

空间人造目标混合光谱材料数目确定方法仿真研究

李庆波, 苗兴晋

北京航空航天大学仪器科学与光电工程学院, 精密光机一体化技术教育部重点实验室, 北京 100191

摘要 在空间人造目标光谱分析领域, 受到观测距离和观测设备空间分辨率的限制, 通常在观测空间人造目标光谱信号时, 目标某个瞬时视场中的多种纯物质材料的光谱特征信息组合在一个像元中, 形成“混合光谱”。因此, 将这些混合光谱分解为每个单一材料的光谱并估计出相应的组成比例是空间人造目标光谱分析研究的重点。大多数现有空间目标光谱分解方法都假设空间人造目标混合光谱中包含的纯物质材料种类个数(即“端元数目”)是先验已知的, 这对于未知空间人造目标而言是不现实的。因此, 纯物质材料数目正确估计对后续光谱数据分析处理的准确性起着至关重要的作用。目前, 现有的端元数目确定方法的设计均在高斯白噪声的假设下进行, 而对于噪声信号的分布存在频谱相关性的情况下, 会提供较差的结果。采用一种基于数据内在维度和似然最大化理论的方法——鲁棒特征值极大似然方法。由于数据内在维数与信号协方差矩阵和信号相关矩阵特征值差异的统计分布特性高度相关, 因此通过分析该特征值差异的统计分布特性, 构建一个极大似然函数, 可以实现空间人造目标混合光谱端元数目的确定。该方法包含两个步骤: 首先, 采用基于多元回归和改进最小噪声分离方法对原始光谱数据进行预处理完成噪声特性估计和噪声白化过程, 从而有效抑制具有频谱相关性的噪声的干扰; 接下来, 通过求解一个离散对数联合似然函数的极大值问题来实现空间人造目标混合光谱端元数目的确定, 该方法完全不需要输入任何参数, 并且运行速度比较快。分别利用实验室实测的五种空间人造目标材料的可见/近红外光谱数据和美国地质勘测局光谱数据构建混合光谱仿真数据进行实验。结果表明, 该方法能有效抑制相关噪声和白噪声的干扰, 空间人造目标纯物质材料数目确定结果具有很好的准确性和稳定性。

关键词 空间人造目标; 光谱分析; 端元数目确定; 内在维度

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)05-1607-05

引言

随着航天领域科学技术的不断发展, 空间人造目标的观测与识别具有重要的研究意义和价值。空间人造目标主要包括大气外绕地运行的各种在轨工作或者停止工作的航天器、人造卫星等。空间人造目标光谱观测是利用光谱仪获取目标的光谱信息, 可以根据目标光谱特征来识别分析目标的材料信息。光谱观测系统灵活、方便, 对空间人造目标观测具有重要意义。由于观测距离和观测设备空间分辨率的限制, 空间人造目标光谱通常是一个“混合光谱”, 为空间人造目标的某个瞬时视场中多种纯物质材料的光谱特征的组合^[2]。因此, 在利用光谱信息分析空间人造目标材料组成和特性之前, 需要确定混合光谱中包含的纯物质材料的数目, 即, “端元数目”。尽管端元数目确定方法受到了广泛关注和研究,

但其仍然是一个具有挑战性的研究领域。

目前, 针对混合光谱中端元数目的确定, 很多算法被开发出来。最著名的是一种基于“虚拟维度”(virtual dimension, VD)的方法^[3], 称为噪声白化(noise whitening)的 Harsanyi-Farrand-Chang(NWHFC)方法。其依赖于 Neyman-Pearson 检测理论, 根据一个输入的虚警概率进行决策。该方法的缺点是需要人工输入一个错误警报参数, 最终结果对该参数非常敏感。另一种方法是最小误差光谱信号子空间识别方法(hyperspectral signal subspace identification by minimum error, HySime)^[4]。该方法不需要输入参数, 在满足最小均方根误差的条件下, 确定包含数据有效信息的最佳线性子空间的维数。但是该方法在信噪比较低、输入光谱数据维度较高时会有较大误差。此外, 有学者提出了一种基于随机矩阵理论(random matrix theory, RMT)的方法^[5]。该方法基于 RMT 的最新研究, 根据一个理论阈值以区分信号协方差矩

收稿日期: 2020-05-19, 修订日期: 2020-09-16

基金项目: 国家自然科学基金项目(61575015)资助

作者简介: 李庆波, 女, 1975年生, 北京航空航天大学仪器科学与光电工程学院副教授 e-mail: qbleebuaa@buaa.edu.cn

阵中由噪声引起的特征值和由信号引起的特征值。但是该方法需要一个由光谱数据统计分布信息所估计的阈值,当该阈值估计误差较大时,算法结果的误差也大。因此,本文提出了一种鲁棒特征值极大似然(robust eigenvalue maximum likelihood, REML)方法实现空间人造目标混合光谱端元数目的确定,该方法能够有效抑制高斯白噪声和频谱相关噪声的干扰,且不依赖于任何输入参数和固定阈值。

1 实验部分

1.1 原理与算法

1.1.1 数据预处理

传统的端元数目确定方法假定噪声为独立同分布的高斯噪声,这对于实际的空间人造目标光谱是不正确的,其往往会受到频谱相关噪声的干扰。因此,需要一个预处理步骤——噪声特性估计和噪声白化,来抑制频谱相关噪声的干扰。本文采用基于多元回归的方法来估计噪声协方差矩阵^[6-7],该方法当存在非高斯白噪声时,效果较好。接下来,使用一个噪声白化步骤将光谱数据转换到一个新的主成分空间来实现抑制频谱相关噪声的影响,具体方法如下:

对于具有 L 个频带的 N 条光谱的原始光谱数据矩阵 $\mathbf{P} \in R^{(L \times N)}$,利用正交分解方法将噪声协方差矩阵估计值 $\hat{\mathbf{S}} \in R^{(L \times L)}$ 分解,得到一个由噪声协方差矩阵特征向量组成的正交矩阵 $\mathbf{K}_P \in R^{(L \times L)}$ 。利用该矩阵对原始光谱数据进行正交变换

$$\mathbf{F} = \mathbf{P}^T \mathbf{K}_P \quad (1)$$

随后,将每个频段的数据(即 \mathbf{F} 的每列)除以相应频段内的噪声的标准差的估计值 $\{\hat{\epsilon}_l\}_{l=1}^L$,这个估计值是由 $\hat{\mathbf{S}}$ 的对角线上的值确定的。可以得到矩阵

$$\begin{aligned} \tilde{\mathbf{F}} &= \mathbf{F} ./ (\mathbf{1}_N \mathbf{e}^T) \\ \mathbf{e} &= [\hat{\epsilon}_1, \hat{\epsilon}_2, \dots, \hat{\epsilon}_L]^T \end{aligned} \quad (2)$$

其中 $./$ 代表矩阵(向量)元素的除法。这就得到了在每个频带中具有相等噪声方差的转换数据,相当于在转换空间中噪声被白化了。接下来计算矩阵 $\tilde{\mathbf{F}}^T$ 的协方差矩阵,同样利用正交变换将该矩阵分解得到的正交矩阵 $\mathbf{K}_F \in R^{(L \times L)}$ 并进行正交变换

$$\tilde{\mathbf{P}} = (\tilde{\mathbf{F}} \mathbf{K}_F)^T \quad (3)$$

则经过上述变换后,原始光谱数据即被转换到一个新的主成分空间中,转换光谱数据由频带不相关的噪声组成,且该方法允许在噪声协方差矩阵估计过程中存在一定误差。

1.1.2 特征值极大似然方法

对于预处理后的空间人造目标混合光谱数据矩阵而言,令 $\{\alpha_l\}_{l=1}^L$ ($\alpha_l \geq \alpha_{l+1}$) 和 $\{\beta_l\}_{l=1}^L$ ($\beta_l \geq \beta_{l+1}$) 分别为其样本的协方差矩阵和相关矩阵的特征值。假设信号源的数目是非随机的未知正常实数,而噪声信号经过上述预处理过程后近似为高斯白噪声,且光谱净信号部分与噪声信号部分之间是相互独立的。根据对信号协方差矩阵和信号相关矩阵中特征值统计分布的相关研究结果^[8-9],对于第 l 个特征值(即第 l 个频带)可以得到

$$p(\beta_l - \alpha_l | 1 \leq l \leq M) \approx N(0, \delta_l^2)$$

$$p(\beta_l - \alpha_l | M < l \leq L) \approx N(\mu_l > 0, \delta_l^2) \quad (4)$$

$$l = 1, 2, \dots, L$$

其中, M 为混合光谱所包含的纯物质材料的数目, μ_l 是未知常数, δ_l^2 的近似值由式(5)给出^[10]

$$\delta_l^2 \approx 2(\alpha_l^2/N + \beta_l^2/N) \quad (5)$$

上述统计分布对应的离散的对数联合似然函数见式(6)

$$LF(l) = \log \left(\prod_{i=1}^L \exp(-m_i^2/2\delta_i^2) / \prod_{i=1}^L \delta_i \right), m_i = \beta_i - \alpha_i \quad (6)$$

该对数联合似然函数将会在 $l=M+1$ 时取得最大值,则混合光谱所包含的端元数目的估计值见式(7)

$$\hat{M} = \underset{1 \leq l \leq L}{\operatorname{argmax}} (LF(l)) - 1 \quad (7)$$

因此,该方法不需要任何输入参数和固定阈值即可完成空间人造目标混合光谱端元数目自动确定。

1.2 方法

1.2.1 仿真实验一

为了分析评价本文提出的空间人造目标混合光谱端元数目自动确定方法(简称 REML),使用实验室测得的五种空间人造目标常用的材料(白漆、太阳能电池板、聚酰亚胺、铝箔、外包装)的可见/近红外光谱特征作为纯物质材料端元光谱进行仿真实验。实验装置示意图如图 1 所示,实验所使用的光源为 Laser-1 卤钨灯,光谱仪选用美国海洋光学公司的 USB4000 光纤光谱仪;使用 QR400-7-VIS-BX 型 Y 型光纤探头测量材料的反射光谱,测量波长范围是 400~1 000 nm,光谱波长分辨率为 0.2 nm。测量参数设置为:积分时间 12

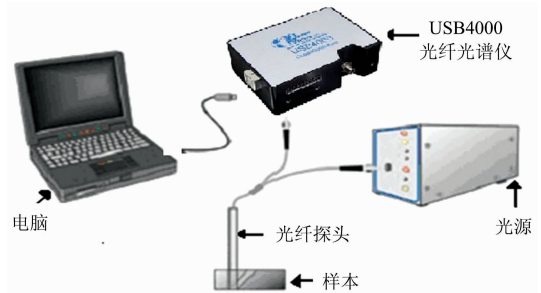


图 1 空间人造目标材料光谱测量实验示意图

Fig. 1 Schematic of spectral signatures measurement experiment of artificial space target materials

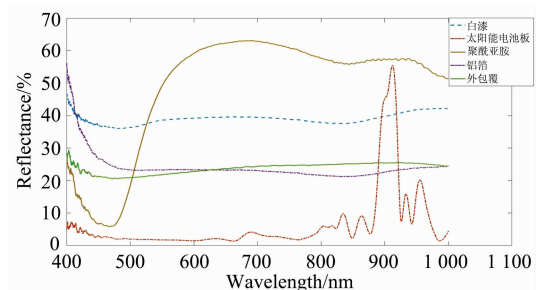


图 2 五种空间人造目标材料的光谱特性

Fig. 2 Spectral signatures of five artificial space target materials

ms, 平滑度 10, 平均次数 16, 每条光谱数据包含 3 091 个波段, 实验测得光谱结果如图 2 所示。

在获取到空间人造目标纯物质材料实验室光谱特性后, 根据迪利克雷(Dirichlet)分布函数生成相应比例参数, 将这五种材料光谱按该比例参数进行混合, 从而形成空间人造目标混合光谱仿真数据集, 共 4 000 条混合光谱。对该数据集分别添加 20 dB, 30 dB 的高斯白噪声和频谱相关噪声, 以模拟实际光谱获取过程中所受到的噪声干扰。将本文提出的 REML 方法和一些常用的空间人造目标端元数目确定方法: 噪声白化的 NWHFC 方法^[3], HySime 方法^[4], 特征间隔法(noise-whitened eigengap approach, NWEGA)^[5], 进行比较, 为了探究算法的准确性和稳定性, 实验中每种算法均重复运行 50 次, 结果取 50 次运行的平均值, 保留两位小数。

1.2.2 仿真实验二

由于未知空间人造目标的组成材料种类和数目多变, 并且不完全已知, 因此, 需要进一步探究本文所提出的方法在空间人造目标混合光谱端元数目变化的鲁棒性。本实验使用美国地质勘测局(United States Geological Survey, USGS)光谱数据库中随机选择 3, 5, 7 和 9 条光谱数据作为纯物质材

料光谱构建混合光谱仿真数据集, 其包含 224 个波段(波长范围是 370~2 500 nm), 共 5 000 条混合光谱。为了模拟数据采集过程, 将具有不同信噪比的频谱相关噪声分别添加到仿真混合光谱数据中。同样, 将本文提出的 REML 方法和一些常用的空间人造目标端元数目确定方法(NWHFC, HySime以及 NWEGA)进行比较, 实验中每个方法均重复运行 50 次, 结果取 50 次运行的平均值, 保留两位小数。

2 结果与讨论

2.1 第一组仿真实验的结果

表 1 和表 2 显示了 NWHFC, HySime, NWEGA 以及本文提出的 REML 算法在不同信噪比、不同类型噪声干扰下对由五种空间人造目标材料生成的混合光谱数据的端元数目的估计结果。其中 NWHFC_1, NWHFC_2 和 NWHFC_3 分别表示输入虚预警参数分别为 $\{10^{-3}, 10^{-4}, 10^{-5}\}$ 的 NWHFC 算法。图 3 表示在信噪比为 20dB 时不同类型的噪声干扰下, REML 算法的对数联合似然函数值关于特征值指数的变化。

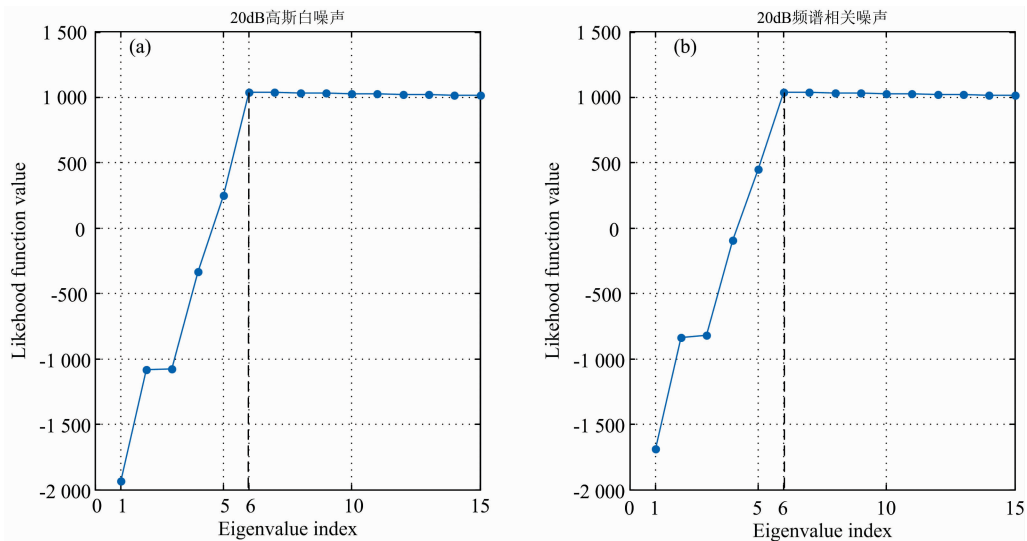


图 3 信噪比为 20 dB 时不同类型噪声干扰时 REML 的对数联合似然函数值

Fig. 3 Logarithmic joint likelihood function values of REML for different types of noise (SNR: 20 dB)

表 1 高斯白噪声干扰下各个算法实验结果比较

Table 1 Comparison of experimental results of various algorithms under the Gaussian white noise

信噪比/dB	REML	HySime	NWHFC_1	NWHFC_2	NWHFC_3	NWEGA
20	5.00	25.82	4.98	4.98	4.84	5.36
30	5.00	15.20	5.00	5.00	4.94	5.36

表 2 频谱相关噪声干扰下各个算法实验结果比较

Table 2 Comparison of experimental results of various algorithms under the spectrally correlated noise

信噪比/dB	REML	HySime	NWHFC_1	NWHFC_2	NWHFC_3	NWEGA
20	5.00	28.88	5.00	5.00	4.84	5.40
30	5.00	22.78	5.00	5.00	5.00	5.36

从表中实验结果可以看出,对于不同种类和信噪比的噪声,HySime 算法的实验结果出现较大误差。REML 算法和 NWHFC 算法对端元数目的确定均取得了较好的实验结果。但是,当信噪比较低时,NWHFC 算法的效果有所下降;对于 NWEGA 算法而言,其运行结果的稳定性较差,这可能是由于其存在一个基于数据统计信息(易受噪声干扰)的硬阈值导致了计算结果的波动性。另外,从图 3 中可以看出,REML 中的对数联合似然函数的值在不同类型噪声干扰下均能够在正确的特征值指数下取得最大值。因此,本文所采

表 3 端元数目不同时频谱相关噪声干扰下各个算法实验结果比较

Table 3 Comparison of experimental results of various algorithms under the spectrally correlated noise with different numbers of endmembers

信噪比 /dB	方法	端元数目			
		M=3	M=5	M=7	M=9
20	REML	3.00	5.00	7.00	8.96
	HySime	91.76	91.32	91.18	90.70
	NWHFC_1	3.00	5.00	4.90	7.76
	NWHFC_2	3.00	5.00	4.58	7.48
	NWHFC_3	3.00	5.00	4.28	7.14
	NWEGA	12.76	14.60	15.94	17.08
30	REML	3.00	5.00	7.00	9.00
	HySime	90.86	90.36	89.88	89.38
	NWHFC_1	3.00	5.00	5.08	8.00
	NWHFC_2	3.00	5.00	4.94	8.00
	NWHFC_3	3.00	5.00	4.80	8.00
	NWEGA	14.08	14.70	15.06	7.22

用的 REML 算法则展现出对频谱相关噪声和白噪声干扰均有较强的鲁棒性。

2.2 第二组仿真实验的结果

第二组实验旨在探究不同算法对混合光谱端元数目变化的鲁棒性,实验结果如表 3 所示,为噪声干扰为频谱相关噪声的实验结果。

就端元数目变化对算法的影响而言,当混合光谱包含的纯物质材料数目变多时,NWHFC 出现了低估现象,NWEGA 存在高估和低估现象,表明这两种算法难以区分出不同信号源的差异性。当虚预警参数发生变化时,NWHFC 算法的实验结果也发生变化,这也是该算法一个局限性。另外,当噪声信号存在频谱相关性时,HySime 算法出现了严重的高估现象,完全无法使用。REML 算法的实验结果具有很好的准确性和稳定性,展现出对端元数目变化较强的鲁棒性。

3 结 论

提出了一种用于确定空间人造目标混合光谱纯物质材料数目的全自动方法,该方法可以有效解决空间人造目标光谱观测过程中所存在的频谱相关噪声的干扰问题。其优势和特点如下:首先,该方法无输入参数,且不基于固定阈值;其次,当光谱信号受到白噪声和频谱相关噪声的干扰时,由于利用了一个有效且允许有一定估计误差的噪声特性估计和噪声白化方法,能够有效抑制噪声对计算结果的干扰。实验结果表明,该方法能够准确且稳定地确定空间人造目标混合光谱中纯物质材料的数目。因此,该方法在利用光谱信息进行空间人造目标材料分析领域具有广阔应用前景。

References

- [1] WANG Xue-yao(王雪瑶). Spacecraft Engineering(航天器工程), 2018, 27(3): 86.
- [2] LI Qing-bo, WU Ke-jiang, NIU Chun-yang(李庆波, 吴科江, 牛春阳). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(9): 2679.
- [3] Chang Chein-I, Du Q. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(3): 608.
- [4] Sánchez S, Plaza A. Journal of Real-Time Image Processing, 2014, 9(3): 397.
- [5] Halimi A, Honeine P, Kharouf M, et al. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(7): 3811.
- [6] Tratar L F, Strmcnik E. Energy, 2016, 109(15): 266.
- [7] Andreou C, Karathanassi V. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(1): 247.
- [8] Raskutti G, Yuan M, Chen H. Annals of Statistics, 2019, 47(3): 1554.
- [9] Ceruti C, Bassi S, Rozza A, et al. Pattern Recognition, 2014, 47(8): 2569.
- [10] Luo B, Chanussot J, Doute S, et al. IEEE Geoscience and Remote Sensing Letters, 2013, 10(1): 24.

Simulation Research on the Method of Determining Material Number of Artificial Space Target Mixed Spectra

LI Qing-bo, MIAO Xing-jin

Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China

Abstract When observing the spectral signal of an artificial space target, because of the long observation distance and the low spatial resolution of observation equipment, the spectral signatures of multiple pure materials in a certain instantaneous scene is combined in one pixel to form a “mixed spectrum”. Therefore, unmixing these mixed spectra into the collection of pure material spectra and estimating the corresponding fractional abundances have been increasingly significant in the field of spectral analysis for artificial space targets. Most existing spectral unmixing methods assume that the number of pure materials (that is, “the number of endmembers”) contained in mixed spectra of an artificial space target is known a priori, which is unrealistic for unknown artificial space targets. Therefore, the exact estimation of the number of endmembers plays a significant role in the accuracy of subsequent spectral analysis and processing. At present, the existing methods of endmember number estimation are mostly proposed under the assumption of Gaussian white noise interference. However, when the distribution of the noise signal is a spectral correlation, poor estimation results will be provided. In this paper, based on the intrinsic dimensions of data and the theory of maximum likelihood, a Robust Eigenvalue Maximum Likelihood (REML) method is proposed. By analyzing the statistical distribution characteristics of differences between the eigenvalues of the signal covariance matrix and those of signal correlation matrix, a maximum likelihood function can be established to estimate the number of endmembers contained in mixed spectra. This method consists of two steps: first, the original spectral data is pre-processed using multiple regression and a modified minimum noise fraction method to complete the noise estimation and whitening process, thereby effectively suppressing the interference of spectrally correlated noise. Then, the number of endmembers is estimated by solving a discrete logarithmic maximum likelihood function. This method does not require any input parameters and runs fast. Simulation experiments are based on synthetic mixed spectral data generated by the visible/near-infrared spectral signatures of five varieties of artificial space target materials measured in the laboratory and U. S. Geological Survey spectral dataset. And, experimental results demonstrate that this method can effectively suppress the interference of spectrally correlated noise and Gaussian white noise, and the estimation results have good accuracy and stability.

Keywords Artificial space target; Spectral analysis; Number of endmembers; Intrinsic dimension

(Received May 19, 2020; accepted Sep. 16, 2020)