边界点修复前的低维流形图

Fig. 6 Manifold Maps Before Repairing the Boundary Points in the Shortest Path Graph

### 2.3.2 模拟坐标的重构与实际光谱向量对比

通过流形重构得到重构光谱向量,并与实际

光谱向量对比。本文采用稳健流形重建方法<sup>[12]</sup>得到高维重构光谱向量。

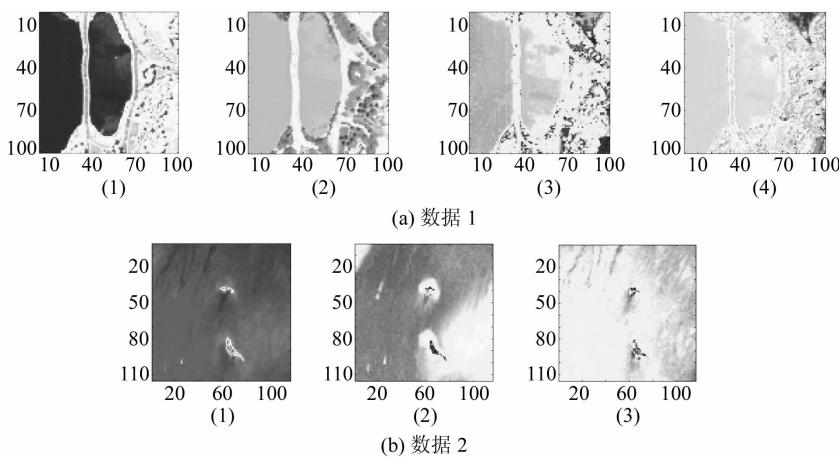


图7 遗失点流形坐标修复后的低维流形图

Fig. 7 Manifold Maps After Repairing Manifold Coordinates of the Boundary Points

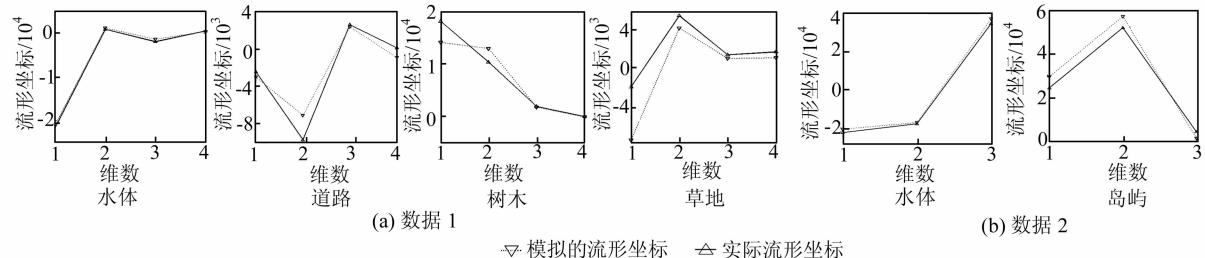


图8 遗失点模拟的流形坐标与实际坐标对比

Fig. 8 Contrast Between the Simulated and Real Manifold Coordinates of the Lost Points in Each Ground Object

将遗失点的模拟坐标代入重构函数,得到重构光谱向量。遗失点的重构光谱曲线与实际光谱曲线的对比如图9所示。图9(a)中,水体的重构光谱曲线与实际最接近,其次是道路,然后是草地,最后是树木。图9(b)中,水体的重构光谱曲线与实际最为吻合,岛屿的差些,但整体较为吻合,

仅在被移除波段附近效果稍差。对数据1,水体、道路、树木、草地重构的光谱向量与实际光谱向量间的光谱角分别为 $1.86^\circ$ 、 $2.79^\circ$ 、 $3.02^\circ$ 、 $2.15^\circ$ ;对数据2,岛屿、水体的光谱角分别为 $5.73^\circ$ 、 $1.66^\circ$ ,这也证实了上述结论,并且与§2.3.1的结论也是一致的。

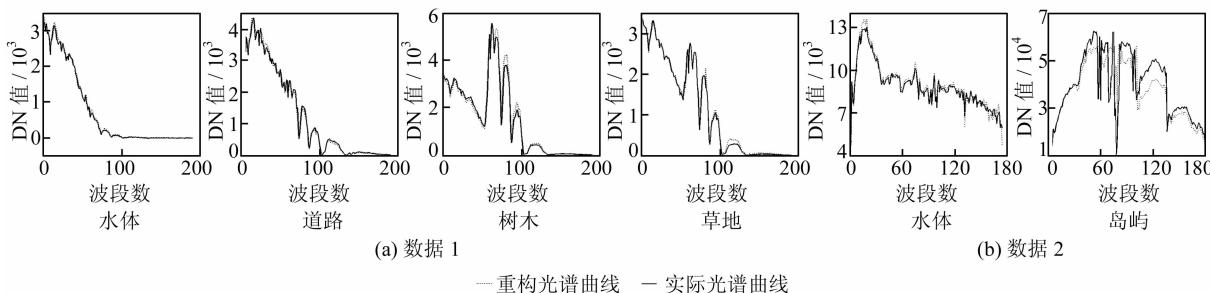


图9 主要地物的遗失点的流形重构光谱曲线与实际光谱曲线对比图

Fig. 9 Contrast Between the Reconstructed and Real Spectral Curves of Lost points in Each Ground Object

### 3 结语

本文采用PLS模拟修复Isomap降维中遗失点的低维流形坐标,实验证明,PLS能很好地修

复遗失点的流形坐标,尤其对影像中占大面积比例的地物效果最好。然而考虑计算量的问题,实验只针对小区域数据,考虑流形拼接来解决大区域数据的Isomap计算问题将是进一步要研究的内容。

## 参 考 文 献

- [1] 黄启宏, 刘钊. 流形学习中非线性维数约简方法概述[J]. 计算机应用研究, 2007, 24(11): 19-25
- [2] Tenenbaum J B, Silva V, Langford J C. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290(5 500): 2 319-2 325
- [3] Chen Y, Crawford M M, Ghosh J. Applying Nonlinear Manifold Learning to Hyperspectral Data for Land Cover Classification[C]. IEEE, 2005: 4 313-4 314
- [4] Chen Y, Crawford M M. Improved Nonlinear Manifold Learning for Land Cover Classification Via Intelligent Landmark Selection[C]. International Symposium on Geoscience and Remote Sensing, Denver, Colorado, USA, 2006
- [5] Fong M. Dimension Reduction on Hyperspectral Images[R]. University of California, Los Angeles, 2007
- [6] Dong Guangjun, Zhang Yongsheng, Song Jin. Dimensionality Reduction of Hyperspectral Data Based on ISOMAP Algorithm[C]. The 8th International Conference on Electronic Measurement and Instruments, Xi'an, China, 2007
- [7] 杜培军, 王小美, 谭昆. 利用流形学习进行高光谱遥感影像的降维与特征提取[J]. 武汉大学学报·信息科学版, 2011, 36(2): 148-152
- [8] Bachmann C M, Ainsworth T L, Fusina R A. Exploiting Manifold Geometry in Hyperspectral Imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3): 441-454
- [9] Bachmanri C M, Ainsworth T L. Modeling Data Manifold Geometry in Hyperspectral Imagery[C]. IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, 2004
- [10] Wold H. Path Models with Latent Variables: the NIPALS Approach [M]. New York: Academic Press, 1975
- [11] Shao C, Huang H. Selection of the Optimal Parameter Value for the ISOMAP Algorithm[C]. The 4th Mexican International Conference on Artificial Intelligence, Monterrey, Mexico, 2005
- [12] van der Maaten L J P, Postma E O, van der Herik H J. Dimensionality Reduction: A Comparative Review[R]. Maastricht University, MICC, 2009
- [13] 孟德宇, 徐晨, 徐宗本. 基于 Isomap 的流形结构重建方法[J]. 计算机学报, 2010, 33(3): 545-555

**第一作者简介:**孙伟伟,博士生,研究方向为 GIS-T 理论及工程应用、高光谱影像非线性降维。

E-mail: sw8525@gmail.com

## Manifold Coordinate Repairing of Lost Points with PLS for Isomap Nonlinear Dimensionality Reduction of Hyperspectral Image

SUN Weiwei<sup>1</sup> LIU Chun<sup>1,2</sup> SHI Beiqi<sup>3</sup> LI Weiyue<sup>1</sup>

(1) Department of Surveying and Geo-informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China)

(2) Key Laboratory of Advanced Engineering Survey of NASM, 1239 Siping Road, Shanghai 200092, China)

(3) Urban Information Research Center, Shanghai Normal University, 100 Guilin Road, Shanghai 200234, China)

**Abstract:** As a manifold learning method, Isomap has been widely used for making nonlinearly reduction for hyperspectral image. However, during the construction process of the shortest path graph, the boundary points, which are not noise points, have always been omitted for the consideration of the stability of the graph. Therefore, the PLS method is introduced to repair and simulate the manifold coordinates of the lost points in the shortest path graph. And the simulated manifold coordinates have been evaluated from two different aspects to verify our method. The results show that the simulated manifold coordinates agree well with the real one. It will be quite useful for further classification or visualization with low dimensional manifold image.

**Key words:** Isomap; PLS; manifold coordinates repairing; the shortest path graph

**About the first author:** SUN Weiwei, Ph.D candidate, majors in GIS-T theory and engineering application, nonlinear dimensionality reduction of hyperspectral image.

E-mail: sw8525@gmail.com