

基于加权欧氏距离的光谱重构训练样本选择

任澳¹, 孔玲君², 刘真¹, 王茜¹

(1.上海理工大学, 上海 200093; 2.上海出版印刷高等专科学校, 上海 200093)

摘要: **目的** 研究光谱反射率重建过程中训练样本的选取, 以提高光谱反射率的重建精度。**方法** 根据检验样本与训练样本间的欧氏距离判断两者的相似度, 并将样本向量各维度分量“标准化”到均值、方差相等, 使得各维度分别满足标准正态分布, 同时将其方差的倒数作为权重赋予训练样本。实验以 Munsell 色卡为总的训练样本集, 经 Mohammadi 方法、Cao 方法以及文中方法选择出的样本作为最终的训练样本, 用 Color Rendition Chart 24 色卡为检验样本, 采用伪逆法分别对选择出来的训练样本进行光谱反射率重建。**结果** 通过 Matlab 软件仿真实验, 文中方法的平均色差为 $0.7918 \Delta E_{ab}^*$, 最大色差为 $1.7148 \Delta E_{ab}^*$, 平均均方根误差为 0.0060, 最大光谱均方根误差为 0.0127。**结论** 基于加权欧氏距离的训练样本选择可以有效地提高光谱的重建精度, 能更好地实现颜色再现。

关键词: 样本选择; 加权欧氏距离; 光谱反射率重建

中图分类号: TS801.3 **文献标识码:** A **文章编号:** 1001-3563(2020)15-0253-07

DOI: 10.19554/j.cnki.1001-3563.2020.15.038

Spectral Reconstruction Training Sample Selection Based on Weighted Euclidean Distance

REN Ao¹, KONG Ling-jun², LIU Zhen¹, WANG Qian¹

(1.Shanghai University of Science and Technology, Shanghai 200093, China;

2.Shanghai Publishing and Printing College, Shanghai 200093, China)

ABSTRACT: The work aims to study the selection of training samples for spectral reflectance reconstruction to improve the reconstruction accuracy of spectral reflectance. The similarity between the test sample and the training sample was determined according to their Euclidean distance and the dimensional components of the sample vector were "normalized" to the average and the variance to be equal, so that the dimensions respectively met the standard normal distribution, and the reciprocal of the variance was given to the training samples as the weight. In the experiment, the Munsell color card was used as the total training sample set, the samples selected by Mohammadi method, Cao et al method and the method presented herein were used as the final training samples, and Color Rendition Chart 24 was used as the test samples. Spectral reflectance reconstruction was performed on the training samples respectively selected by pseudo-inverse method. Through the Matlab software simulation experiment, the average color difference of the proposed method was $0.7918 \Delta E_{ab}^*$, the maximum color difference was $1.7148 \Delta E_{ab}^*$, the average root-mean-square error was 0.0060, and the maximum spectral root-mean-square error was 0.0127. The selection of training samples based on weighted Euclidean distance can effectively improve the reconstruction accuracy of the spectrum and achieve better color reproduction.

收稿日期: 2019-11-25

基金项目: 上海理工大学科技发展项目 (2018KJFZ023); 绿色制版与柔印标准化实验室资助项目 (LGPSFP-03)

作者简介: 任澳 (1995—), 男, 上海理工大学硕士生, 主攻多光谱图像复制。

通信作者: 孔玲君 (1972—), 女, 博士, 上海出版印刷高等专科学校教授, 主要研究方向为图文信息处理与色彩再现、数字印刷及质量评价等。

KEY WORDS: sample selection; weighted Euclidean distance; spectral reflectance reconstruction

近年来,随着光谱成像技术的迅速发展,光谱成像技术在卫星遥感领域的研究在全球已经趋于成熟,但是在颜色复制领域的运用和研究还处于起步阶段^[1-3]。利用光谱成像技术可以保证图像在不同的媒介之间进行颜色的高保真复制,由于光谱反射率是物体表面的本征属性,不受照明条件和观测条件的影响,因此在准确描述物体表面的颜色信息方面起着重要的作用,可以更客观、更精确地对颜色进行表述^[4-6]。目前研究面临的关键问题是如何重建,以获得高精度的光谱反射率,由于光谱获取系统直接得到的是相机的光谱响应值,还必须使用光谱重建算法将成像通道响应值转换为光谱反射率,这个过程称为光谱图像的光谱反射率重建^[3]。在光谱反射率重建过程中,多通道相机到光谱反射率的转换模型是基于训练样本建立的,因此选择合适的训练样本对光谱反射率重建至关重要。经过相关科研人员的不断探索,已经提出的很多有效的选择训练样本的方法,包括 Hardeberg^[7-8]提出的最小条件数方法,即比较所选训练样本的光谱反射率矩阵的最大最小奇异值之比; Cheung 等^[9]提出的使新加入的训练样本之间的差异最大化的选择方法; Mohammadi^[10]提出的基于相关距离聚类的方法,将距离最小的聚为一类,再通过光谱反射率向量夹角最小选择最有代表性的样本;沈会良等^[11]的方法从选择的训练样本集与总样本集的光谱等效性出发;Cao 等^[12]提出了较小色差组加权的方法,这些方法的共同目的是选择出有典型代表性的训练样本集,达到少而精的效果。上述方法选择出的样本并不是最优训练样本组合,仅单一地基于光谱反射率空间或色彩空间选取,并没有考虑到与测试样本光谱特性相似度越高的训练样本应赋予较大权重,且训练样本应随着检验样本动态变化。文中在这些方法的基础上提出一种基于样本间加权欧式距离的训练样本选择方法来进行光谱反射率重建工作,并对其重构效果进行评价。

1 光谱反射率重建原理

首先假设光谱成像系统是理想的线性光电转换系统,则在实际照明环境下第 k 个通道像素响应值可以由式 (1) 得到。

$$S_k = \int_{\lambda_{\min}}^{\lambda_{\max}} l(\lambda) r(\lambda) f(\lambda) \tau(\lambda) d\lambda + b_k + n_k \quad (1)$$

式中: λ 为波长变量; S_k 为第 k 个通道的响应值; $l(\lambda)$ 为光源相对功率; $r(\lambda)$ 为物体表面光谱反射率; $f(\lambda)$ 为 CCD 相机的灵敏度函数; $\tau(\lambda)$ 为滤光片的透射率; b_k 为 CCD 相机暗电流噪声; n_k 为相机固定模式噪声; k 为第 k 个通道,文中选用 6 通道, k 取 1~6。

为了更加简明表达系统模型,令 $G_k(\lambda) = f(\lambda) \tau(\lambda) l(\lambda)$, 从而式 (1) 写成矩阵形式可以表示为:

$$\mathbf{S} = \mathbf{G}\mathbf{r} + \mathbf{B} + \mathbf{N} \quad (2)$$

式中: $\mathbf{G}$ 为光谱成像系统的灵敏度矩阵; $\mathbf{r}$ 为物体表面光谱反射率矩阵; $\mathbf{S}$ 为某一图像像素的响应值矩阵; $\mathbf{B}$ 和 $\mathbf{N}$ 为第 k 个通道的附加噪声矩阵,经暗电流和光照不均匀性校正后的光谱成像系统,可认为 $\mathbf{B}$ 和 $\mathbf{N}$ 均为 0, 因此式 (2) 可简化为:

$$\mathbf{S} = \mathbf{G}\mathbf{r} \quad (3)$$

光谱反射率重建就是式 (3) 的一个逆过程,输入重建色块的响应值,输出重建的光谱反射率,见式 (4)。

$$\hat{\mathbf{r}} = \mathbf{Q}\hat{\mathbf{S}} \quad (4)$$

式中: $\hat{\mathbf{r}}$ 为重建的光谱反射率; $\hat{\mathbf{S}}$ 为重建色块的响应值; $\mathbf{Q}$ 为多通道数字响应信号到光谱反射率之间的转换矩阵,研究光谱反射率重建的核心就是需要得到这个转换矩阵,它是由已知的训练样本的响应值以及光谱反射率通过相应的数学算法得到的,目前最常用的计算方法有伪逆法^[13]、wiener 估计法^[14]以及主成分分析^[15-16]等方法,伪逆法因其计算过程简单而在光谱反射率重建过程中使用较为频繁,其方法见式 (5)。

$$\hat{\mathbf{r}} = \mathbf{r}_{\text{train}} [\mathbf{S}_{\text{train}}]^{-1} \mathbf{S}_{\text{test}} \quad (5)$$

式中: $\hat{\mathbf{r}}$ 为重建的光谱反射率; $\mathbf{r}_{\text{train}}$ 为已知的训练样本的光谱反射率; $[\]^{-1}$ 为伪逆运算符号; $\mathbf{S}_{\text{train}}$ 为训练样本的响应值; $\mathbf{S}_{\text{test}}$ 为检验样本的响应值。文中基于加权欧氏距离选择训练样本,通过伪逆法来提高光谱反射率重建精度。

2 训练样本选择方法

光谱重构过程中选择合适的训练样本至关重要,需要估算不同样本光谱特性之间的相似性,而相似性可用样本间的“距离”来衡量。文中选择用欧式距离大小来衡量样本间的相似性,并针对欧氏距离在选择样本中的缺点而提出加权欧氏距离。

欧氏距离是指 2 点之间的直线距离^[17-18]。求 2 个 n 维向量 $a(x_{11}, x_{12}, \dots, x_{1n})$ 与 $b(x_{21}, x_{22}, \dots, x_{2n})$ 间的欧氏距离 D_{EUC} 可表示为:

$$D_{\text{EUC}} = \sqrt{\sum_{k=1}^n (x_{1k} - x_{2k})^2} \quad (6)$$

将欧氏距离看作训练样本与检验样本的相似度,距离越近就越相似。由于数据各维分量的分布不一样,这种度量方式将样本每个位置的权重对总体的影响认为是均等的,会有时导致不能满足实际的要求。且数据各个维度之间的尺度不一样,直接通过计算欧

氏距离来判定向量是否相似会产生误差。

文中提出的加权欧氏距离是将所有维度分别进行处理,先将各个维度分量都“标准化”到均值、方差相等,使得各个维度分别满足标准正态分布。求 2 个 n 维向量 $a(x_{11}, x_{12} \dots x_{1n})$ 与 $b(x_{21}, x_{22} \dots x_{2n})$ 间的加权欧氏距离 D_{SEUC} 可表示为:

$$D_{\text{SEUC}} = \sqrt{\sum_{k=1}^n \left(\frac{x_{1k} - x_{2k}}{S_k} \right)^2} \quad (7)$$

式中: S_k 为第 k 维度的标准差,方差的倒数就是一个权重,加权欧氏距离又可称为标准欧式距离。

Mohammadi 方法训练样本数量为 24, Cao 等研究了 4 种加权模型,其中选择了不同数量的训练样本,当训练样本数量为 3 时,其 RMSE 值达到最小。文中根据样本间距离大小进行相似性大小排序,选取前 P 个样本用来组成与测试样本相似性较高的训练样本子集。训练样本数量的选择基于色差与均方根误差最小化,当训练样本数量为 25 左右时重构误差最小。

3 实验与分析

3.1 数据获取及评价指标

实验选用 Munsell 1269 色卡作为训练样本集, Color Rendition Chart24 色卡(以下简称 RC24 色卡)作为检验样本集。在实验中使用 X-Rite 的 Eye-one pro 分光光度计在 D50 光源, 2° 视角下测量训练样本和检验样本的光谱反射率值,采样波段为 400 ~ 700 nm,以 10 nm 为采样间隔。同时用多光谱相机获取样本的 6 通道系统响应值,并使用 Matlab 2016 进行仿真实验。

采用 ΔE^*_{ab} 、均方根误差 RMSE 作为光谱反射率重建精度的评价指标^[19],其中用 RMSE 来衡量原始光谱反射率与重建光谱反射率之间的偏差。假设一个颜色的 n 个原始光谱反射率为 $R(r_1, r_2 \dots r_n)$,其重建得到的光谱反射率为 $\hat{R}(\hat{r}_1, \hat{r}_2 \dots \hat{r}_n)$,则 RMSE 可表示为:

$$R_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - \hat{r}_i)^2} \quad (8)$$

式中: r_i 为每一点像素的重构光谱反射率; $\hat{r}_i$ 为原始的光谱反射率; n 为样本维度,文中在波段 400 ~ 700 nm 以 10 nm 为间隔进行采样,故这里 n 取 31。

3.2 实验结果及分析

为方便记录色块信息,实验将 RC24 色卡中的色块依次按照从左到右,从上到下顺序进行编号,并用 1[#]—24[#]表示。为了使选出的色块更具有代表性,这里

选取 24 个色块中的 3[#], 6[#], 9[#], 16[#], 20[#]等 5 个色块用于实验结果分析,这 5 个色块基本上按照均匀间隔,且覆盖 ab 空间中的 4 个象限及中心点。

3.2.1 样本选择的有效性

1) 基于光谱反射率曲线及误差波动曲线分析。图 1 列出了使用文中方法、Mohammadi^[10]方法以及 Cao 等^[12]研究的 4 种加权模型中效果最好的平方反比加权模型方法,重建得到 5 个色块的光谱反射率曲线和原始光谱反射率曲线,以及光谱重建误差波动曲线。为了更容易观察光谱反射率重建误差波动情况,在每个光谱反射率曲线右边,对应着此样本光谱反射率重建误差波动曲线。

从图 1a—e 中的光谱反射率曲线可以看出,文中方法重建得到的光谱反射率曲线较为平滑,与实际光谱反射率曲线基本拟合,且走势基本相同。其中样本 20[#]重构效果最好,而样本 9[#]重构效果稍微逊色一些。从图 1f—j 中的误差波动曲线可以看出,文中方法重构的光谱反射率误差波动幅度整体明显较好,除了样本 9[#]在 550 ~ 650 nm 间误差波动幅度稍大,其他波段基本接近理想误差,而与 Mohammadi 和 Cao 等的方法相比较,无论是得到的光谱反射率曲线还是误差波动曲线文中方法重构的效果都明显优于他们。由此反映出文中提出的训练样本选择方法重构效果更好。

2) 基于样本在 ab 空间中的分布分析。训练样本的空间分布会影响样本的光谱重构精度,根据样本的光谱反射率值在 D50 光源、 2° 视角下计算得到其对应的 CIEXYZ 三刺激值,进而转换成 CIELAB 值。检验样本与训练样本的 a^* , b^* 分布见图 2。

由图 2 可以看出, Mohammadi 方法和 Cao 等方法没有考虑检验样本的光谱特性,直接从整个 Munsell 样本集选择代表性样本。筛选出的训练样本虽在 4 个象限中均有分布,但未随着检验样本的变化而发生改变。用文中方法筛选得到的训练样本均匀分布在检验样本的周围,筛选出来的训练样本也随检验样本的改变而动态变化。其中检验样本 20[#]筛选的训练样本均匀分布在其周围,且呈环状分布;检验样本 9[#]周围的训练样本分布相对离散。从前文分析可知,检验样本 20[#]光谱反射率重构效果最好,检验样本 9[#]的重构精度稍微差一点,由此反映出若筛选的训练样本在 ab 空间中均匀分布在检验样本周围,则光谱反射率重构效果相对较好。

从以上实验结果可以看出,文中方法选择的训练样本随着检验样本动态变化,根据每个检验样本的光谱特性不同从训练样本集中选择出光谱特性最相似的代表性训练样本,且筛选出的训练样本在 ab 空间中均匀分布在检验样本周围,这样光谱重构效果才好。

3.2.2 光谱反射率重构精度对比

3 种方法重构得到的均方根误差值与色差值见表 1。

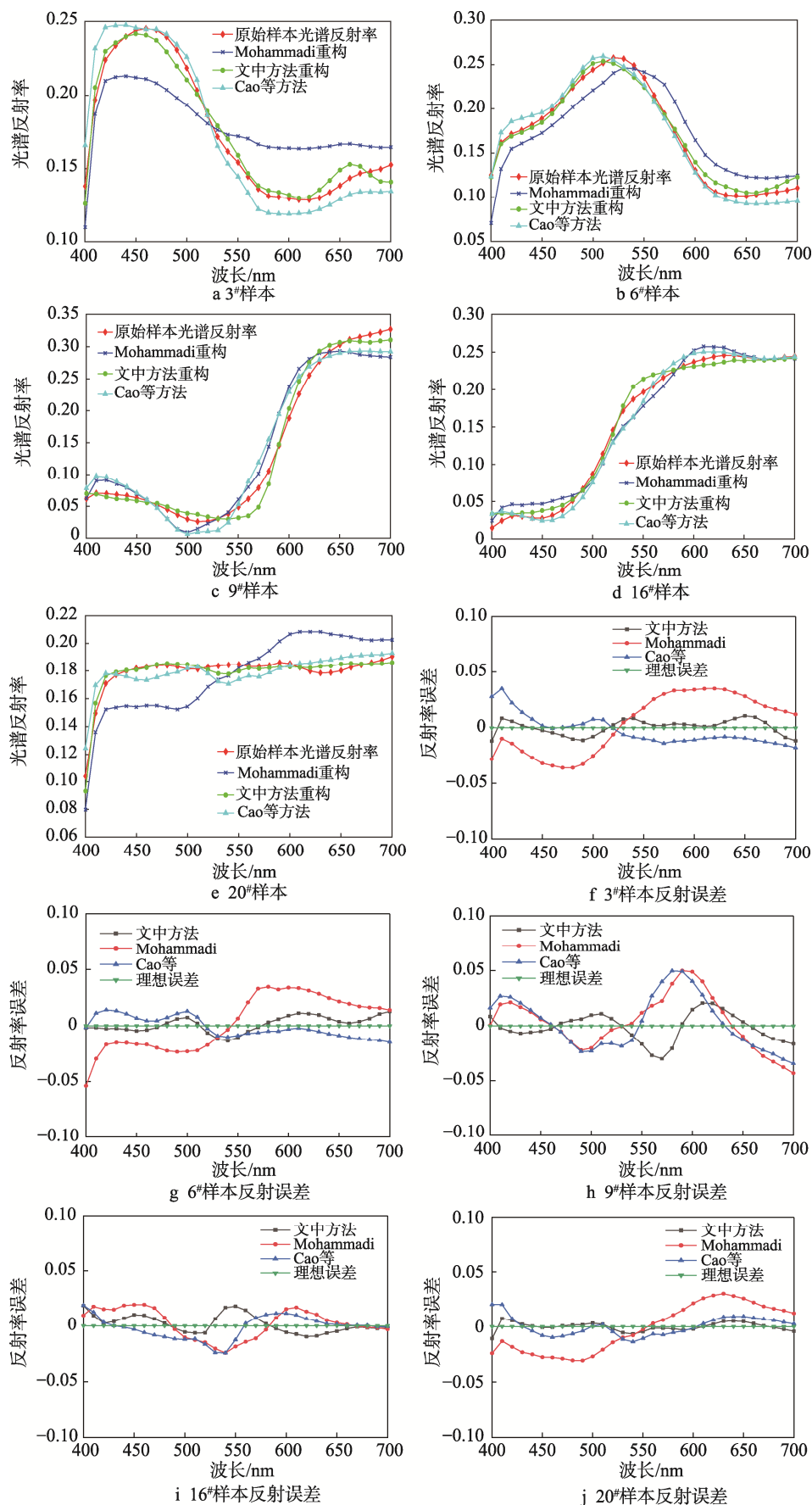


图1 光谱反射率曲线及误差波动曲线

Fig.1 Spectral reflectance curve and error fluctuation curve

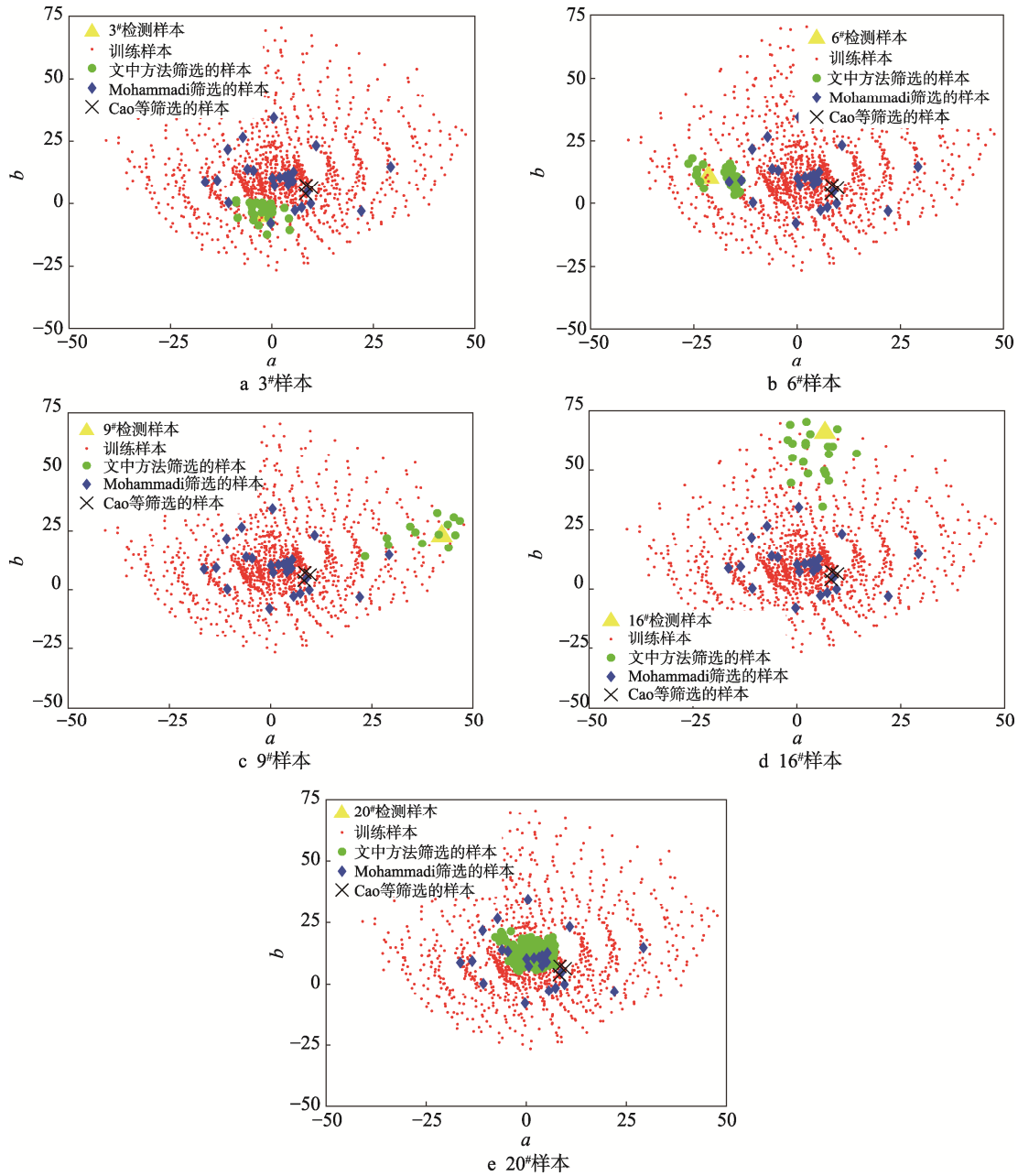


图 2 训练样本与检验样本在 ab 空间中的分布
Fig.2 Distribution of training samples and test samples in ab space

表 1 3 种训练样本选择方法的重构精度比较
Tab.1 Comparison of reconstruction accuracybetween threetraining sample selection methods

方法	ΔE_{ab}			RMSE		
	均值	最大值	最小值	均值	最大值	最小值
Mohammadi 方法	2.8136	9.2690	0.0489	0.0216	0.0263	0.0132
Cao 等方法	3.4808	10.5640	0.9150	0.0507	0.0935	0.0179
文中方法	0.7918	1.7148	0.0060	0.0060	0.0127	0.0039

从表 1 可以看出,文中提出的基于加权欧式距离选择训练样本的方法,得到重建光谱的平均色差值为 $0.7918 \Delta E_{ab}^*$,最大色差值为 $1.7148 \Delta E_{ab}^*$,最小色差值为 $0.0060 \Delta E_{ab}^*$,平均均方根误差值为 0.0060 ,比 Mohammadi 方法得到的平均均方根误差降低了 0.0156 ,平均色差值降低了 $2.0218 \Delta E_{ab}^*$,比 Cao 等方

法的光谱重构精度也有较大提升。

结果表明,不论是从色度精度方面还是光谱精度方面,文中方法均优于 Mohammadi 方法和 Cao 等方法,说明了基于文中提出的加权欧式距离选择训练样本方法的光谱重建精度有显著提高。

4 结语

文中对光谱反射率重建过程中训练样本的选取这一重要环节进行研究,提出了基于加权欧式距离选择训练样本的方法。首先把向量所有维度分别进行处理,先将各个维度分量都“标准化”到均值、方差相等,使得各个维度分别满足标准正态分布,然后将方差的倒数赋予该训练样本作为权重大小,利用加权后的欧氏距离来选择训练样本进行光谱重建。其次考虑到如果训练样本的光谱特性与检验样本的光谱特性越相似,则重构精度就越高,根据每一个检验样本光谱特性选取与之相似度较高训练样本,生成自适应矩阵来进行光谱反射率重建。从实验结果可以看出,基于加权欧式距离的光谱反射率重构选择训练样本的方法获得的光谱精度显著提高,优于 Mohammadi 方法和 Cao 等的方法,对光谱反射率重建过程中训练样本的选择有一定的参考意义。

参考文献:

- [1] 龙艳群,王慧琴,王可,等. 基于样本优化选取的光谱重建方法研究[J]. 影像科学与光化学, 2017, 35(1): 88—96.
LONG Yan-qun, WANG Hui-qin, WANG Ke, et al. Spectral Reconstruction Method Based on Optimized Sample Selection[J]. Image Science and Photochemistry, 2017, 35(1): 88—96.
- [2] JING X, YUE T X, ZHENG L. Research on Spectral Reflectance Reconstruction Algorithms for Munsell Color Card[C]// International Conference on Computer Science and Information Technology (CSAIT), 2014: 663—669.
- [3] 何颂华,李玉梅. 光谱颜色复制技术[M]. 北京: 中国农业出版社, 2018: 10.
HE Song-hua, LI Yu-mei. Spectral Color Reproduction Technology[M]. Beijing: China Agriculture Press, 2018: 10.
- [4] RIBES A, SCHMITT F, BRETTEL H. Calibration and Spectral Reconstruction for an Art Painting Multispectral Acquisition System[C]// Conference on Color in Graphics, 2015: 498—530.
- [5] 杨萍. 多基色数字相机的光谱重构理论与技术[M]. 北京: 中国轻工业出版社, 2017: 3.
YANG Ping. Spectral Reflectance Reconstruction Based on Multicolor Digital Camera: Principle and Technology[M]. Beijing: China Light Industry Press, 2017: 3.
- [6] 易文娟,孙刘杰,陈志文,等. 基于聚类分析的光谱重建样本分析[J]. 包装工程, 2019, 40(17): 249—255.
YI Wen-juan, SUN Liu-jie, CHEN Zhi-wen, et al. Spectral Reconstruction Sample Analysis Based on Cluster Analysis[J]. Packaging Engineering, 2019, 40(17): 249—255.
- [7] HARDEBERG J Y. Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches[J]. Color Research & Application, 2002, 27(4): 304—305.
- [8] HARDEBERG J Y, SCHMITT F, BRETTEL H, et al. Multispectral Color Image Capture Using Liquid Crystal Tunable Filter[J]. Optical Engineering, 2002, 41(10): 2532—2548.
- [9] CHEUNG V, WESTLAND S. Method for Optimal Color Selection[J]. Journal of Imaging Science and Technology, 2006, 50(5): 481—488.
- [10] MOHAMMADI M, NEZAMABADI M, BEMS R S, et al. Spectral Imaging Target Development Based on Hierarchical Cluster Analysis[C]// 12th Color Imaging Conference, Color Science and Engineering Systems Scottsdale, 2004: 59—64.
- [11] SHEN H L, ZHANG H G, XIN J H, et al. Optimal Selection of Representative Colors for Spectral Reflectance Reconstruction a Multispectral Imaging System[J]. Applied Optics, 2008, 47(13): 2494—2502.
- [12] CAO B, LIAO N F, CHENG H B. Spectral Reflectance Reconstruction from RGB Images Based on Weighting Smaller Color Difference Group[J]. Color Research and Application, 2017, 42(3): 327—332.
- [13] 曾茜,孔玲君,占文杰. 基于向量夹角样本选择的光谱反射率重建[J]. 包装工程, 2018, 39(15): 216—220.
ZENG Xi, KONG Ling-jun, ZHAN Wen-jie. Spectral Reflectance Reconstruction Based on Vector Included Angle Sample Selection[J]. Packaging Engineering, 2018, 39(15): 216—220.
- [14] 李守礼. 面向复制的多光谱获取关键技术研究[J]. 科技风, 2017(1): 1.
LI Shou-li. Research on Key Technologies of Replication-oriented Multispectral Acquisition [J]. Technology Style, 2017(1): 1.
- [15] SHANG M, YANG L, LIU D, et al. Emission Spectral Reconstruction Based on Principal Component Analysis Applied to Fluorescence Full-color Prints[J]. Nanoence and Nanotechnology Letters, 2019, 11(10): 1349—1356.

- [16] 李婵, 万晓霞, 刘强, 等. 基于主成分分析的光谱重建训练样本选择方法研究[J]. 光谱学与光谱分析, 2016, 36(5): 1400—1405.
LI Chan, WAN Xiao-xia, LIU Qiang, et al. Study on Selection Methods of Spectral Reconstruction Training Samples Based on Principal Component Analysis[J]. Spectroscopy and Spectral Analysis, 2016, 36(5): 1400—1405.
- [17] 陈静杰, 车洁. 基于标准欧式距离的燃油流量缺失数据填补算法[J]. 计算机科学, 2017, 44(6): 109—112.
CHENG Jing-jie, CHE Jie. Fuel Flow Missing Data Padding Algorithm Based on Standard Euclidean Distance[J]. Computer Science, 2017, 44(6): 109—112.
- [18] 刘占宁, 宋宇辰, 孟海东, 等. 样本点与距离权重对距离幂次反比法的影响[J]. 矿业研究与开发, 2018, 38(9): 95—99.
LIU Zhan-ning, SONG Yu-chen, MENG Hai-dong, et al. Influence of Sample Point and Distance Weight on Distance Power Inverse Ratio Method[J]. Mining Research and Development, 2018, 38(9): 95—99.
- [19] 刘浩学. CIE 均匀颜色空间与色差公式的应用[J]. 北京印刷学院学报, 2003, 11(3): 3—9.
LIU Hao-xue. The Application of CIE Uniform Color Space and Its Color Difference Formula[J]. Journal of Beijing Institute of Graphic Communication, 2003, 11(3): 3—9.