

# 利用流形学习进行高光谱遥感影像的降维与特征提取

杜培军<sup>1</sup> 王小美<sup>1,2</sup> 谭 琨<sup>1</sup> 夏俊士<sup>1</sup>

(1 中国矿业大学国土环境与灾害监测国家测绘局重点实验室, 徐州市大学路 1 号, 221116)  
(2 黄河水利职业技术学院测绘工程系, 开封市东京大道西段, 475004)

**摘 要:** 基于最新的非线性降维方法——流形学习的理论, 从高光谱遥感数据内在的非线性结构出发, 采用全局化的等距映射(Isomap)方法进行降维, 取得了优于常用的 MNF 方法的结果。把光谱角和光谱信息散度与测地距离相结合用于 Isomap 算法, 结果在冗余方差和光谱规范化特征值方面优于采用传统欧氏距离计算邻域的 Isomap 方法。实验表明, 流形学习是一种有效的高光谱遥感数据特征提取方法。  
**关键词:** 高光谱遥感; 降维; 流形学习; Isomap 算法  
**中图法分类号:** P237.3

高光谱遥感数据有几十甚至上百个波段, 具有信息量大、信息相关性强、信息冗余多等特点, 因此高维数据处理一直是高光谱遥感信息处理与实际应用的难题<sup>[1-3]</sup>。为了解决高维数据处理的问题, 通常利用特征提取与特征选择进行降维处理<sup>[4-7]</sup>。传统的降维方法主要是线性方法, 其中应用最广泛的特征提取算法是主成分分析(principal component analysis, PCA)和最大噪声分离(maximum noise fraction, MNF)。然而, 线性降维方法主要适用于数据空间分布是线性结构的情况, 而高光谱遥感数据中, 由于光谱分辨率很高, 导致很多异物同谱的地物在高维空间中会分布在不同的但却可能很接近的区域, 即高光谱数据在高维空间中会分布成许多小的数据云团, 这些云团之间并不是线性分布的。近年来兴起并在模式识别领域得到广泛应用的流形学习(manifold learning)在非线性降维方面具有明显的优越性<sup>[8-11]</sup>。本文试图将流形学习应用到高光谱遥感数据中, 结合高光谱遥感数据内在的特性, 探讨流形学习方法应用于高光谱遥感信息处理的实现方法, 以期的高光谱遥感数据非线性特征提取引入新的方法。

目前, 常用的流形学习算法主要包括等距映射 Isomap<sup>[8]</sup>、局部线性嵌入 LLE<sup>[9]</sup>、拉普拉斯映

射法<sup>[12]</sup>、局部切空间排列算法<sup>[13]</sup>等。其中, Isomap 和 LLE 是两种最具代表性且得到广泛应用的非线性降维方法; Isomap 是建立在多维尺度空间(multi-dimensional scaling, MDS)<sup>[14]</sup>基础上, 力求保持数据点的内在几何性质, 即保持两点间的测地距离<sup>[8]</sup>。本文在高光谱遥感影像降维中选用 Isomap 算法。

## 1 高光谱遥感数据的结构

对高光谱遥感数据, 由于图像上的像元表示了不同的地物, 数据的分布应成为云团式的聚类结构, 同一类地物的数据聚成一团, 不同类地物的数据则形成多个大小不一的云团, 在原始高维空间中, 这些云团是可区分的<sup>[7]</sup>。特征提取的目的是希望在低维空间中仍能表现不同类别地物所构成的云团, 从而实现在低维空间中对地物的分类操作。高光谱数据具有内在的非线性特征, 它的非线性源包括地面散射的非线性、表述模型为双向反射分布函数(BRDF)、最小反射率波长明显的非线性变化、像素内水体变化的衰减作用、一个像素内的多种散射和亚像元成分的异质性等, 这些非线性特征在传统的算法中并没有被充分考虑<sup>[15]</sup>。线性特征提取方法无法体现出高光谱遥

感数据的非线性特征, 而流形学习的算法特点和适用于高维数据非线性特征分析的优势, 已在高光谱遥感影像分析与数据处理中得到了初步的应用。图 1 显示了高光谱数据内在的非线性特征, 图 1(a) 中, 48、22、3 波段对应的中心波长分别为  $0.975\ 1\ \mu\text{m}$ 、 $0.710\ 9\ \mu\text{m}$ 、 $0.477\ 3\ \mu\text{m}$ , 分别对应植被、耕地(收割后)、裸地。图 1(c) 中, 在流形坐标系统下, 两点间距离沿着流形线测量, 而不是通常用的欧氏距离。

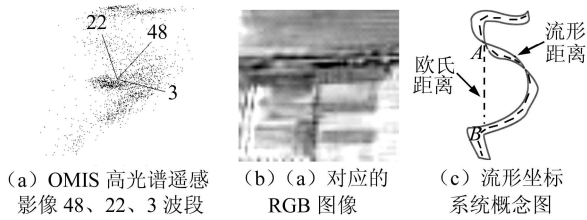


图 1 高光谱数据的内在非线性特征  
Fig. 1 Nonlinear Inherent Structure of Hyperspectral Data

## 2 基于 Isomap 算法的高光谱遥感数据降维

### 2.1 Isomap 算法中的距离计算及其改进

Isomap 在计算最短距离时花费时间较多, 存储量大, 用于数据量很大的高光谱数据时效率很低。计算最短路径的方法一般有两种: Floyd 算法和 Dijkstra 算法。利用 Floyd 算法计算最短路径时, 其复杂度为  $O(N^3)$ ; 利用斐波纳契堆作为优先队列的 Dijkstra 算法计算最短路径时, 其复杂度为  $O(kN^2\lg N)$  ( $k$  为邻域大小), 此后还要对  $N \times N$  大小的矩阵进行特征值求解, 其复杂度为  $O(N^3)$ <sup>[16]</sup>。为了减少 Isomap 的计算时间, Silva 和 Tenenbaum<sup>[16]</sup> 提出了 L-Isomap 方法。这种方法从样本中选出  $n$  个点作为标记点, 在计算距离矩阵  $D_c$  时, 并不是所有的点之间的距离都计算, 而是计算样本点与标记点之间的距离, 可以节省计算距离的时间。本文也采用 L-Isomap 方法进行高光谱特征提取。

Isomap 算法还需要确定邻域参数  $k$  来构建邻域。研究中,  $k$  是通过残差来选择的<sup>[13]</sup>。然而, 这种根据最终映射“质量”来判断对应的邻域大小是否合适的后验式策略比较耗时。

在流形学习中, 无论哪一种流形学习算法都需要计算近邻点, 一般用  $\varepsilon$  邻域或  $k$  邻域来确定近邻点, 在计算距离时, 常用欧氏距离。但对于遥感数据来说, 因为光谱之间的距离或相似性度量可以采

用不同的方法, 因此研究中将一些常用的光谱度量指标来代替欧氏距离, 以实验特征提取的效果。

#### 2.1.1 光谱角

光谱角 (spectral angle, SA) 是对地物光谱波形相似度的一种度量, 它将每条光谱都视为波段空间的一个矢量, 通过计算两个光谱之间的“角度”来确定它们之间的相似性。光谱角可以通过下式确定光谱  $A$  和光谱  $B$  之间的相似性:

$$\cos \alpha = \frac{AB}{|A||B|} = \frac{\sum_{i=1}^N A_i B_i}{\sqrt{\sum_{i=1}^N A_i^2} \sqrt{\sum_{i=1}^N B_i^2}} \quad (1)$$

式中,  $N$  是波段数。两光谱之间的夹角越小, 说明光谱之间的相似性越大。光谱角与矢量的模无关, 即不受向量本身长度的影响。光谱角对乘性干扰具有良好的抗干扰性, 不受光照度条件的变化。

#### 2.1.2 光谱信息散度

光谱信息散度 (spectral information divergence, SID) 可以用来衡量两个光谱之间的相似性。假设高光谱两个像元  $N$  个波段的光谱分别为  $A = (A_1, A_2, \dots, A_N)$ ,  $B = (B_1, B_2, \dots, B_N)$ , 则光谱信息散度为<sup>[17]</sup>:

$$\text{SID}(A, B) = D(A \| B) + D(B \| A) \quad (2)$$

式中,

$$\begin{aligned} D(A \| B) &= \sum_{i=1}^N p_i \log_i(p_i/q_i) \\ D(B \| A) &= \sum_{i=1}^N q_i \log_i(q_i/p_i) \\ p_i &= A_i / \sum_{i=1}^N A_i, \quad q_i = B_i / \sum_{i=1}^N B_i \end{aligned}$$

### 2.2 实验结果

#### 2.2.1 实验数据

实验中采用两个高光谱遥感数据, 一个是在美国印地安纳州某农林混合实验场获取的 220 波段的 AVIRIS 高光谱图像 ( $R: 46(0.821\ 79\ \mu\text{m})$ ,  $G: 14(0.527\ 85\ \mu\text{m})$ ,  $B: 6(0.449\ 07\ \mu\text{m})$ ), 波长范围为  $0.4 \sim 2.5\ \mu\text{m}$ , 图像空间分辨率为  $25\ \text{m}$ , 在去除水汽吸收波段和低信噪比波段后, 保留了其中 200 个波段进行数据处理。另一个是延安市枣园镇的 OMIS 高光谱遥感影像 ( $R: 18(0.663\ 9\ \mu\text{m})$ ,  $G: 9(0.553\ 1\ \mu\text{m})$ ,  $B: 2(0.465\ 0\ \mu\text{m})$ ), 波段为 80 个, 波长范围为  $0.4 \sim 1.7\ \mu\text{m}$ 。

#### 2.2.2 流形学习降维结果与分析比较

最大噪声分离 (MNF) 本质上是两次层叠的

主成分变换,是一种较为常用的高光谱遥感降维方法。用 MNF 变换作为线性降维方法来进行对照实验。MNF 实验通过 ENVI 软件实现,流形学习算法 Isomap 则通过 MATLAB 编程实现。

图 2(a) 为 AVIRIS 高光谱遥感影像及其用欧氏距离计算邻域的 Isomap 和 MNF 降维后的前 5 个成分的图像,进行 Isomap 降维时,  $k$  值取 30。图 2(b) 为 OMIS 高光谱遥感影像及其用欧氏距离计算邻域的 Isomap 和 MNF 降维后的前 5 个成分的图像,进行 Isomap 降维时,  $k$  值取 15。从图 2 可以看出,流形学习降维结果的前 5 个分

量信息丰富,而且细节表现清楚。图 3(a) 为通过对 MNF 与用欧氏距离的 Isomap 的规则化谱(  $\lambda/\lambda_1$  )之间的比较,可以看出,Isomap 的规则化特征谱在最初的 5~ 10 个 Isomap 成分下降了两个数量级,而 MNF 则比 Isomap 多几倍的成分才能达到相同的标准。图 3(b) 为通过对 MNF 与用欧氏距离的 Isomap 降维维数与信息量关系的比较,可以看出,流形学习降维在前 5~ 10 个成分上信息量达到了 90% 以上,而 MNF 需要超过 30 个成分以上才可以达到同样的标准,说明流形学习在信息量保持上优于 MNF。

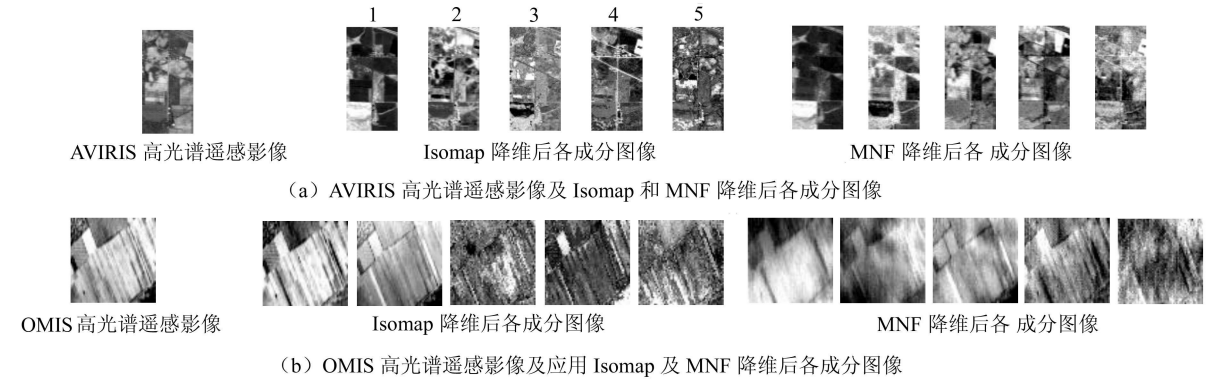


图 2 高光谱遥感影像流形学习降维处理结果与传统 MNF 方法降维处理结果比较

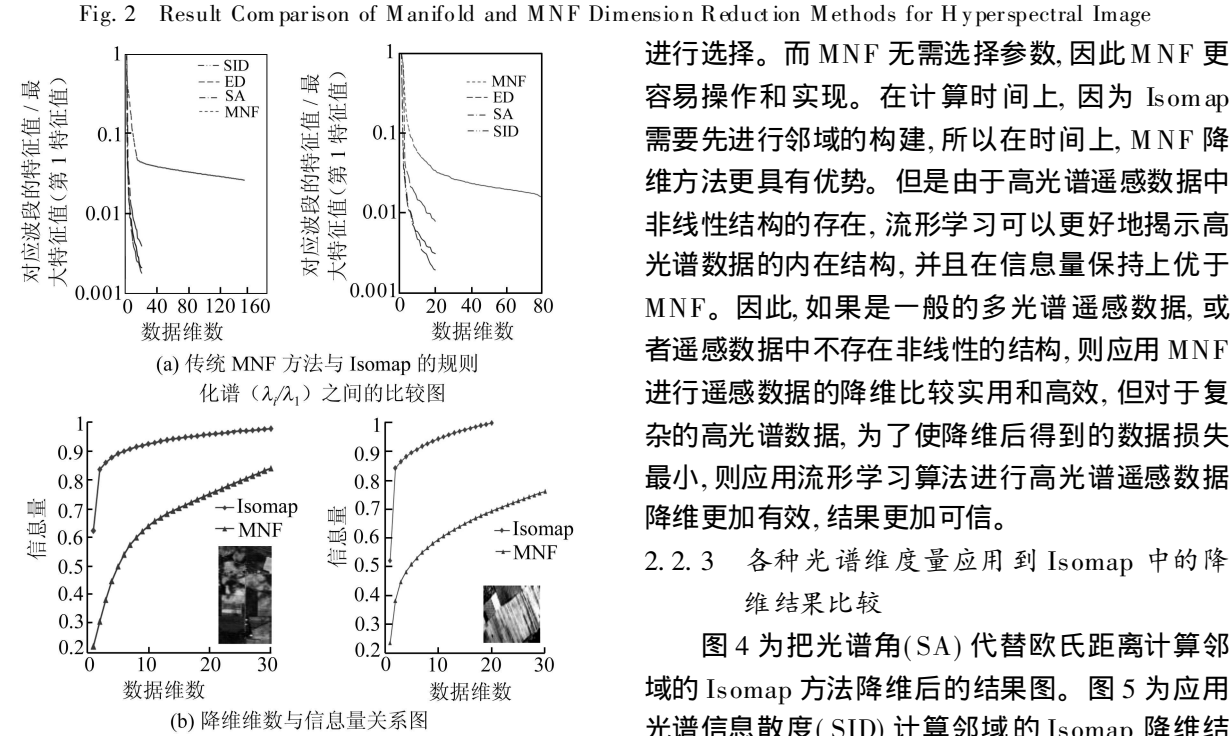


图 3 流形学习降维结果与传统 MNF 方法降维结果比较  
Fig. 3 Result Comparison between Manifold and MNF

Isomap 算法进行高光谱遥感数据降维需要选择邻域参数  $k$  和降维后的维数  $d$ , 参数  $k$  的选择对降维结果有很大的影响, 本文用后验式方法

进行选择。而 MNF 无需选择参数, 因此 MNF 更容易操作和实现。在计算时间上, 因为 Isomap 需要先进行邻域的构建, 所以在时间上, MNF 降维方法更具有优势。但是由于高光谱遥感数据中非线性结构的存在, 流形学习可以更好地揭示高光谱数据的内在结构, 并且在信息量保持上优于 MNF。因此, 如果是一般的多光谱遥感数据, 或者遥感数据中不存在非线性的结构, 则应用 MNF 进行遥感数据的降维比较实用和高效, 但对于复杂的高光谱数据, 为了使降维后得到的数据损失最小, 则应用流形学习算法进行高光谱遥感数据降维更加有效, 结果更加可信。

### 2.2.3 各种光谱维度量应用到 Isomap 中的降维结果比较

图 4 为把光谱角(SA)代替欧氏距离计算邻域的 Isomap 方法降维后的结果图。图 5 为应用光谱信息散度(SID)计算邻域的 Isomap 降维结果图。

在各种度量之间再对降维后的前 20 个成分进行比较。AVIRIS、OMIS 高光谱遥感数据利用欧氏距离计算邻域的 Isomap 方法光谱规范化特征值最小分别为 0.003 9 和 0.007 8, 利用 SA 和 SID 计算邻域的 Isomap 方法光谱规范化特征值

最小分别为 0.002 1 和 0.001 8、0.003 0 和 0.001 9。可以看出, SID 和 SA 计算邻域的 Isomap 方法在光谱规范化特征值上优于用欧氏距离计算邻域的 Isomap 方法。

在冗余方差上, AVIRIS、OMIS 高光谱遥感数据利用欧氏距离计算邻域的 Isomap 方法的冗余方差最小值分别为 0.015 4 和 0.016 4, 而用 SA 和 SID 计算邻域的 Isomap 方法的冗余方差最小值分别为 0.005 5 和 0.003 5、0.007 94 和 0.005 05。可以看出, 利用 SA 和 SID 计算邻域的 Isomap 降维结果比用欧氏距离的 Isomap 降维结果的冗余方差小得多。因此, 采用 SA 和 SID 作为距离度量能够有效地提高 Isomap 算法的降维性能。

利用 SID 计算邻域的 Isomap 方法得到的光谱规范化特征值和冗余方差都小于利用 SA 计算邻域的 Isomap 方法得到的结果, 说明利用 SID 计算邻域的 Isomap 方法降维后的数据信息量更集中于前面的几维数据上。从图 4 和图 5 的比较也能看出, 利用 SID 计算邻域的 Isomap 方法降维后得到的图像更清楚、平滑, 细节更丰富。但从计算时间上来看, SID 的计算时间要比欧氏距离和 SA 的计算时间长, 这对数据量较大的遥感数据来说是一个限制因素。

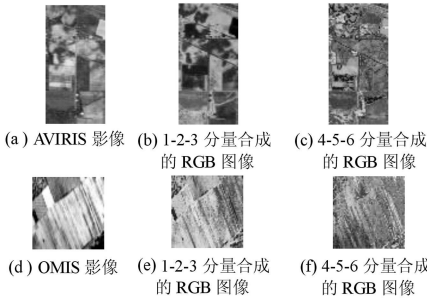


图 4 应用 SA 度量的 Isomap 算法降维结果  
Fig. 4 Results of Isomap Using Spectral Angle as Neighborhood Distance Metric

3 结 语

本文在分析高光谱数据非线性结构的基础上, 把最新的非线性降维方法——流形学习应用于高光谱数据, 选择全局的 Isomap 算法来进行高光谱数据的降维, 并针对高光谱遥感数据的特点, 将光谱角、光谱信息散度用于改进距离计算。将 Isomap 流形学习降维结果与主流的 MNF 降维结果比较, 发现 Isomap 流形学习降维在光谱规范化特征值方面优于 MNF 变换。此外, Isor

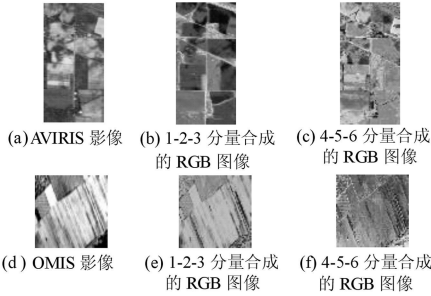


图 5 应用 SID 度量的 Isomap 算法降维结果  
Fig. 5 Results of Isomap Using SID as Neighborhood Distance Metric

map 流形学习降维在信息量的保持方面也取得了很好的结果, 并且可以揭示非线性数据的整体结构。通过将光谱角和光谱信息散度这两个光谱维度量特征与测地线距离相结合, 发现无论是在冗余方差、信息量, 还是在光谱规范化特征值方面, 都优于欧氏距离计算邻域的 Isomap 降维结果。因此, 建议在今后的研究中, 可以直接采用光谱角或是光谱信息散度计算邻域。如何实现大尺度的高光谱遥感数据中的 Isomap 降维, 是以后进一步研究的问题。

参 考 文 献

[1] 童庆禧, 张兵, 郑兰芬. 高光谱遥感——原理、技术与应用[M]. 北京: 高等教育出版社, 2006

[2] Hughes G F. On the Mean Accuracy of Statistical Pattern Recognition[J]. IEEE Trans Inf Theory, 1968, IT-14(1): 55-63

[3] 杜培军, 陈云浩, 王行风, 等. 遥感科学与进展[M]. 徐州: 中国矿业大学出版社, 2007: 369-370

[4] Kumar S, Ghosh J, Crawford M M. Best Bases Feature Extraction Algorithms for Classification of Hyperspectral Data[J]. IEEE Trans Geosci and Rem Sens, 2001, 39(7): 1368-1379

[5] Hsu P H. Feature Extraction of Hyperspectral Images Using Wavelet and Matching Pursuit[J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2007, 62: 78-92

[6] Du Qian, He Yang. Similarity based Unsupervised Band Selection for Hyperspectral Image Analysis[J]. IEEE Geoscience and Remote Sensing Letters, 2008, 5(4): 564-568

[7] 张连蓬. 基于投影寻踪和非线性主曲线的高光谱遥感图像特征提取及分类研究[D]. 青岛: 山东科技大学, 2003

[8] Tenenbaum J, Silva D D, Langford J. A Global Geometric Framework for Nonlinear Dimensionality Reduction[J]. Science, 2000, 290(5500): 2319-

2 323

[ 9] Roweis S, Saul L. Nonlinear Dimensionality Reduction by Locally Linear Embedding [ J]. Science, 2000, 290(5 500): 2 323- 2 326

[ 10] 马瑞,王家璇,宋亦旭. 基于局部线性嵌入(LLE)非线性降维的多流形学习[J]. 清华大学学报(自然科学版), 2008, 48( 4): 582 585

[ 11] Junping Z, Li S Z, Jue W. Manifold Learning and Applications in Recognition in Intelligent Multimedia Processing with Soft Computing[ M]. Heidelberg: Springer Verlag, 2004

[ 12] Mikhail B, Parth N. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation [ J]. Neural Computation, 2003, 15( 6): 1 373- 1 396

[ 13] Zhang Zhenyue, Zha Hongyuan. Principal Manifolds and Nonlinear Dimensionality Reduction Via Tangent Space Alignment [ J]. SIAM Journal of Scientific Computing, 2004, 26(1): 313 338

[ 14] Trevor C, Michael C. Multidimensional Scaling [ M]. London: Chapman & Hall, 1994

[ 15] Bachmann C M, Ainsworth T L, Fusina R A. Exploiting Manifold Geometry in Hyperspectral Imagery[J]. IEEE Trans Geosci and Rem Sens, 2005, 43( 3): 441 454

[ 16] De Sliva V, Tenenbaum J. Global Versus Local Methods in Nonlinear Dimensionality Reduction [ C]/Becker S, Thrun S, Obermayer K. Advances in Neural Information Processing Systems 15. Cambridge: MIT Press, 2002

[ 17] Chang C I. Spectral Information Divergence for Hyperspectral Image Analysis [ OL]. <http://www.google.com>, 2003

第一作者简介: 杜培军, 教授, 博士, 博士生导师, 主要研究方向为遥感图像处理与模式识别、高光谱遥感、资源环境遥感与信息系统。  
E mail: dupjrs@cumt.edu.cn

# Dimensionality Reduction and Feature Extraction from Hyperspectral Remote Sensing Imagery Based on Manifold Learning

DU Peijun<sup>1</sup> WANG Xiaomei<sup>1,2</sup> TAN Kun<sup>1</sup> XIA Junshi<sup>1</sup>

(1 Key Laboratory for Land Resources and Disaster Monitoring of SBSM, China University of Mining and Technology, 1 Daxue Road, Xuzhou 221116, China)  
(2 Survey Engineering Department, Yellow River Conservancy Technical Institute, West Dongjing Road, Kaifeng 475004, China)

**Abstract:** Manifold learning, as the novel nonlinear dimensionality reduction algorithm, is applied to dimensionality reduction and feature extraction of hyperspectral remote sensing information. In order to address inherent nonlinear characteristics of hyperspectral image, Isometric mapping (Isomap), the most popular manifold learning algorithm, is employed to dimensionality reduction of hyperspectral image, and the experimental results show that it outperforms traditional MNF transform. In order to include spectral information into manifold learning, spectral angle (SA) and spectral information divergence (SID), instead of Euclidean distance, are applied to derive the neighborhood distances in Isomap algorithm, and the result is better than that using Euclidean distance in terms of residual variance and normalized spectral eigenvalue. It is concluded that manifold learning is effective to dimensionality reduction and feature extraction from hyperspectral remote sensing imagery.

**Key words:** hyperspectral remote sensing; dimensionality reduction; manifold learning; Isomap

About the first author: DU Peijun, professor, Ph. D, Ph. D supervisor, majors in remote sensing image processing and pattern recognition, hyperspectral remote sensing, resources and environment remote sensing and information system.  
E mail: dupjrs@cumt.edu.cn