

文章编号: 1001-3806(2018)03-0417-05

结合 K-L 散度和互信息的无监督波段选择算法

王琪, 杨枕*, 张俭峰, 向英杰, 田张男
(空军航空大学 航空航天情报系, 长春 130022)

摘要: 波段选择是重要的高光谱图像降维手段。为了达到降维的目的, 提出结合 K-L 散度和互信息的无监督波段选择算法, 并进行了理论分析和实验验证。首先选出信息熵最大的波段作为初始波段, 然后将散度与互信息量的比值定义为联合散度互信息(KLMI)准则, 选择 KLMI 值大且信息量也大的波段加入波段子集中, 选出信息量大且相似度低的波段集合, 最终利用 k 最近邻分类算法实现了基于最大方差主成分分析算法、聚类算法、互信息算法和本文中方法的真实高光谱数据分类实验。结果表明, 本文中的算法总体分类精度和 κ 系数均达到 0.8 以上, 高于其它算法; 大多数地物的分类精度均得到提升, 具有较好的分类性能。该算法是一种实用的高光谱图像降维算法。

关键词: 遥感; 波段选择; K-L 散度; 互信息; 分类

中图分类号: TP751.1 **文献标志码:** A **doi:**10.7510/jgjs.issn.1001-3806.2018.03.024

Unsupervised band selection algorithm combined with K-L divergence and mutual information

WANG Qi, YANG Guang, ZHANG Jianfeng, XIANG Yingjie, TIAN Zhangnan

(Department of Aerospace Intelligence, Aviation University of Air Force, Changchun 130022, China)

Abstract: Band selection is an important method of dimensionality reduction of hyperspectral images. In order to reduce the dimensionality, an unsupervised band selection algorithm combining K-L divergence and mutual information was proposed. And theoretical analysis and experimental verification were carried out. Firstly, the band with the largest information entropy was selected as the initial band. Then, the ratio of divergence to mutual information was defined as the criterion of joint K-L divergence mutual information (KLMI). The band which has large KLMI and information entropy was selected to band subset. Then bands with large information and low similarity were obtained. Finally, the real hyperspectral data classification experiments based maximum-variance principle component analysis (MVP), affinity propagation (AP), mutual information (MI) and the proposed method were realized by using k -nearest neighbor classifier. Experimental results show that, the accuracy of the proposed algorithm is higher than that of other algorithms. The overall classification accuracy and kappa coefficient κ are over 0.8. Classification accuracy of the most objects is improved. The proposed method has outstanding performance on classification and is a practical dimensionality reduction algorithm of hyperspectral image.

Key words: remote sensing; band selection; K-L divergence; mutual information; classification

引言

相比于普通的遥感图像, 高光谱图像增加了光谱维, 不仅包含了丰富的空间信息还具有高光谱分辨率。高光谱传感器能够同时获取目标区域的 2 维几何空间信息与 1 维光谱信息, 因此高光谱数据具有图像立方

体的形式和结构, 体现出图谱合一的特点和优势^[1-2]。高光谱图像丰富的光谱和空间信息对目标探测及地物分类具有重要作用。然而, 高光谱图像维数众多, 导致数据量增加, 相邻波段之间信息冗余度大, 存在“维数灾难”, 使得高光谱影像的降维、目标检测等面临很大的难题, 不利于数据处理^[3]。如何高效充分地利用高光谱图像丰富的信息, 又能提高数据处理的效率, 是目前研究的重点。高光谱图像降维是解决这一问题的常用方法。

针对高光谱图像数据冗余问题, 常用降维方法来解决。波段选择是寻找与强化最具可分性的光谱波段的过程, 它能在不损失重要信息的情况下有效降低维

基金项目: 吉林省教育厅“十二五”科研资助项目(2015448); 吉林省科技发展计划资助项目(20140101213JC)

作者简介: 王琪(1993-), 女, 硕士研究生, 现主要从事高光谱图像波段选择的研究。

* 通讯联系人。E-mail: yg2599@126.com

收稿日期: 2017-06-12; 收到修改稿日期: 2017-06-28

数^[4]。波段选择方法是对高光谱源数据进行波段特征排序搜索和对光谱波段进行特征聚类^[5],是对特定对象选择光谱特征空间中的一个子集,这个子集是一个简化了的光谱特征空间,但它包括了该对象的主要特征光谱,并且在一个含有多种目标对象的组合中,该子集能够最大限度地区别于其它地物。波段选择按是否需要先验知识可分为有监督波段选择和无监督波段选择。有监督波段选择是指已知目标或背景的先验知识条件下,最大限度地保留探测目标的信息。无监督波段选择是指在没有先验知识的情况下,选出信息量大的波段。由于高光谱图像目标或背景等先验信息通常未知,所以无监督波段选择算法具有较强的实用意义。

当前对波段选择的研究有:基于搜索的方法,如结合遗传算法的蚁群算法,这类较为复杂,运算时间较长;基于排序的思想,如最大方差主成分分析(maximum-variance principle component analysis, MVPCA)方法^[6]、聚类(affinity propagation, AP)算法^[7]、基于互信息(mutual information, MI)的方法^[8-9]、基于最大信息量的无监督波段选择算法^[10],这类算法是利用某一指标(如信息量)进行排序忽略了波段间的相似性、冗余度,导致选择结果不准确。为了提高波段选择精度,本文中提出了一种结合 K-L (Kullback-Leibler)散度和互信息的无监督波段选择算法,利用 K-L 散度和互信息的比值定义了联合 K-L 散度-互信息(joint K-L divergence-mutual information, KLMI)准则。首先按信息熵大小对波段排序,然后通过最大化波段间 K-L 散度和最小化波段间互信息量,选出高信息量低相似度的波段。本文中的算法既能保留原始波段的有用信息特征又考虑了波段间差异性,有利于提高目标探测效率和分类精度。

1 算法描述

1.1 K-L 散度

K-L 散度同光谱相关系数、光谱角制图、离散度都是常见的相似性度量算法,分别从不同角度衡量波段间的相似性。K-L 散度是一种在信息论中得到广泛应用的信号相似性度量^[11]。 P 和 Q 是两个离散随机信号的概率分布函数, $P = [P_1, \dots, P_i, \dots, P_N]^T$, $Q = [Q_1, \dots, Q_i, \dots, Q_N]^T$,定义 Q 相对于 P 的 K-L 散度为:

$$D_{K-L}(P \cdot Q) = \sum_{i=1}^N P_i \ln \frac{P_i}{Q_i} \quad (1)$$

当且仅当 $\sum_{i=1}^N P_i = \sum_{i=1}^N Q_i = 1$ 时,(1)式成立。

(1)式表示:用 Q 中元素表示 P 中元素所需的额外的信息量的大小。K-L 散度越大,说明用 Q 中元素表示 P 中元素的难度越大,即 Q 和 P 之间相似度越低。从本质上讲,K-L 散度是一种概率意义上的非对称距离,也可以理解为两个信号所包含的信息量之间的差^[12]。

高光谱图像的每一个波段可以看成是一个随机变量,因此可以利用 K-L 散度衡量两波段之间的相似性。假设一组高光谱图像数据共有 L 个波段,表示为 $X = [x_1, x_2, \dots, x_L]$, $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,N}]^T$ 是第 i 个波段列向量,其中 N 为像素个数。对 x_i 进行归一化处理,得到 $\bar{x}_i = \frac{x_i}{\sum_{n=1}^N x_{i,n}}$, $x_{i,n}$ 为第 i 个波段列向量的第 n

个分量。因此第 j 波段相对于第 i 个波段的 K-L 散度为:

$$D_{K-L}(\bar{x}_i \cdot \bar{x}_j) = \sum_{n=1}^N x_{i,n} \ln \frac{\bar{x}_{i,n}}{\bar{x}_{j,n}} \quad (2)$$

由(2)式可知,K-L 散度越大,用第 j 波段表示第 i 个波段的难度越大,两波段间信息量之差越大,相似度越低。

1.2 互信息

互信息是信息论中的一个概念,它描述了两个系统的统计相关性,或者说一个系统存在于另一个系统中的信息量^[13],可表示为:

$$I(X, Y) = \int_X \int_Y P(x, y) \log_2 \frac{P(x, y)}{P(x)P(y)} dx dy \quad (3)$$

式中, $P(x, y)$ 是随机变量 X 和 Y 的联合概率密度, $P(x)$, $P(y)$ 分别是变量 X 和 Y 的边缘概率密度。

互信息的实质是当某一随机变量已知的情况下,另一随机变量的不确定性的减小值,不仅可以衡量两随机变量共有信息量的多少,还可以衡量随机变量之间的相关程度。MI 值越大,两变量之间的共有信息越多,相关度越大;当两变量相互独立时,MI 的值为0。

对于高光谱图像而言,给定的两波段图像 X 和 Y ,由(3)式可得图像 X 和 Y 的互信息为:

$$I(X, Y) = \int_X \int_Y P_{i,j}(x, y) \log_2 \frac{P_{i,j}(x, y)}{P_i(x)P_j(y)} dx dy \quad (4)$$

式中,变量 x 为图像 X 中的元素,变量 y 为图像 Y 中的元素, $P_i(x)$, $P_j(y)$ 为变量 x 和变量 y 在第 i 个状态和第 j 个状态下的边缘概率密度, $P_{i,j}(x, y)$ 为变量 x, y 的联合概率密度。

1.3 联合散度互信息 (KLMI) 的波段选择

对于高光谱图像波段选择而言,既要满足信息量最大的要求又要使地物目标之间的可分性最大化。为了选取高信息量的波段,需要计算波段信息量并进行排序,将排在前面的单个波段进行组合获得最优波段集合。衡量波段信息量的大小可以利用 K-L 散度指标。因此,可以选出 K-L 散度值较大的波段组成集合,根据(2)式最大化 K-L 散度可用如下公式表示:

$$\max \left(\frac{1}{N} \sum_{n=1}^N x_{i,n} \ln \frac{\overline{x_{i,n}}}{x_{j,n}} \right) \quad (5)$$

但是,高信息量的波段组合并不一定能最大程度上保持图像的原始波段信息,因为这些波段可能是邻近波段,相似性很大,提供的额外信息量几乎没有。有的波段虽然信息量不是很丰富,但是能提供与其它波段互补的信息。因此,就需要选择信息互补且相似度较低的波段。任意波段间的相似度可用互信息 $I(X_i, X_j)$ ($1 \leq i, j \leq N$) 衡量,根据(4)式,最小化互信息可用如下公式表示:

$$\min \left[\frac{2}{N^2 - N} \sum_{1 \leq i, j \leq N} I(X_i, X_j) \right] \quad (6)$$

将(5)式和(6)式做比值,定义了联合散度互信息 (KLMI) 准则:

$$\max \left(\frac{(N-1) \sum_{n=1}^N x_{i,n} \ln \frac{\overline{x_{i,n}}}{x_{j,n}}}{2 \sum_{1 \leq i, j \leq N} I(X_i, X_j)} \right) \quad (7)$$

由上述分析可知, KLMI 值用于衡量各个波段之间的相似度, KLMI 值越大,说明波段间相似度、冗余性越低,在该波段集合上的信息互补性和可分性越好。

本文中提出的 KLMI 波段选择方法,通过最大化 K-L 散度和最小化波段之间的互信息来进行波段选择。首先计算原始波段中信息熵最大的波段作为初始波段,然后选择与初始波段的 KLMI 最大且自身信息熵也越大的波段作为第 2 个波段,组成波段子集。然后选择与波段子集中所有波段 KLMI 值越大且自身信息熵越大的波段归入所选波段子集,以此类推,选择包含 k 个波段的子集作为结果。

1.4 算法步骤

本文中提出的 KLMI 波段选择算法步骤主要由 4 步组成:(1)利用信息熵指标选择信息量最大的波段 b_1 作为初始波段,产生波段子集 $\Phi_1 = \{b_1\}$;(2)利用联合散度互信息 (KLMI) 准则计算剩余波段与 Φ_1 中元素的相似度,找出 KLMI 值最大时对应的波段,同时

考虑波段自身的信息熵,选择 KLMI 值越大且信息熵大的波段 b_2 ,将 b_2 波段加入子集 Φ_1 中,即 $\Phi_2 = \Phi_1 \cup \{b_2\}$;(3)循环步骤(2),直到波段集合 Φ 中波段的数量满足分类需求;(4)更新波段集合 Φ , Φ 中元素即为波段选择结果。

2 实验验证

2.1 实验数据

实验中采用的是于 1992 年由 AVIRIS 传感器获取的高光谱图像数据 Indian Pines,成像地区为美国印第安纳西北地区,该数据共 220 个波段,波长范围为 $0.4\mu\text{m} \sim 2.5\mu\text{m}$,光谱分辨率为 10nm,空间分辨率为 17m,图像大小为 $145\text{pixel} \times 145\text{pixel}$,该数据常用于高光谱图像的分类研究。去除水的吸收带和噪声波段,处理后保留了 200 个有效波段用于本实验。Purdue 大学给出了一份关于该地区的实地调查报告^[14]。该地区主要被农作物(约占总面积 2/3,包括大豆、玉米、小麦、干草堆)和植被(约占总面积的 1/3,包括树林、草地等)所覆盖。除农作物与植被外,还有铁路、公路、高速公路、房屋和无线电发射塔等地物。实验数据的假彩色合成和地面真实标记如图 1 所示。

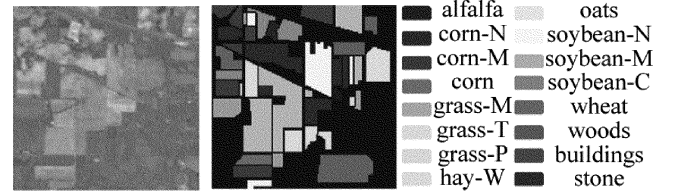


Fig. 1 Synthetic false color image and real markings image

2.2 波段选择结果

根据本文中算法步骤,进行实验验证,计算图像各波段的信息熵,得出第 112 波段熵值最大,因此选择第 112 波段作为初始波段 b_1 ,最终选出 10 个波段。同时为了更好地比较本文中算法性能,实验还实现了常用的 MVPCA 方法、AP 方法、MI 方法及 K-L 散度方法的波段选择结果。然后计算各波段选择算法结果中各波段的信息熵的总和,信息熵总和及波段选择结果如表 1 所示。

由表 1 可知,本文中算法选择的波段信息熵之和最大,即信息量最大,而 AP 算法多所选波段包含的信息量最少。同时 MVPCA, MI, AP 及 K-L 散度方法 4 种方法得到的结果,波段比较集中,说明它们之间的冗余度较大;本文中方法所选波段分布范围较广,且远离光谱范围的边缘,效果较好。本文中方法所选前 10 个波段如图 2 所示。

Table 1 Comparison of band selection results and information

band selection algorithm	MVPCA	MI	AP	K-L	proposed algorithm
band selection results	8,20,36,62,63,69,70,72,92,95	9,32,30,23,13,89,57,97,100,110	112,113,31,114,121,125,156,127,164,51	15,49,33,87,21,63,101,99,71,56	112,42,31,40,108,51,120,150,21,9
sum of entropy	67.1277	67.4139	67.0094	67.1058	67.7815

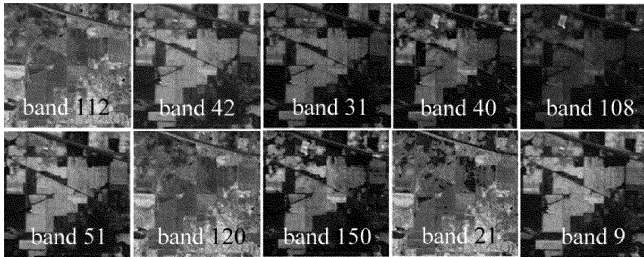


Fig. 2 The results of band selection

2.3 分类结果

Purdue 大学给出的 Indian Pines 数据实地调查指出该区域有 16 类不同地物,每一类地物中,分别随机选取 5% 和 10% 数量的样本作为训练样本,剩余样本作为测试样本,将波段选择后的结果进行分类实验,本实验中采用 k 最近邻 (k -nearest neighbor, KNN) 分类算法,经过多次实验反复调整 k 值,选取最优近邻个数为 $k=7$ 时分类精度最高且噪声数据干扰降到最低。各算法总体精度与波段维数的关系如图 3 所示,横坐标为波段维数,纵坐标为分类精度。总体分类精度公式为:总体分类精度 (overall accuracy, OA) = 正确分类的像元数/像元总数。

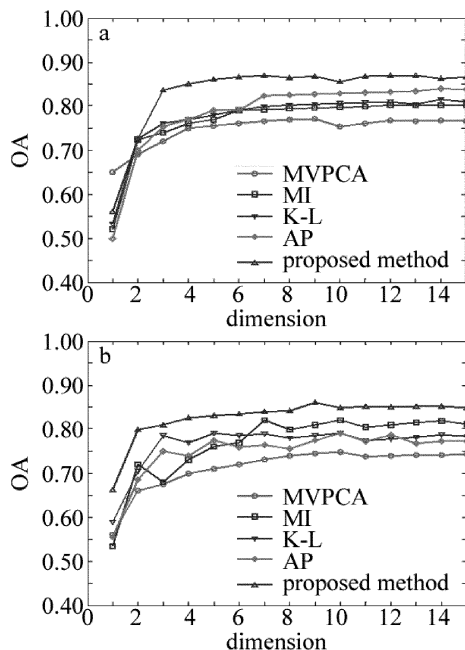


Fig. 3 Overall classification accuracy of different samples with various bands
a—5% samples b—10% samples

由图 3 可知,随着训练样本数量的增加,目标的先验知识就越多,分类的性能越好,分类的总体精度也就越高。在 5% 和 10% 的样本比例下,且维数大于 2 时,本文中算法相比于其它 3 种波段选择算法均取得了较好的分类效果。在样本比例仅为 5% 的条件下,本文中算法的分类精度就能达到 82%。

表 2、表 3 中给出了训练样本分别为 5% 和 10% 的情况下,不同算法中各类地物的分类精度、总体精度 (OA) 及 κ 。 κ 的计算公式^[15]为: $\kappa = (\text{总体精度} - \text{期望精度}) / (1 - \text{期望精度})$ 。

Table 2 Classification accuracy of various types of objects (5% samples)

class	training samples	test samples	MVPCA	MI	AP	algorithm in this paper
alfalfa	23	23	0.53	0.43	0.38	0.94
corn-N	34	1394	0.49	0.65	0.64	0.71
corn-M	27	803	0.49	0.55	0.64	0.69
corn	25	212	0.37	0.40	0.47	0.66
grass-M	27	456	0.80	0.83	0.91	0.94
grass-T	26	704	0.90	0.90	0.92	0.99
grass-P	14	14	0.57	0.56	0.86	0.86
hay-W	27	451	0.97	0.98	0.98	1.00
oats	10	10	0.27	0.36	0.60	0.93
soybean-N	30	942	0.62	0.63	0.76	0.84
soybean-M	31	2424	0.64	0.77	0.79	0.81
soybean-C	28	565	0.38	0.52	0.47	0.58
wheat	25	180	0.91	0.89	0.98	0.99
woods	30	1235	0.90	0.93	0.94	0.96
buildings	27	359	0.38	0.53	0.47	0.77
stone	26	67	0.86	0.86	0.90	0.96
OA			0.65	0.71	0.76	0.82
κ			0.61	0.69	0.70	0.81

$\kappa > 0.8$ 时说明分类精度高, κ 在 0.4 ~ 0.8 之间,分类精度中等, $\kappa < 0.4$ 时,分类效果较差。在相同样本比例情况下,本文中方法的 OA 及 κ 均优于其它 3 种算法,分类效果比其它 3 类算法效果好。这是因为本文中算法采用 KLMI 准则进行波段选择,实现了波段的高信息量及波段间的低冗余度,使得同类别数据间的相似性突出,且不同类别的数据之间的差异性更加明显。同时,对于大多数地物,本文中方法的分类精度优于其它算法,但是存在少数地物分类精度不高。

Table 3 Classification accuracy of various types of objects (10% samples)

class	training samples	test samples	MVPCA	MI	AP	algorithm in this paper
alfalfa	23	23	0.28	0.71	0.7	0.79
corn-N	89	1428	0.64	0.60	0.74	0.74
corn-M	73	830	0.75	0.60	0.71	0.81
corn	66	237	0.49	0.49	0.49	0.64
grass-M	71	730	0.96	0.86	0.88	0.82
grass-T	81	478	0.73	0.92	0.94	0.98
hrass-P	14	28	0.47	0.57	0.83	0.62
hay-W	71	478	0.89	0.96	0.99	0.99
oats	10	10	0.44	0.47	0.58	0.47
soybean-N	76	896	0.76	0.74	0.71	0.72
soybean-M	112	2343	0.66	0.71	0.85	0.87
soybean-C	68	525	0.52	0.47	0.72	0.78
wheat	67	138	0.97	0.93	0.94	0.98
woods	89	1176	0.86	0.90	0.96	0.99
buildings	68	318	0.72	0.45	0.62	0.91
stone	47	46	0.20	0.91	0.88	0.91
OA			0.71	0.72	0.81	0.85
κ			0.63	0.69	0.75	0.82

随着训练样本数量的增加,各类地物的分类精度、OA 和 κ 均有所提高。综上所述,本文中算法取得了较好的分类效果,性能优于其它 3 种算法。

3 结 论

本文中提出了一种结合 K-L 散度和互信息的无监督波段选择算法,将 K-L 散度与互信息的比值定义为新的联合散度互信息准则,通过最大化散度和最小化波段间的互信息量,从原始波段中选出高信息量且低相似度的波段集合。首先基于信息熵选择初始波段,然后利用 KLMI 准则进行后续波段的选择,最终将本文算法及 MVPCA, AP, MI 这 3 种无监督波段选择算法应用于高光谱图像分类实验。实验结果表明,本文中算法相比于其它算法在分类实验中取得了较好的效果,算法具有较为优越的性能。该算法存在的不足为波段选择数量阈值需要人为设定,算法没有实现自动确定阈值功能,这是下一步研究的重点,进一步提高算法的智能性和自动化水平。

参 考 文 献

- [1] ZHANG B. Current progress of hyperspectral remote sensing in China [J]. Journal of Remote Sensing, 2016, 20(5):1062-1090 (in Chinese).
- [2] BIOUCAS-DIAS J M, PLAZA A, CAMPS-VALLS G, *et al.* Hyperspectral remote sensing data analysis and future challenges[J]. IEEE Geoscience & Remote Sensing Magazine, 2013, 1(2):6-36.
- [3] XIANG Y J, YANG G, ZHANG J F, *et al.* Dimensionality reduction for hyperspectral imagery manifold learning based on spectral gradient angles[J]. Laser Technology, 2017, 41(6):921-926 (in Chinese).
- [4] QIN F J, ZHANG A W, WANG Sh M, *et al.* Hyperspectral band selection based on spectral clustering and inter-class separability factor [J]. Spectroscopy and Spectral Analysis, 2015, 35(5):1357-1364 (in Chinese).
- [5] TAN Y Q. Unsupervised band selection for hyperspectral image based on multiobjective optimization [D]. Xi'an: Xidian University, 2014: 19-21 (in Chinese).
- [6] MYRONENKO A, SONG X. Point set registration: Coherent point drifts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(12):2262-2275.
- [7] CHENG H, WANG Zh Q, ZHANG Y Y. Research on geometric rectification of aerial images [J]. Journal of Northeast Normal University, 2009, 41(3):50-54 (in Chinese).
- [8] BATTITI R. Using mutual information for selecting features in supervised neural net learning [J]. IEEE Transactions on Neural Networks, 1994, 5(4):537-550.
- [9] KWAK N, CHOI Ch H. Improved mutual information feature selector for neural networks in supervised learning [C]//Proceeding of 1999 International Joint Conference on Neural Networks. New York, USA: IEEE, 1999:1313-1318.
- [10] LIU X S, GE L, WANG B, *et al.* An unsupervised band selection algorithm for hyperspectral imagery based on maximal information [J]. Journal of Infrared and Millimeter Waves, 2012, 31(2):166-176 (in Chinese).
- [11] SHI Y, EBERHART R. A modified particle swarm optimizer [J]. IEEE Xplore, 1998, 7803(5):69-73.
- [12] KULLBACK S. Information theory and statistics [M]. New York, USA: John Wiley and Sons, 1959:55-70.
- [13] ZHOU Y, LI X R, ZHAO L Y. Modified linear-prediction based band selection for hyperspectral image [J]. Acta Optica Sinica, 2013, 33(8):0828002 (in Chinese).
- [14] LANDGREBE D. Multispectral data analysis: A signal theory perspective [R]. West Lafayette, USA: Purdue University, 1998:56-89.
- [15] YANG J, HUA W Sh, LIU X, *et al.* Band selection algorithm for hyperspectral imagery based on K-L divergence and spectral divisibility distance [J]. Journal of Applied Optics, 2014, 35(1):71-75 (in Chinese).