

基于样本选择的光谱重构研究

易文娟, 张雷洪

(上海理工大学, 上海 200093)

摘要: **目的** 为了提高使用主成分分析法重构光谱反射率的重构精度。**方法** 利用 Matlab 进行仿真实验, 选择 3 种不同色卡作为训练样本, 使用主成分分析法探究主成分个数和样本间隔对重构结果的影响。**结果** 主成分个数为 4 时, 贡献率均超过 99%; 样本间隔为 10 nm 时, RC24 色卡重构效果最好, 其平均色差 $2.37 \Delta E_{ab}^*$ 平均均方根误差为 0.0185。**结论** 训练样本的选择会影响光谱重构精度, RC24 色卡具有数据量小、重建精度较高的特点, 在颜色复制领域可以优先选择。

关键词: 光谱反射率; 主成分分析法; 训练样本; 重建精度

中图分类号: TP801.3 **文献标识码:** A **文章编号:** 1001-3563(2018)13-0233-06

DOI: 10.19554/j.cnki.1001-3563.2018.13.037

Spectral Reconstruction Based on Sample Selection

YI Wen-juan, ZHANG Lei-hong

(University of Shanghai for Science and Technology, Shanghai 200093, China)

ABSTRACT: The work aims to improve the reconstruction accuracy of spectral reflectance with principal component analysis (PCA). Matlab was used to conduct simulation experiments. 3 different color cards were selected as the training samples and the effect of the number of principal components and the sample interval on the reconstruction results was investigated by the method of PCA. When the number of principal components was 4, the contribution rate was over 99%. When the sample interval was 10 nm, the RC24 color card had the best reconstruction effect, its average color difference was $2.37 \Delta E_{ab}^*$ and the average root-mean-square error was 0.0185. The selection of training samples will affect the accuracy of spectral reconstruction. With the characteristics of small amount of data and high accuracy of reconstruction, RC24 color card can be preferred in the field of color reproduction.

KEY WORDS: spectral reflectance; principal component analysis (PCA); training sample; reconstruction accuracy

光谱成像技术可以不受环境光源照射条件、显示设备色彩特性的影响, 可以全面准确地保存物体的颜色信息^[1-5], 因此光谱匹配在信息传输、高保真图像复制、艺术品修复再现等方面有广阔的应用前景。光谱反射率重构的常用算法有伪逆法 (Pseudo-inverse)^[3-5]、主成分分析法 (PCA)^[3-6]、维纳估计法^[5, 7]、R 矩阵法^[8]等, 其中主成分分析法是一种基于学习的重建方法对高维度的光谱数据进行实现数据变量降维处理^[2], 对实验条件要求较低且重建精度也能达到应用要求, 因此一直是研究的热点。

目前国外进行多光谱图像色彩再现研究的主要机构有美国罗切斯特理工学院、日本千叶大学、日本东京工业大学等, 我国武汉大学、北京理工大学、浙江大学等高校也对光谱复制做出了不同方向的研究。陈艺奕^[8]采用伪逆法、主成分分析法等方法分别对数码相机的物体表面色光谱重构, 结果表明主成分分析法具有较高的精度。何成栋^[9]采用正交试验方法重构检验样本分析颜色三属性对光谱重建精度的影响, 结果表明样本选择对主成分分析法的影响大于对伪逆法算法的影响。于海琦^[10]的实验表明 RC 色卡的 7 个

收稿日期: 2018-01-09

基金项目: 易文娟 (1995—), 女, 上海理工大学硕士生, 主攻多光谱复制。

作者简介: 张雷洪 (1981—), 男, 博士, 上海理工大学副教授, 主要研究方向为压缩感知重构算法、关联成像算法以及高光谱重构算法。

主成分对光谱图像重构是达到较高颜色复制精度的最优选择,但是主成分数量过多会带来数据冗余耗时较长的问题。目前基于主成分分析法的研究日益增多^[8-13],主流的样本选择方法是选择某种具有典型代表性的训练样本集,从表征样本集颜色的不同属性着手分析^[5],但关于不同样本的选择及其重建波动性的相关研究较少,文中采用主成分分析法探讨不同数量的训练样本和样本间隔对主成分分析法重构结果的影响,实现对大量数据的压缩和特征提取,有助于进一步提高光谱反射率的重构精度。

1 主成分分析法

主成分分析法通过降维可以减少数据集的维数同时保持数据集对方差的贡献特征,实现降低空间维度和计算复杂度^[6]。用 M 表示光谱反射率 R 和相机响应值 V 之间的转换矩阵,则光谱反射率 R 可表示为:

$$R = MV \quad (1)$$

$n \times m$ 维 (n 为光谱维数, m 为样本数) 的光谱反射率样本集 r 可表示为 $r = (r_1, r_2, r_3, \dots, r_m)$, 也可以看做是 j 个彼此正交的 n 维特征向量 e_j 的线性组合, 由线性代数基础:

$$r = \sum_{i=1}^j e_i a_i = E \cdot A \quad (2)$$

$E = (e_1, e_2, e_3, \dots, e_j)$ 为提取的特征向量, $A = (a_1, a_2, a_3, \dots, a_j)^T$ 为对应的转换系数。将 $r = MV$ 代入式 (1) 可得:

$$M = \sum_{i=1}^j e_i a_i \cdot V^{-1} \quad (3)$$

2 样本选择

窄带光谱成像系统中的光谱反射率重构大多是通过训练样本获得系统数字响应信号到光谱反射率值的转换矩阵, 再进一步完成物体表面光谱反射率重建^[5], 因此作为光谱重建中建立多通道信号与光谱反射率之间转换矩阵的基础, 训练样本的选择对光谱重建有着重要的作用。

训练样本选择是指对样本集颜色的不同属性分析^[5], 目前大多是选择具有某一训练样本集进行光谱重建工作, 本文选择多种训练样本集并分析其重建结果。实验选用标准色卡 ColorChecker Color Rendition Chart (以下简称 RC24 色卡) 和 ColorCheckerSG (以下简称 SG140 色卡) 以及 MUNSELL 色卡。

RC24 色卡的各个色块可以在各种光源下与其代表的自然物体的颜色相匹配, 常用于色彩再现流程为还原各种照明条件和介质上的图像色彩提供参照。

SG140 色卡有 140 种高饱和度的亚光色彩, 常用于比较数码还原的色彩与真实场景或者与试标板的色差。MUNSELL 色卡包含了 1269 高光泽的颜色, 被广泛的应用到艺术设计、包装产品设计以及质量控制等行业^[2]。

由于海琦^[10]的实验可知 MUNSSELL 和 RC24 色卡的颜色总体分布趋势相似但均匀性不一致且 RC24 色卡色精度最高。D50 条件下 3 种色卡的 LAB 三维分布见图 1。L 为匹配人类亮度的感知即心理亮度; ab 为心理色度, a 由负值到正值表示绿-红的变化, b 由负值到正值表示蓝-黄的变化。由图 1 可以得到 MUNSSELL 分布较密集, SG140 色卡分布广泛但不够均匀, RC24 色卡的分布广泛且均匀, 因此实验选择 RC24 色卡 (见图 2) 作为检验样本。

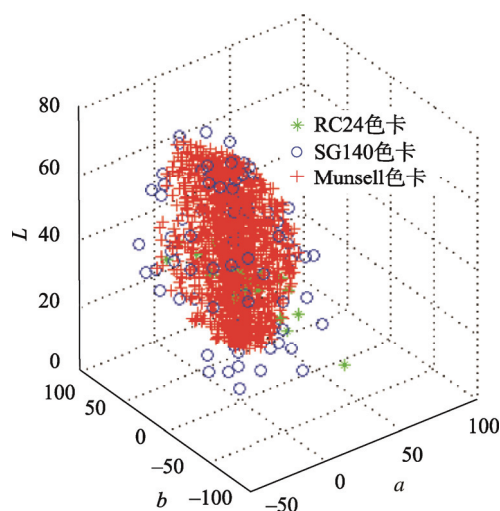


图 1 D50 条件下 3 种色卡色度比较

Fig.1 Chromaticity comparison of 3 color cards under D50 light source



图 2 RC24 色卡

Fig.2 RC24 color card

3 评价方法

实验根据主成分个数的累计贡献率探究主元分析法 (PCA) 中主成分个数对实验结果的影响, 累计贡献率公式为:

$$V_k = \frac{\sum_{i=1}^k W_i}{\sum_{i=1}^n W_i} \tag{4}$$

式中： V_k 为前 k 个主成分的累计贡献率； W_i 为第 i 个主成分的贡献量； n 为总主成分个数。

3.1 均方根误差

均方根误差能够评估重构值和真实值的物理差异，均方根误差对一组测量中极值的反映非常灵敏，但是它没有考虑人眼视觉特性的影响，只能反映光谱曲线形状之间的差异。

均方根误差（RMSE）的计算公式为：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_i - \bar{r}_i)^2} \tag{5}$$

式中： r_i 为每一点像素的重构光谱反射率； $\bar{r}_i$ 为原始光谱反射率； n 为样本维度。

3.2 色差

评估光谱颜色复制精度的最终标准是在各种照明环境下的人类彩色视觉。CIE2000 色差公式的使用标准条件是 D65 的光源，但是目前典型的用于跨媒体彩色复制工艺流程中的国际标准化组织的参考色域（ISORCG）为 D50 光源^[11]，因此为了后续的顺利输出实验采用 D50 光源。根据 CIE1976 色差公式，CIE Lab 均匀颜色空间中的颜色 $A(L_1, a_1, b_1)$ 和颜色 $B(L_2, a_2, b_2)$ 两者色差值 ΔE_{ab} 表示为：

$$\Delta E_{ab} = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \tag{6}$$

4 仿真实验

实验选用 GretagMacbeth 公司 2 套标准色卡 ColorChecker Color Rendition Chart（以下简称 RC24 色卡）和 ColorCheckerSG（以下简称 SG140 色卡）以及 MUNSELL3 种色卡作为训练样本，D50 光源、2° 视场作为实验条件，使用 MegaVision 测量实验样本的光谱反射率，在 Matlab 软件上进行仿真实验，实验过程如下所述。

1) 探究主元个数对光谱反射率重构的影响，设定主元个数 $i(i=1,2...9)$ 进行实验，取得不同主成分个数的特征值和累计贡献率，使用 RMSE 和 ΔE_{ab} 比较重构结果。

2) 探究样本间隔对重构结果的影响，取 10, 20, 30 nm 为间隔进行采样，使用 RMSE 和 ΔE_{ab} 比较重构结果。

4.1 主成分个数对重构结果的影响

该实验选取累计贡献率 99.5% 作为评价标准，这样既能保持绝大部分训练样本的信息又达到减少变

量提高计算速度，实验结果见表 1。

表 1 主元个数对不同训练样本的影响
Tab.1 The influence of the number of principal components on different training samples

主元个数	SG140 色卡		RC24 色卡		MUNSELL1269 色卡	
	特征值	累计贡献率/%	特征值	累计贡献率/%	特征值	累计贡献率/%
1	82.755	82.755	66.148	66.148	76.766	76.766
2	12.611	95.366	25.153	91.301	15.832	92.598
3	3.629	98.994	8.249	99.550	5.961	98.558
4	0.532	99.526	0.414	99.964	0.757	99.315
5	0.255	99.781	0.024	99.988	0.369	99.684
6	0.072	99.853	0.011	99.999	0.119	99.803
7	0.061	99.914	0.000	99.999	0.090	99.893
8	0.037	99.952	0.000	99.999	0.039	99.932
9	0.026	99.977	0.000	99.999	0.032	99.964

由表 1 可得，主成分个数增加，特征值降低，累计贡献率增加；主成分个数由 2 增至 3 时累计贡献率有明显的增加，此后累计贡献率虽有增加但增加幅度不大，因此选择主成分个数为 3—8 可以很好地反应训练样本的信息；当主成分个数为 4 时 RC24 色卡的 V_k 超过 99.9%，此时 SG140 色卡和 Munsell 色卡 V_k 也超过 99%，此时包含了训练样本的绝大部分信息可以对训练样本进行降维。

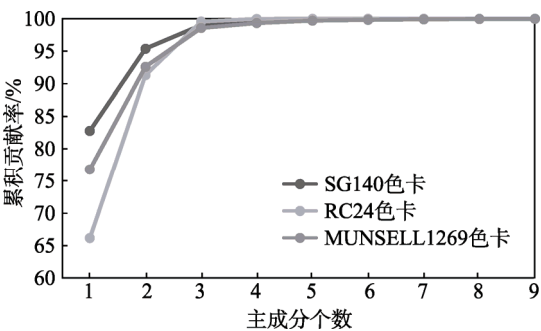


图 3 主成分个数与累计贡献率的关系
Fig. 3 Relationship between principal component number and accumulative contribution rate

将不同主成分个数时 3 种色卡的累计贡献率见图 3，可得知，当主成份个数增至 3 时累积贡献率有明显上升然后趋于平稳，当主成分个数为 3 时 RC24 色卡的累计贡献率最大，此时已经包含训练样本的大部分信息并可以保持较小的计算量，不同主元个数得到的重构色差和重构均方根误差见表 2—3。

分析表 2, 3 可得，重构的色差、均方根误差随着主成分个数增加而减小。当主成分个数为 3 时平均色差显著降低，主成分个数由 4 增加至 9 时色差变化不

表 2 不同主元个数的色差评价
Tab.2 Evaluation of chromatic aberration of different principal components

主元个数	SG140 色卡			RC24 色卡			MUNSELL1269 色卡		
	MEAN	STD	MAX	MEAN	STD	MAX	MEAN	STD	MAX
1	21.58	11.70	38.28	24.35	10.09	47.75	22.62	10.40	42.78
2	14.48	10.90	37.03	16.21	10.86	37.64	15.66	11.25	38.58
3	4.15	2.86	11.98	2.43	1.44	6.54	3.60	1.88	9.39
4	4.34	2.84	11.47	2.37	1.77	6.93	3.75	1.99	9.89
5	4.42	2.90	11.50	2.25	1.89	6.72	3.76	2.03	9.94
6	4.47	2.90	11.53	2.27	1.95	6.79	3.77	2.05	10.01
7	4.48	2.89	11.52	2.27	1.95	6.79	3.77	2.04	10.00

表 3 不同主元个数的均方根误差评价
Tab.3 Evaluation of root-mean-square error of different principal components

主元个数	SG140 色卡			RC24 色卡			MUNSELL1269 色卡		
	MEAN	STD	MAX	MEAN	STD	MAX	MEAN	STD	MAX
1	0.0716	0.0343	0.1328	0.0786	0.0266	0.1299	0.0734	0.0289	0.1221
2	0.0449	0.0279	0.1026	0.0475	0.0269	0.0966	0.0457	0.0274	0.0927
3	0.0245	0.0147	0.0651	0.0198	0.0130	0.0511	0.0226	0.0134	0.0509
4	0.0237	0.0138	0.0591	0.0185	0.0123	0.0483	0.0226	0.0133	0.0513
5	0.0236	0.0137	0.0591	0.0173	0.0123	0.0479	0.0224	0.0134	0.0514
6	0.0232	0.0138	0.0590	0.0170	0.0124	0.0477	0.0224	0.0134	0.0514
7	0.0232	0.0139	0.0590	0.0170	0.0124	0.0477	0.0227	0.0132	0.0513

明显；主成分个数为 1 或 2 时，SG140 色卡重构得到的 RMSE 值最小，主成分个数大于 3 之后 RC24 色卡重构得到的 RMSE 值最小；当主成分个数不小于 2 时，RC24 色卡可以得到最小的色差平均值，Munsell 色卡得到的色差在均值附近更为紧密但色差均值略大；主成分个数为 4 时，RC24 色卡可得到最小的色差平均值 $2.37 \Delta E_{ab}^*$ 和最小的 RMSE 平均值 0.0185。由于色差会受到光源和标准观察者的影响，光源和标准观察者对整个光谱带上不同波长的光谱的作用是不同的，但均方根误差并没有考虑人眼视觉的特性，它是一种纯光谱曲线形状之间的差异，且由于 CIE1976 色差公式的不均匀性，因此有出现重构的均方根误差较小但色差较大的情况；当主成分个数为 4 时，此时 SG140 色卡累计贡献率比 Munsell 色卡累积贡献率高，但由于 RC24 色卡与 Munsell 色卡的基底和颜料相似且 Munsell 色卡色域更广泛，Munsell 色卡得到色差平均值 $3.75 \Delta E_{ab}^*$ 和 RMSE 平均值 0.0226

优于 SG140 色卡的重构效果。

4.2 样本间隔对重构结果的影响

波长范围及采样间隔对光谱表征精度的影响程度和所表示光谱的物体介质类型有关^[14-15]，因此有必要探讨采样间隔对光谱重构的影响。在 D50 光源下 3 种训练样本在不同间隔的重构色差见表 4，均方根误差见表 5。

由表 4 和表 5 可得，D50 光源的光谱曲线较为平直，因此可以比较出随着样本间隔减小，3 种色卡的重构色差和均方根误差的平均值、标准差和最大值降低。样本间隔相同时，RC24 色卡的平均均方根误差和平均色差值最小，SG140 色卡的平均均方根误差和平均色差值最大。由于采样间隔不同，光谱空间的维度也不同，这对图像复制的获取计算和传输存储至关重要。当间隔为 10 nm 时，RC24 色卡的重构均方根误差最小，其平均均方根误差 0.0185，平均色差为

表 4 不同间隔的色差评价
Tab.4 Evaluation of chromatic aberration of different intervals

间隔/nm	SG140 色卡			RC24 色卡			Munsell 色卡		
	MEAN	STD	MAX	MEAN	STD	MAX	MEAN	STD	MAX
10	4.33	2.84	11.47	2.37	1.77	6.93	3.74	1.99	9.88
20	5.39	3.75	13.59	2.93	2.34	9.45	4.11	2.02	9.57
30	5.82	4.07	14.69	3.21	2.55	10.30	4.37	2.19	10.24

表 5 不同间隔的均方根误差评价
Tab.5 Evaluation of root-mean-square error of different intervals

间隔/nm	SG140 色卡			RC24 色卡			Munsell 色卡		
	MEAN	STD	MAX	MEAN	STD	MAX	MEAN	STD	MAX
10	0.0237	0.0138	0.0591	0.0185	0.0123	0.0483	0.0226	0.0133	0.0513
20	0.0346	0.0200	0.0867	0.0270	0.0179	0.0709	0.0319	0.0187	0.0710
30	0.0439	0.0252	0.1099	0.0343	0.0224	0.0899	0.0387	0.0227	0.0844

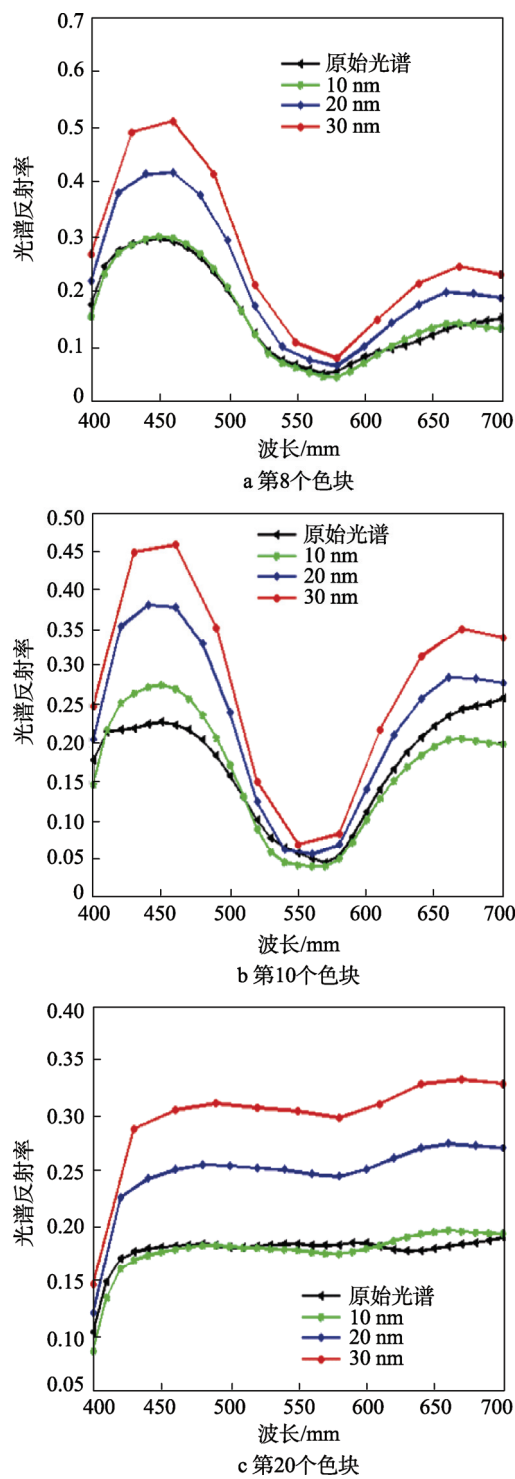


图 4 不同间隔时部分色块的重构曲线
Fig.4 Reconstruction curves of partial color blocks at different intervals

2.37 ΔE_{ab}^* ，此时重构效果良好；由图 4 可知，不同样本间隔对不同波段、色块的重构曲线影响程度不同。从第 20 个色块的 3 条重构曲线可以看出不同间隔对其影响明显，对第 8 和 10 个色块进行分析得到 550 nm 附近样本间隔的选择对该色块的重构影响并不明显即采样间隔对黄绿波段的影响较小，但对 450 nm 波段附近影响较大即采样间隔对紫红色波段的影响明显。

5 结语

主成分分析法的研究越来越受到关注，因此有必要探究样本选择对此方法重构结果的影响。文中使用 3 种训练样本研究了主成分个数和样本间隔对重构结果的影响，实验表明不同色块数量的训练样本会影响光谱重构精度，确定了主成分个数为 4，采样间隔 10 nm 时可以对检验样本进行良好的重构。此时训练样本选择 RC24 色卡可以得到最小的色差平均值 2.371 ΔE_{ab}^* 和最小的均方根误差平均值 0.0198。该实验的研究成果有助于进一步研究不同波段的颜色重构从而进一步提高光谱反射率的重构精度，在跨媒体出版、彩色艺术品维护修复、高保真数字图像复制以及彩色纺织印染等相关领域有一定的建议作用。

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. Study on Spectral Reconstruction Method Based on Optimized Selected Samples[J]. Imaging Science and Phorochemistry, 2017, 35(1): 88—96.

[2] 兑利涛. 多光谱图像的光谱反射率重建技术研究[D]. 南昌: 华东交通大学, 2017.
DUI Li-tao. Spectral Reflectance Reconstruction of Multispectral Images[D]. Nanchang: East China Jiaotong University, 2017.

[3] WANG H W, LI J, CHEN G X. Study on Key Issues of Multi-spectral Color Reproduction Technique[J]. Advanced Materials Research, 2011, 174: 93—96.

[4] ZHU Y, LI B, XU X. Spectral Reconstruction and Accuracy Appraisal Based on Pseudo Inverse Method[C]//

- Photonics and Optoelectronics (SOPO), 2012 Symposium on IEEE, 2012: 1—3.
- [5] 梁金星, 万晓霞, 王琪. 光谱重建训练样本选择方法研究[J]. 包装工程, 2016, 37(7): 125—130.
LIANG Jin-xing, WAN Xiao-xia, WANG Qi. Selection Method of Training Samples for Spectral Reconstruction[J]. Packaging Engineering, 2016, 37(7): 125—130.
- [6] HARIFI T, AMIRSHAHI S, AGAHIAN F. Recovery of Reflectance Spectra from Colorimetric Data Using Principal Component Analysis Embedded Regression Technique[J]. Optical Review, 2008, 15(6): 302—308.
- [7] SONG J, KIM C, YOO Y. Vein Visualization Using a Smart Phone with Multispectral Wiener Estimation for Point-of-care Applications[J]. Biomedical & Health Informatics IEEE Journal of, 2015, 19(2): 773—778.
- [8] 陈奕艺, 徐海松, 张显斗, 等. 基于数码相机的光谱重构研究[J]. 光学学报, 2009, 29(5): 1416—1419.
CHEN Yi-yi, XU Hai-song, ZHANG Xian-dou, et al. Study of Spectral Reconstruction Based on Digital Camera[J]. Acta Optica Sinica, 2009, 29(5): 1416—1419.
- [9] 何成栋, 黄新国, 张姗姗. 基于颜色属性的光谱重建训练样本正交优化[J]. 包装学报, 2016, 37(3): 66—70.
HE Cheng-dong, HUANG Xin-guo, ZHANG Shan-shan. Orthogonal Optimization of Spectral Reconstruction Training Samples Based on Color Properties[J]. Packaging Journal, 2016, 37(3): 66—70.
- [10] 于海琦, 刘真, 张雷洪, 等. 样本特征对光谱图像重构影响的研究[J]. 包装工程, 2014, 35(13): 144—149.
YU Hai-qi, LIU Zhen, ZHANG Lei-hong, et al. Effects of Sample Characteristics on Spectral Image Reconstruction[J]. Packaging Engineering, 2014, 35(13): 144—149.
- [11] 李天一. 实际物体色色域的研发[D]. 鞍山: 辽宁科技大学, 2016.
LI Tian-yi. The Development of The Real Object Colour Gamut[D]. Anshan: University of Science and Technology Liaoning, 2016.
- [12] ZHANG Y, ZHOU S S. Research on Reconstruction of Spectral Reflectance Based on Principal Component Analysis[J]. Applied Mechanics and Materials, 2013, 262: 53—58.
- [13] JING X, YUE T X, ZHENG L. Research on Spectral Reflectance Reconstruction Algorithms for Munsell Color Card[M]. Springer India, 2014.
- [14] SHAMS-NATERI A. Wavelength Intervals Effect on Reflectance Spectra Reconstruction[J]. Optica Applicata, 2012, 42(4): 737—742.
- [15] ZHANG X D, WANG Q, LI J C, et al. The Interim Connection Space Based on Human Color Vision for Spectral Color Reproduction[J]. Optical Society of America. Journal A: Optics, Image Science, and Vision, 2012, 29(6): 1027.