

高光谱成像的垃圾分类识别研究

赵冬娥^{1,2}, 吴 瑞², 赵宝国², 陈媛媛²

1. 中北大学电子测试技术国家重点实验室, 山西 太原 030051

2. 中北大学信息与通信工程学院, 山西 太原 030051

摘 要 高光谱成像因光谱分辨率高、图谱合一、可实现快速无损检测等特点现已广泛应用于农业、医学、遥感等领域。现有的对可回收生活垃圾检测与分类的方法, 都存在检测时间长, 分类效率低, 而大量多种垃圾无法同时快速分拣等问题。考虑到不同类别的生活垃圾由于其主要组成分子结构的差异, 对不同波长的光有不同的吸收特性。高光谱图像在记录待分类垃圾的空间信息的同时, 可以获得垃圾对不同波长的光的反射率光谱信息, 通过建立识别分类模型对反射率光谱信息进行分析可以实现对高光谱图像中待分类垃圾的识别与分类。收集常见纸质、塑料、木质三种材料的可回收的垃圾样本, 包括塑料瓶、食品包装袋、塑料玩具(饰品)碎片、一次性筷子、雪糕棒、木制家具碎片、木制包装盒、废旧课本、广告纸、办公用纸等多种物品共 30 个样本, 进行清洗和裁剪处理, 避免样本表面污渍对样本反射率产生影响。利用高光谱成像系统采集样本在近红外(780~1 000 nm)范围内的高光谱图像, 其中 18 个样本做训练样本集, 12 个样本做测试样本集。对采集的样本图像数据做预处理, 包括去噪声以及黑白校正反演反射率信息等处理; 通过主成分分析(PCA)方法对训练样本集感兴趣区域(ROI)进行分析, 提取到的特征波段为 795.815, 836.869, 885.619, 916.409, 929.239, 934.37, 957.463, 972.858 和 988.253 nm; 在特征波段下分别提取这三种类别垃圾的参考光谱, 通过光谱角度填图法(SAM)对测试样本 ROI 区域内提取的测试样本点集在特征波段下与参考光谱进行匹配, 由匹配程度进行样本点归类, 分析结果表明, 测试样本集中纸制样本(A 类别)、塑料样本(B 类别)、木制样本(C 类别)的分类准确度分别为 100%, 98% 和 100%, 测试样本点集整体的分类准确度为 99.33%; 通过 Fisher 判别方法分析训练样本集得出判别函数式和判别准则, 对测试样本点集分类, 评价结果为 A, B 和 C 类样本分类准确度分别为 100%, 100% 和 97%, 测试样本点集整体分类准确度为 99%。通过 SAM 和 Fisher 两种判别方法对测试样本集的光谱图像进行目标物的检测与分类, 结果表明, 利用 SAM 判别方法在可回收垃圾的高光谱图像中实现检测与分类有更高的分类准确度, 可达到 99.33%。同时, 也验证了使用高光谱成像进行可回收垃圾快速分类的科学性以及可行性, 对未来系统化、机械化、智能化地解决生活中可回收垃圾的分类具有一定的实用意义。

关键词 高光谱成像; 垃圾分类; PCA; SAM; Fisher

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)03-0917-06

引 言

依靠人工辨识对垃圾进行分拣归类^[1], 是当下大多数地区处理垃圾的手段, 不仅耗时耗力, 而且受限于从事相关工作的工人的专业判断, 容易造成垃圾的错误分类。许多发达国家在 20 世纪 60 年代就致力于采用科技手段代替人工法对垃圾进行分类处理^[2-3]。现在, 国内外已经有多种较为完整的垃圾分类回收体系, 大多体系都是基于红外扫描^[4]、浮

选^[5]、静电分离^[6]等方式利用不同材料的垃圾的能量吸收、磁性、重量、密度等特性对垃圾进行分类, 提高了部分材料的垃圾的分类效率。然而仍然无法满足当下大量不同材料垃圾的快速分类处理的需求^[7]。

高光谱成像系统采用高分辨率光谱相机, 在探测物体空间特征的同时能获得上百个波段带宽小于 10nm 的光谱信息^[8]。考虑到不同材料的垃圾对不同波长的光的吸收特性与其本身的分子结构的差异^[9], 垃圾的颜色会影响可见光波长范围内的光的反射, 采集生活垃圾在近红外光谱(780~

收稿日期: 2018-08-14, 修订日期: 2018-12-28

基金项目: 国家自然科学基金青年基金项目(61605176)资助

作者简介: 赵冬娥, 女, 1970 年生, 中北大学信息与通信工程学院教授 e-mail: zhaodong@nuc.edu.cn

1 000 nm)的高光谱图像,通过主成分分析法(principal components analysis, PCA)^[10]、光谱角填图法(spectral angle mapping, SAM)^[11]以及 Fisher 判别分析^[12]建立垃圾的识别分类模型,验证高光谱成像在垃圾分类领域的实用性。

1 实验部分

1.1 材料

透明及半透明的物体对光的反射较少,通过分析反射率图谱完成分类识别比较困难,实验样本选择常见的不透明的可回收生活垃圾:塑料瓶、塑料包装袋、塑料玩具(饰品)碎片、一次性筷子、雪糕棒、木制家具碎片、木制包装盒、废旧课本、广告纸、包装纸、办公用纸等三种材料各 10 个样本。对收集的待测垃圾样本进行清洗和裁剪等处理,使样本表面干净且形状规则,避免表面污渍对系统光源的吸收和反射产生的影响。每类样本取六个组成训练样本集,其余样本组成测试样本集。

1.2 高光谱图像采集

实验选用美国 Headwall 公司产的 Hyperspec VNIR-A 型号高光谱成像光谱仪。主要由 1 600×2 500 高分辨率 CCD 相机、全反射式光栅、狭缝及镜头组成。为避免不同种类的垃圾样本的颜色对光谱信息的影响,采用 2.5 nm(平均间隔)波段间隔对实验样本在近红外 777.91~1 003.59 nm 共 89 个波段进行光谱成像。

表 1 高光谱成像仪参数

Table 1 Hyperspectral imaging spectrometer

| Hyperspec VNIR-A 高光谱成像光谱仪 | |
|---------------------------------|-------------|
| 光谱范围/nm | 380~1 000 |
| 光谱分辨率/nm | 2~3 |
| 光谱通道数 | 837 |
| 空间通道数 | 1 004 |
| 狭缝宽度/ μm | 25 |
| 数值孔径 | F/2.0 |
| 光谱采样值/(nm·pixel ⁻¹) | 0.74 |
| Smile & Keystone | 2×2.3×3 484 |
| 探测器类型 | CCD |
| 像元间距/ μm | 7.4 |
| 最大帧频/fps | 90 |
| 相机接口 | Cameralink |
| 重量/kg | ~2.8 |

实验在暗室中进行,通过调节卤素灯光源亮度,避免图像过饱和;设置相机积分时间为 40 ms,凝视时间为 39 ms;设置待测样本与相机狭缝的距离为 30 cm,调节镜头焦距并计算设置移动载物平台的推扫速度为 2.1 cm·s⁻¹,使采集到的样本图像清晰且不失帧,保证样本光谱信息和空间信息的完整性。

1.3 识别分类模型的建立

分类识别模型的流程,见图 1。

采集样本高光谱图像数据,通过黑白校正、ROI 区域的

选择、去掉首尾波段等方法去除环境噪声和系统噪声;对训练样本集感兴趣区域进行主成分分析提取特征波段,实现数据的降维处理;通过 SAM、Fisher 判别分类方法对测试样本集进行归类,并通过分类准确度评价归类结果,从而验证高光谱成像在垃圾分类领域的实用性。

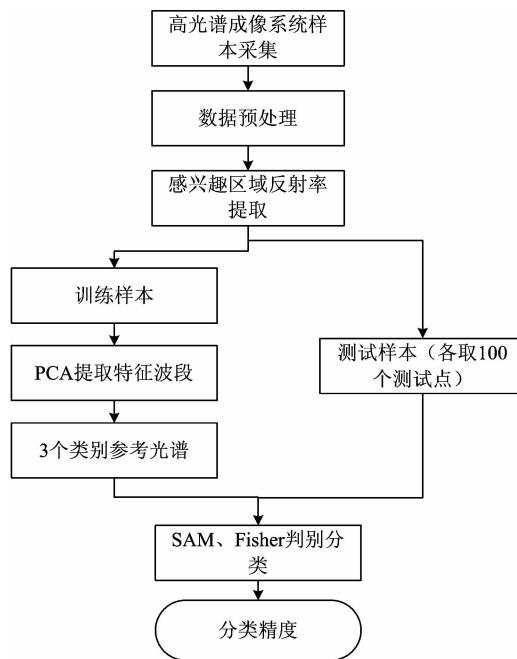


图 1 识别分类模型流程图

Fig. 1 Classification recognition model

2 数据分析

2.1 图像数据预处理

为消除采集过程中,光谱相机在首尾波段产生的噪声,选择 780.41~1 001.09 nm 共 87 个波段的数据信息进行处理。为消除相机自身暗电流带来的图像噪声,通过参考白板比较法进行黑白校正得到样本的反射率,公式为

$$R_{\text{norm}} = \frac{R_{\text{raw}} - R_{\text{dark}}}{R_{\text{white}} - R_{\text{dark}}} \quad (1)$$

式(1)中, R_{norm} 为校正后的光谱图像, R_{raw} 为原始图像光谱, R_{dark} 为拧上镜头盖后采集的全暗参考图像, R_{white} 为反射率为 99% 的标准参考白板得到的参考图像。

由于垃圾样本形状大小各不相同,且部分表面发生强反射,故选择均匀、没有异常点、形状规则的区域作为样本图像分析的 ROI,减少实验样本数据的噪声。

2.2 PCA 特征提取

采集到的垃圾样本的高光谱图像以近乎连续的波段记录了样本的空间信息和 780.41~1 001.09 nm 部分近红外光谱信息,数据量大,波段信息冗余,计算耗时。主成分分析方法目的是提取对原始信息贡献率大的主成分,以较少的维度信息代替原始多维信息,去除冗余波段,实现数据的降维。

$$Z_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{im}x_m \quad (2)$$

其中, Z_i 为第 i 个主成分, a_m 为对应训练样本集标准化矩阵

特征值 λ_i 的特征向量。

$$\alpha_i = \frac{\lambda_i}{\sum \lambda_i} \quad (3)$$

通过计算各主成分贡献率 α_i ，前 3 个主成分累加贡献率至 96.59%，第 4 个及之后的主成分贡献率均小于 1%。表明前 3 个主成分解释了原始数据全部信息的 96.59%。

$$L(Z_i, X_j) = \sqrt{\lambda_j} a_{ij} \quad (4)$$

式(4)中, $L(Z_i, X_j)$ 为表示主成分 Z_i 与原始变量 X_j 相关程度的载荷系数, 不受原始光谱冗余信息影响, 有助于识别分类的有效波长^[13]。对 Z_1, Z_2 和 Z_3 主成分对应载荷系数分析可得: 795.815, 836.869, 885.619, 916.409, 929.239, 934.37, 957.463, 972.858 和 988.253 nm 这九个波长下三组载荷系数均存在波峰(波谷), 且绝对值相对较大, 表明 PCA 提取特征空间 $H = [795.815, 836.869, 885.619, 916.409, 929.239, 934.37, 957.463, 972.858, 988.253 \text{ nm}]$ 可代表原始图谱信息用于垃圾样本分类。

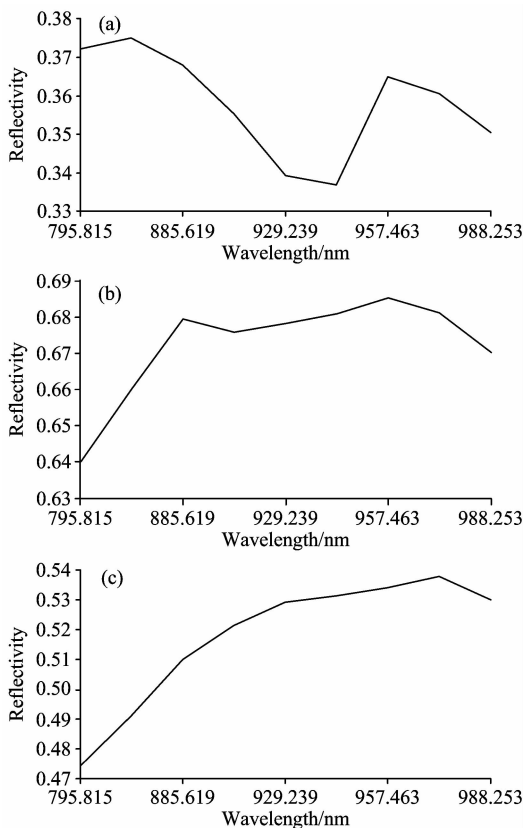


图 2 (a), (b) 和 (c) 分别为塑料、木制、纸垃圾样本的参考光谱

Fig. 2 (a), (b), and (c) are reference spectra of plastic, wood, and paper

2.3 光谱匹配判别分析

光谱角匹配法是将图像中每一个像元的 n 个光谱维响应当作 n 维空间矢量, 通过计算待测矢量与参考光谱转化的参考矢量之间的广义夹角 θ 来表征匹配程度, 夹角越小, 表明待测像元光谱与参考光谱匹配程度越高, 如式(5)

$$\theta = \cos^{-1} \frac{\vec{n}_c \cdot \vec{n}_d}{|\vec{n}_c| |\vec{n}_d|} \quad (5)$$

其中, n_c 为匹配类别参考光谱向量, n_d 为待测像元光谱向量。

$$\min(\theta_1, \theta_2, \dots, \theta_i) = \theta_k \rightarrow k \text{ 类}, (k \leq i) \quad (6)$$

其中, θ_i 为待测像元光谱与 i 类别参考光谱, θ_k 为其中最小广义夹角, 表明待测像元与 k 类别最匹配。

分别选择三类训练样本的特征空间内 ROI 平均反射率光谱曲线反射率, 构成三种类别垃圾的参考光谱, 如图 2 所示。

在 12 个测试样本图像中选择感兴趣区域, 并分别随机提取 25 个样本测试点, 构成容量为 300 个测试点的样本测试点集。对三种类别各自提取的 100 个样本点分别与三条参考光谱进行 SAM 判别分析, 如图 3。

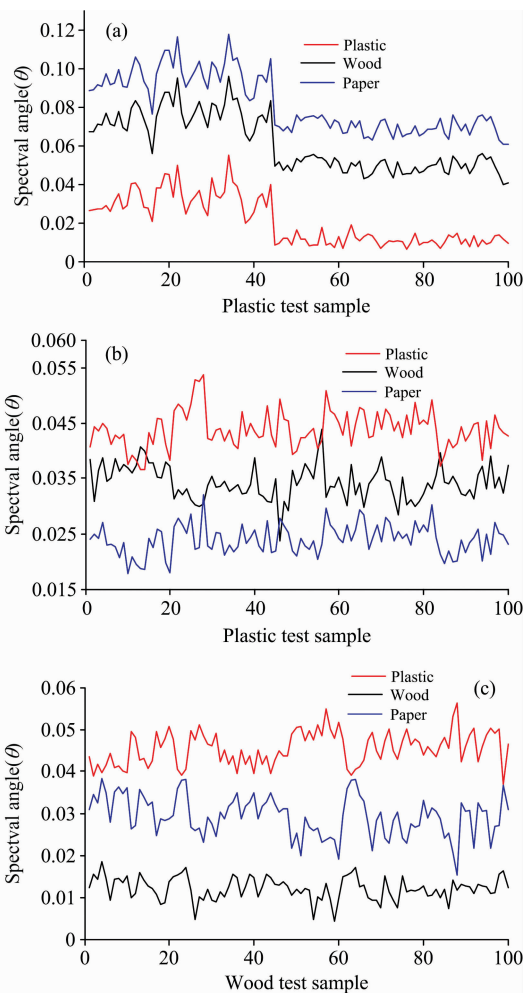


图 3 (a), (b), (c) 分别为塑料、纸、木类别测试样本点 SAM 判别分析结果

Fig. 3 (a), (b), and (c) show the results of SAM discriminant analysis of plastic, paper, and wood test samples

由图 3 可知, 塑料、木制品、纸制品 3 个类别的 100 个测试样本点分别与三组参考光谱匹配程度分类时, 编号 17 和 33 两个纸类别样本点被错误分为木制类别, 其余测试点均被分至正确类别。

2.4 Fisher 判别分析

Fisher 判别分析法是根据已知类别的样本原始信息或能代表已知类别样本信息的特征空间信息，建立线性判别函数 Y [式(7)]，将多维空间样本投影至一维空间，并通过判别准则对未知类别样本进行分类判别，进而验证样本分类信息的分类准确度。

$$Y(x) = \sum_{i=1}^p C_i x_i, \quad i = 1, 2, \dots, p \quad (7)$$

利用训练样本集得出判别塑料与其他类别的函数 Y_1 (阈值为 -0.4212) 以及判别木类别与纸类别的函数 Y_2 (阈值为 65.2191)：

$$Y_1 = -3.8439x_1 + 17.4284x_2 + 14.9068x_3 + 5.8967x_4 - 19.0925x_5 - 11.5119x_6 - 4.7543x_7 + 2.0040x_8 - 3.9972x_9 \quad (8)$$

$$Y_2 = 289.2366x_1 - 79.7434x_2 + 916.0392x_3 + 314.0980x_4 + 583.2292x_5 - 280.5590x_6 - 472.1794x_7 - 821.0909x_8 - 263.0350x_9 \quad (9)$$

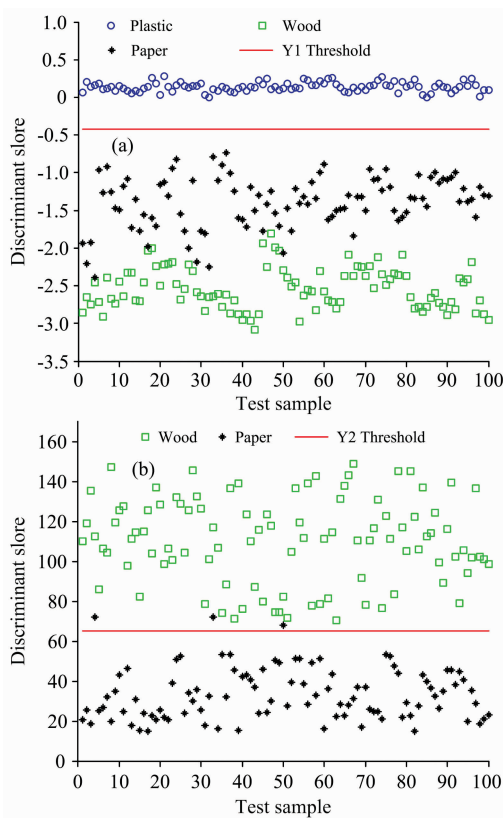


图 4 (a) 判别塑料与其他类别；(b) 判别除塑料类别的木类与纸类

Fig. 4 (a) identifies plastics and other categories; (b) identifies wood and paper in the plastic category

通过 Y_1 和 Y_2 以及对应阈值对测试样本集进行 Fisher 判别分类，结果如图 4。

由图 4 可知，三个纸类别测试样本点被误分至木类别，其余测试点均被分至正确类别。

2.5 结果评价

由图 3 和图 4 可知，SAM 光谱匹配法、Fisher 判别法对塑料(A)、纸制品(B)、木制品(C)样本测试点集分类准确度如表 3 和表 4 所示。

表 3 SAM 分类结果评价

Table 3 Classification result evaluation

| 样本 | 测试点数 | 样本点归类 | | | 准确度 /% |
|----|------|-------|----|-----|--------|
| | | A | B | C | |
| A | 100 | 100 | 0 | 0 | 100 |
| B | 100 | 0 | 98 | 2 | 98 |
| C | 100 | 0 | 0 | 100 | 100 |

表 4 Fisher 分类结果评价

Table 4 Classification result evaluation

| 样本 | 测试点数 | 样本点归类 | | | 准确度 /% |
|----|------|-------|-----|---|--------|
| | | A | B | C | |
| A | 100 | 100 | 0 | 0 | 100 |
| B | 100 | 0 | 100 | 0 | 100 |
| C | 100 | 0 | 97 | 3 | 97 |

由统计结果可知，通过 SAM 光谱匹配算法对测试样本集与训练样本集参考光谱进行匹配分类，分类准确度分别为 100%，98% 和 100%，测试样本点集整体分类准确度达到 99.33%；通过 Fisher 判别分析方法对训练样本集分析得出判别函数 Y_1 和 Y_2 以及判别准则(阈值 y_1 和 y_2)，并对测试样本集分类，准确度分别为 100%，100% 和 97%，测试样本点集整体分类准确度达到 99%。

3 结论

采集了常见的可回收的生活垃圾，如废旧课本、办公室用纸、塑料瓶、塑料包装袋、废旧木制家具、一次性筷子等多个纸质、塑料、木质的垃圾样本，通过建立识别分类模型，在近红外(780.41~1001.09 nm)波段提取样本的特征波段，并利用 SAM 光谱匹配算法和 Fisher 判别分析算法对测试样本 ROI 区域内测试点分类，分类准确度均达到 99% 以上。

结果表明，利用高光谱成像在实验室环境下，可以将部分常见的可回收垃圾进行识别分类，对未来改进和创新垃圾分类技术提供了依据。

References

- [1] Hollstein F, Wohllebe M, Arnaiz S. Identification and Sorting of Plastic Film Waster by NIR-Hyperspectral-Imaging. International Conference on Near Infrared Spectroscopy, 2015.
- [2] WANG Dong, LIANG Yong, MA Gang-ping, et al(王 栋, 梁 勇, 马刚平, 等). Environmental Engineering(环境工程), 2017, 35(2): 138.
- [3] Lu X N, Liu F, He Z Z, et al. Science China Technological Sciences, 2018, 61(9): 1.
- [4] Yang J, Xu L, Gao M G, et al. Acta Physica Sinica, 2013, 62(14): 140705.
- [5] WEI Guo-xia, LIU Han-qiao, XU Xian, et al.(魏国侠, 刘汉桥, 徐 仙, 等)Modern Chemical Industry(现代化工), 2014, 34(11): 46.
- [6] Silveira A V M, Santana M P, Tanabe E H, et al. International Journal of Mineral Processing, 2017, 169(11): 91.
- [7] YANG Yu-qiang, WANG Kun, HUANG Huan-lin, et al(杨煜强, 王 坤, 黄焕林, 等). Acta Scientiae Circumstantiae(环境科学学报), 2015, 35(2): 570.
- [8] LIU Li-xin, LI Meng-zhu, ZHAO Zhi-gang, et al(刘立新, 李梦珠, 赵志刚, 等). Chinese Journal of Lasers(中国激光), 2018, 45(2): 1.
- [9] GUO Xing-wang, GUAN He-qing, LIU Ying-tao, et al(郭兴旺, 管和清, 刘颖韬, 等). Infrared and Laser Engineering(红外与激光工程), 2017, 46(1): 51.
- [10] TAN Nian, SUN Yi-dan, WANG Xue-shun, et al(谭 念, 孙一丹, 王学顺, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(11): 3370.
- [11] LIU Chang, LI Jun-wei(刘 畅, 李军伟). Infrared and Laser Engineering(红外与激光工程), 2015, 44(10): 3141.
- [12] WU Tong, TAN Chao(吴 同, 谭 超). Advances in Analytical Chemistry(分析化学发展), 2016, 6(1): 1.
- [13] LIU Song-tao, CHANG Chun, MA Xin-xing, et al(刘松涛, 常 春, 马新星, 等). Laser & Infrared(激光与红外), 2013, 43(11): 1316.

Research on Garbage Classification and Recognition Based on Hyperspectral Imaging Technology

ZHAO Dong-e^{1,2}, WU Rui², ZHAO Bao-guo², CHEN Yuan-yuan²

1. National Key Laboratory for Electronic Measurement Technology, North University of China, Taiyuan 030051, China

2. School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

Abstract Hyperspectral imaging technology is profoundly applied into the fields of agriculture, medicine and remote sensing due to its high spectral resolution, merged image-spectrum, and fast non-destructive testing. While the method used now has the defects of long-term testing period, poor efficiency and sorting asynchrony. Spectral image can identify and classify the target garbage by establishing a recognition and classification model and analyzing reflectance spectrum information based on the facts that different materials of domestic garbage, due to their different molecular structures, will absorb different wavelengths of light and the hyperspectral image can obtain the spatial information and the reflectance spectral information from different-wavelength illumination of the target garbage. Collected recyclable garbage samples of common paper, plastic and wood materials, including plastic bottles, food packaging bags, plastic toys (jewelry) pieces, disposable chopsticks, ice cream bars, wooden furniture pieces, wooden boxes, waste textbooks, advertising paper, office paper and other items, 30 in total. And cleaned and cut them to avoid the influence of sample surface stains on the sample reflectivity. Hyperspectral imaging systems were used to acquire hyperspectral images of the sample in the near-infrared (780 ~1 000 nm) formed 18 training samples and 12 test samples. Pre-processed the collected sample image by de-noising and black-and-white correction inversion of reflectivity information. Then analyzed the region of interest of training samples by principal components analysis. The characteristic band extracted were 795.815, 836.869, 885.619, 916.409, 929.239, 934.37, 957.463, 972.858, 988.253 nm; Next, matched and categorized the characteristic band of the ROI with reference spectra of the three types of garbage from the characteristic band by spectral angle mapping. The result illustrated that the classification precision of paper (A class), plastic (B class) and wood (C class) were 100%, 98% and 100% respectively and the average was 99.33%; at last, sorted the test samples by Fisher linear discrimination. The classification precision of class A, B, C were 100%, 100% and 97% respectively and the average was 99%. After a series of testing and classification by SAM and Fisher as the narrated above, the results showed that aforesaid manipulation of

hyperspectral image for recyclable garbage by SAM can get more accurate results which is 99.33%. meanwhile, the research can testify that it's feasible to apply the scheme of hyperspectral imaging to assort garbage, which is significant to methodically and automatically recycle garbage in the future.

Keywords Hyperspectral imaging; Garbage classification; PCA; SAM; Fisher

(Received Aug. 14, 2018; accepted Dec. 28, 2018)