文章编号:1004-4213(2011)08-1132-5

基于高光谱散射图像技术的 UVE-LLE 苹果粉质化分类

汪泊锦,黄敏,朱启兵,王爽

(江南大学 轻工过程先进控制教育部重点实验室,江苏 无锡 214122)

摘　要:利用高光谱散射图像技术研究了苹果的粉质化无损检测.提出了一种无信息变量消除法和局部线性嵌入相结合的苹果粉质化分类的新方法.经无信息变量消除法筛选后的波段降为全谱的23.5%.将波段选择后的原始图像数据用局部线性嵌入降维作为偏最小二乘判别分析的输入变量并建模.无信息变量消除法与局部线性嵌入相结合算法和局部线性嵌入降维方法得到的粉质化分类测试准确度分别是79.0%和79.0%;无信息变量消除法与平均反射法相结合和平均反射法特征提取得到的是77.4%和75.8%.结果表明,无信息变量消除法与局部线性嵌入想结合的方法可以大大地降低高光谱散射图像的数据量,同时保证了分类准确度,为在线检测、分类和高光谱数据的存储提供了一种实时、有效的方法.

关键词:粉质化;高光谱散射图像;无信息变量消除法;局部线性嵌入法;偏最小二乘判别分析

中图分类号:O657.3　　**文献标识码**:A　　**doi**:10.3788/gzxb20114008.1132

0　引言

苹果的粉质化是指苹果非正常软化、汁液减少和果肉质地发棉等一系列生理失调现象,它是苹果内部品质如硬度、糖度和汁液含量等参量的综合指标.传统的基于人工品尝的苹果粉质化测量方法具有对人员依赖性强、对苹果的破坏性等特点而难以在线运用.研究苹果粉质化的无损测量方法已成为国内外的一个重要课题,出现了多种苹果粉质化的无损测量方法[1-3].高光谱散射图像技术是美国科学家 Lu 提出的一种无损测量农产品内部品质的新方法,已被成功地用于苹果内部的糖度、硬度等成分预测[4-6].本课题组利用高光谱散射图像技术,并结合平均反射特征提取方法(mean reflectance, MEAN)对苹果的粉质化分类进行了初步的研究[7].但是这种方法存在两个问题:1)高光谱散射图像波段多,存在大量冗余信息波段,给在线的实时运用带来困难;2)平均反射特征提取方法是通过对每个波段的数据求平均值来获得苹果粉质化的分类特征,这一方法虽然简单,但是容易导致图像特征的丧失.针对这两个问题,本文首先采用无信息变量消除法(Uninformative Variables Elimination, UVE)[8]对特征波段进行选择,在此基础上,利用选择的特征波段构建数据矩阵,并采用局部线性嵌入方法(Locally Linear Embedding, LLE)[9],对数据矩阵进行特征提取.实验结果表明:采用无信息变量消除法与局部线性嵌入相结合的方法(UVE-LLE)可显著降低高光谱漫反射图像的特征波段,其分类结果要优于全波段条件下的平均反射特征提取方法.

1　实验材料与数据采集

1.1　苹果样本和高光谱系统

实验使用的 580 个"Red Delicious"苹果样本中 180 个样本采摘于美国密歇根州立大学(Michigan State University, MSU)的教学研究中心,并存储在一定的条件下(2% O_2, 3% CO_2 和 0℃);剩下的 400 个苹果样本买自当地的商店(Commercial Packinghouse, CP).总的样本被分为两个部分:240 个样本(180 来自 CP,60 来自 MSU)被储藏在 4℃的温度下;340 个样本(220 来自 CP,120 来自 MSU)被储藏在 20℃温度和 95%的湿度的条件下.高光谱系统由图像光谱仪、高光谱摄像头、250 W 的光纤卤素灯和一套输送装置等部件组成.样本和高光谱系统的详细介绍见参考文献[10].

基金项目:国家自然科学基金(No. 60805014)和中央高校基本科研业务费专项资金(No. JUSRP20913,No. JUSRP21132)资助

第一作者:汪伯锦(1985－),男,硕士研究生,主要研究方向为高光谱图像技术. Email:wp2011@yahoo.cn

导师(通讯作者):黄敏(1974－),女,副教授,主要研究方向为农产品无损检测. Email:huangmzqb@163.com

收稿日期:2011-02-25;**修回日期**:2011-04-19

1.2　数据采集和高光谱散射图像的特点

对 580 个“Red Delicious”样本，每个样本采集 10 幅图像，然后将这些图像取平均，平均后的图像保存起来留作以后处理. 由于采集时的操作和样本问题，6 个样本的图像被剔除. 利用 Magness Taylor (MT) 质地分析仪(model TA. XT2i, Stable Micro Systems, Inc., Surrey, U. K)来测量样本的硬度和汁液含量等数据. 根据硬度和汁液含量来对苹果的粉质化程度进行分类，具体的分类标准如表 1.

表 1　用破坏性仪器法对苹果种类的分类

Table 1　Apple classes according to the destructive instrumental parameters

	Two-class	
	Juiciness<5 cm²	Juiciness≥5 cm²
Hardness<40 KN/m	Mealy	Nonmealy
Hardness≥40 KN/m	Nonmealy	Nonmealy

利用高光谱系统采集到的高光谱散射图像及其特点见图 1，其中高光谱散射图像是对称的. 图 1(a) 在 675 nm 波段处的亮度有明显的降低是叶绿素对光谱的吸收导致的，对应图 1(b)中最低的曲线. 因为是“Red Delicious”苹果，其在 600～1 000 nm 波段范围内特征比较明显. 由于光谱分辨率为 5 nm，共得到 81 个波段. 考虑到噪音的影响，选取空间距离为从空间零点左右各 10 mm，共 20 mm 的空间范围，空间分辨率是 0.20 mm/pixel，得到每个波段包含 101 个像素点的数据量，即原始图像的维数为 81×101.

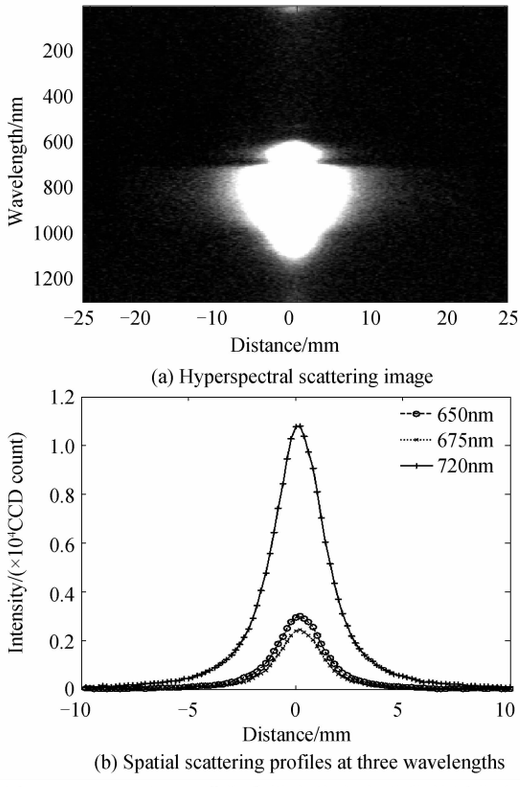

图 1　“Red Delicious”苹果样本的高光谱散射图像和三个典型波段的散射图像

Fig. 1　Hyperspectral scattering images of an “Red Delicious” apple and spatial scattering profiles at three wavelengths

MEAN 算法[11]是对每个波段的数据做平均处理，这样每个样本被处理后得到的是含有 81 个数据的列向量.

2　UVE 算法和 LLE 算法的基本原理

2.1　UVE 算法原理

UVE[12-13]是一种通过分析偏最小二乘(Partial Least Squares, PLS)回归系数，用来消除无信息的变量. 在 PLS 模型中，光谱矩阵 $\boldsymbol{X}$ 和浓度矩阵 $\boldsymbol{Y}$ 之间的数学关系式

$$\boldsymbol{Y}=\boldsymbol{X}\boldsymbol{b}+\boldsymbol{e} \tag{1}$$

式中 $\boldsymbol{b}$ 是系数向量，$\boldsymbol{e}$ 是误差向量. 该算法就是将随机产生的噪音矩阵和光谱矩阵合并，每次去除一个样本做交叉验证，得到的所有系数向量组成回归系数矩阵 $\boldsymbol{B}$，然后计算每次产生的系数向量 $\boldsymbol{b}$ 的均值 mean($\boldsymbol{b}$)和标准偏差 $S(\boldsymbol{b})$的比值 C_i，既系数向量的稳定性.

$$C_i=\frac{\mathrm{mean}(\boldsymbol{b}_i)}{S(\boldsymbol{b}_i)} \tag{2}$$

2.2　LLE 算法原理

LLE 是一种利用局部的线性表示来代替实际数据的非线性部分，就本质而言，它是将流形上的近邻点映射到低维空间的近邻点. LLE 的优点是在降维的同时能够很好地保持原始数据的数据流形，因而，它特别适合图像等非线性数据的处理.

LLE 算法主要分为三个步骤：1)计算每个样本点 $\boldsymbol{X}_i$ 的 K 个近邻点 $\boldsymbol{X}_{ij}$ ($j=1,2,\cdots,k$)(用这 K 个近邻点线性表示该样本点)；2)计算重构权值矩阵 $\boldsymbol{W}_{ij}$ ($j=1,2,\cdots,k$)，使得 K 个近邻点能够最好地线性表示样本点，条件是使重构误差函数 $\varepsilon(\boldsymbol{W})$最小，且满足 $\sum_{j=1\to k}\boldsymbol{W}_{ij}=1$；3)由得到的重构权值矩阵 $\boldsymbol{W}_{ij}$ 来计算 $\boldsymbol{X}_i$ 的低维空间向量 $\boldsymbol{Y}_i$，此时要使嵌入损失函数 $\Phi(\boldsymbol{Y})$最小，其中式(4)中的 $\boldsymbol{Y}_{ij}$ ($j=1,2,\cdots,k$)为 $\boldsymbol{Y}_i$ 的 K 个近邻点.

$$\varepsilon(\boldsymbol{W})=\sum_i\left|\boldsymbol{X}_i-\sum\nolimits_j\boldsymbol{W}_{ij}\boldsymbol{X}_{ij}\right|^2 \tag{3}$$

$$\Phi(\boldsymbol{Y})=\sum_i\left|\boldsymbol{Y}_i-\sum\nolimits_j\boldsymbol{W}_{ij}\boldsymbol{Y}_{ij}\right|^2 \tag{4}$$

2.3　LLE 泛化算法

LLE 算法是利用样本点的 K 个近邻点来表示该点. 在建模之前，训练样本已经确定，因此可以对其直接利用 LLE 进行特征提取. 但是，在实际应用中测试样本是未知的，因此在对测试样本进行 LLE

降维时，必须用训练样本来实现，即测试样本的 K 个近邻点要在训练样本中选择. 具体的 LLE 泛化算法的实现步骤是：1)计算训练样本先后由 UVE、LLE 特征提取后的矩阵 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$；(2)对于一个测试样本，先求其在 UVE 特征提取后得到的向量 $\boldsymbol{z}_o$，再求其在 $\boldsymbol{X}$ 中的 K 个最近邻点(即 K 个向量，K 不大于 $\boldsymbol{X}$ 中包含的样本个数)，并用这 K 个近邻点线性表示，求得系数矩阵 $\boldsymbol{\omega}$，即：$\boldsymbol{z}_0=K_0\times\omega$，其中 K_0 是 K 个向量组成的矩阵；3)这 K 个向量对应在 $\boldsymbol{Y}$ 中的 K 个向量组成的矩阵是 $\boldsymbol{K}_0'$，由 $\boldsymbol{\omega}$ 和 $\boldsymbol{K}_0'$ 计算泛化的测试样本 $\boldsymbol{z}_0'$，即：$\boldsymbol{z}_0'=\boldsymbol{K}_0'\times\boldsymbol{\omega}$.

2.4 基于 UVE-LLE 的苹果粉质化分类算法流程

Step1：利用平均反射特征提取方法对去除奇异样本后的 574 个样本进行原始特征提取；每一样本图像共得到 81 维原始特征；

Step2：利用 Kennard-Stone 算法[14]，对所有样本进行划分，共分为 450 个训练样本和 124 个测试样本；

Step3：将训练样本作为 UVE 的输入，筛选出的最优波段在全波段中的位置用 $S_m(m=1,2,\cdots,N)$表示，其中 N 是最优波段的个数；

Step4：将训练样本 i 的原始图像中对应最优波段 $S_m(m=1,2,\cdots,N)$的数据首尾相接，形成 $101\times N$ 维向量(每个波段包含 101 个像素点的数据量) $X_i(i=1,2,\cdots,h)$，$h=450$ 为总的训练样本数；

Step5：利用 LLE 算法对训练样本 i 的特征 $X_i(i=1,2,\cdots,h)$进行特征提取，获得提取后的特征 $Y_i(i=1,2,\cdots,h)$；

Step6：将特征 $Y_i(i=1,2,\cdots,h)$和破坏性仪器获得的粉质化结果输入偏最小二乘判别分析(Partial Least Squares Discriminant Analysis, PLSDA)建立分类模型；

Step7：对测试样本 $k(k=1,2,\cdots,t,t=124$，为测试样本总数)的原始图像，将选择的最优波段 $S_m(m=1,2,\cdots,N)$首尾相接，形成测试样本向量 Z_k，利用 LLE 泛化算法进行特征提取，获得测试样本的特征为 Z_k'，将特征 Z_k' 输入已建立好的 PLSDA 分类模型，获得其分类预测结果.

3 实验结果与讨论

3.1 UVE 选取有效波长

用 UVE 算法对所有的 81 个波段进行筛选. 选择的结果见图 2.

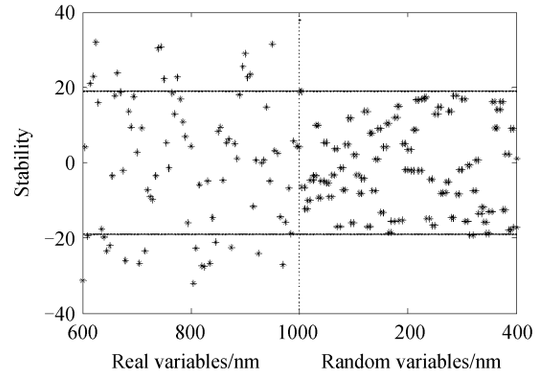

图 2 UVE 算法波段选择

Fig. 2 Wavelengths selected by UVE

图中包含 162 个点，中间竖线左边的 81 个点对应原始图像的 81 个波段，右边的是添加的 81 个随机变量点. 两条水平虚线是阈值，两条虚线之外的点对应的是筛选得到的最优波段，内部的是无信息波段. 本文阈值选取的是最大随机变量点的 0.99. 经 UVE 筛选后得到 19 个波段，分别为：610、615、625、630、635、640、665、670、675、680、685、805、810、835、840、935、940、950、995 nm. 可以看出，通过 UVE 波段选择后，大幅压缩了波段数量，减小了高光谱数据的存储量.

3.2 分类结果

表 2 给出了利用不同特征提取方法进行苹果粉质化分类的结果. 其中 LLE 是对原始图像直接用 LLE 进行特征提取后，利用 PLSDA 分类得到的结果；UVE-MEAN 是利用 UVE 特征提取后，再用 MEAN 进行特征提取，结果输入 PLSDA 分类的结果；而 MEAN 则是对原始数据直接用 MEAN 进行特征提取后，输入 PLSDA 得到的分类结果.

表 2 四种算法得到的分类准确度比较

Table 2 Classification accuracy by four algorithms

Algorithm	Model Accuracy	Test Accuracy
UVE-LLE($k=18,d=46$)	79.8	79.0
LLE($k=20,d=46$)	79.6	79.0
UVE-MEAN	75.3	77.4
MEAN	74.9	75.8

LLE 算法中有 K 和 d 两个参量：其中 K 表示近邻点的个数，它是唯一的自由参量；d 表示要降到的维数. 这两个参量大多是根据经验在一定的范围中选择. 本文 K 选取的范围是 10～20，d 的范围是 40～50，并且根据留一交叉验证产生的交叉验证错误分类率(Classification Error Rate for Cross-validation, CLASSERRCV)的最小值确定 K 和 d 的值.

从表 2 可以看到，用 LLE 特征提取方法得到的分类准确度要优于传统的平均反射特征提取方法. 这主要是因为相比于简单的平均反射特征提取方法，LLE 能够较大程度保留原始数据的非线性特征，从而提高了分类准确度. 利用 UVE 和采用全波段所得到的准确度没有大的波动. 因此，使用 UVE 可以保证苹果的粉质化分类准确度. 而该算法可以使原来的 81 个波段减少到 19 个波段，这使数据量有了较大的降低.

3.3　UVE 中随机噪音矩阵和 LLE 中 K、d 对准确度的影响

UVE 算法要在光谱矩阵中加入噪音矩阵. 噪音矩阵是随机产生的，因此每次得到的最优波段个数不固定，绝大部分在 14～19 之间，其中 19 个波段出现的概率最大，并且主要的波段都包含在各次的结果中. 波段的变化对准确度的影响较小.

LLE 选择不同 K 和 d 的值，对测试准确度有影响. 当 K 较小时，将会影响数据之间的类别信息的正确性；当 K 较大时，将会破坏该算法中局部线性的前提条件. 因此 K 值的选取不宜过小或过大，同样 d 的值也是如此.

4　结论

UVE 能够对高光谱的波段进行选择，去除无用信息的波段，结合 LLE 可以实现高光谱数据的有效特征提取，最后利用 PLSDA 建模并做测试. 实验结果表明：UVE-LLE 算法在模型准确度和测试准确度方面都有较好的结果，更重要的是该算法能够为在线检测和分类并为高光谱数据的存储提供了一个可行的方法.

此外，也出现了许多新的问题. 例如本文中用 UVE 选取的 19 个波段并不是固定的，因为该算法中包含一个随机生成的矩阵，所以并不是每次被选择的波段都是完全相同的. LLE 是由 K、d 两个参量决定，不同的 K、d 对结果有影响. 因此，K、d 值的选定还要进一步研究.

致谢：本文作者对美国农业部 Postharvest Engineering Laboratory 的 Dr. Lu 在实验工作中的指导深表感谢.

参考文献

[1] BARREIRO P, MOYA A, CORREA E, *et al*. Prospects for the rapid detection of mealiness in apples by nondestructive NMR relaxometry[J]. *Applied Magnetic Resonance*, 2002, **22**(3): 387-400.

[2] MEHINAGIC E, ROYER G, BERTRAND D, *et al*. Relationship between sensory analysis, penetrometry and visible-NIR spectroscopy of apples belonging to different cultivars[J]. *Food Quality and Reference*, 2003, **14**(5-6): 473-484.

[3] BECHAR A, MIZRACH A, BARREIRO P, *et al*. Determination of mealiness in apples using ultrasonic measurements[J]. *Biosystems Engineering*, 2005, **91**(3): 329-334.

[4] LU R, PENG Y. Hyperspectral scattering for assessing peach fruit firmness[J]. *Biosystems Engineering*, 2006, **93**(2): 161-171.

[5] NOH H K, LU R. Hyperspectral laser-induced flurescence imaging for assessing apple fruit quality[J]. *Postharvest Biology and Technology*, 2007, **43**(2): 193-201.

[6] HUANG M, LU R. Apple mealiness detection using hyperspectral scattering technique[J]. *Postharvest Biology and Technology*, 2010, **58**(3): 168-175.

[7] LU R. Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images[J]. *Sensing and Instrumentation for Food Qualigy and Saftety*, 2007, **1**(1): 19-27.

[8] CENTNER V, MASSART D L, DE Noord O E, *et al*. Elimination of uninformative variables for multivariate calibration[J]. *Analytical Chemistry*, 1996, **68**(21): 3851-3858.

[9] SAM T Roweis, LAWRENCE K Saul. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, **290**(5500): 2323-2326.

[10] ZHAO Gui-lin, ZHU Qi-bing, HUANG Min. LLE-SVM classification of apple mealiness based on hyperspectral scattering image[J]. *Spectroscopy and Spectral Analysis*, 2010, **30**(10): 2739-2743.
赵桂林，朱启兵，黄敏. 基于高光谱图像技术的苹果粉质化 LLE-SVM 分类[J]. 光谱学与光谱分析，2010，**30**(10)：2739-2743.

[11] LU R, HUANG M, QIN J. Analysis of hyperspectral scattering characteristics for predicting apple fruit firmness and soluble solids content[C]. *SPIE*, 2009, **7315**: 73150I-1.

[12] CHEN Bin, CHEN Dan. The application of uninformative variables elimination in near-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2005, **4**: 26-30.
陈斌，陈蛋. 无信息变量消除法在近红外光谱测定的应用[J]. 光谱仪器与分析，2005，**4**：26-30.

[13] WANG Shuang, HUANG Min, ZHU Qi-bing. Optimal wavelength selection of hyperspectral scattering images based on UVE-PLS projection analysis[J]. *Acta Photonica Sinica*, 2011, **40**(3): 428-432.
王爽，黄敏，朱启兵. 基于无信息变量和偏最小二乘投影分析的高光谱散射图像最优波段选择[J]. 光子学报，2011，**40**(3)：428-432.

[14] MACHO S, IUSA R, CALLAO M P, *et al*. Monitoring ethylene content in heterophasic copolymers by near-infrared spectroscopy standardization of the calibration model[J]. *Anslytica Chimica Acta*, 2001, **445**(2): 213-220.

UVE-LLE Classification of Apple Mealiness Based on Hyperspectral Scattering Image

WANG Bo-jin, HUANG Min, ZHU Qi-bing, WANG Shuang
(*Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu* 214122, *China*)

Abstract: Hyperspectral scattering is a promising technique for noninvasive measurement of apple mealiness. An uninformative variable elimination (UVE) coupled with locally linear embedding (LLE) algorithm was proposed for assessing apple mealiness. After the algorithm, the number of effective wavelengths decreased to 23.5% of full wavelengths of hyperspectral scattering images. LLE was utilized to reduce the dimensionality of images composed of effective wavelengths. Partial least squares discriminant analysis was applied to develop classification model. Compared with mean reflectance (75.8%) and UVE coupled with mean reflectance algorithm (77.4%), LLE and UVE coupled with LLE model yielded better results (79.0%). UVE coupled with LLE model with the presevation of classification accuracy only used 23.5% wavelength of LLE model. Therefore, it provides a useful algorithm for online classification and data saving.

Key words: Mealiness; Hyperspectral scattering images; Uninformative Variables Elimination (UVE); Locally Linear Embedding (LLE); Partial Least Squares Discriminant Analysis (PLSDA)