

# 一种基于色域分析与聚类分析的基色筛选

李金城<sup>1</sup> 刘 真<sup>1</sup> 陈广学<sup>2</sup> 杨晟炜<sup>1</sup> 余节约<sup>1</sup>

(<sup>1</sup> 南京林业大学轻工科学与工程学院, 江苏 南京 210037)  
(<sup>2</sup> 华南理工大学轻工与食品学院, 广东 广州 510640)

**摘要** 提出了一种基于色域分析与聚类分析结合的基色筛选算法。该方法对传统 4 色色域和目标高保真色域进行分析, 筛选出 4 色色域外的特征点, 并对其光谱反射率进行主成分分析, 确定基色的个数  $m$ 。以  $m$  作为分类数, 对集合中特征色光谱反射率进行聚类分析, 在  $m$  个分类光谱中选择和特征色平均光谱距离最远的  $m$  个光谱作为基色光谱, 在基色库中选择和  $m$  个光谱相关距离最小的实际专色作为最终基色。实验结果表明, 提出的基色选择算法能够准确估计出新基色的光谱反射率, 最大程度实现了目标高保真色域的再现。

**关键词** 视觉光学; 视觉与色彩; 多基色高保真复制; 基色估计; 色域分析

中图分类号 O436 文献标识码 A doi: 10.3788/AOS201232.0633001

## Colorant Selection Based on Gamut Analysis and Cluster Analysis

Li Jincheng<sup>1</sup> Liu Zhen<sup>1</sup> Chen Guangxue<sup>2</sup> Yang Shengwei<sup>1</sup> Yu Jieyue<sup>1</sup>

(<sup>1</sup> College of Light Industry Science and Engineering, Nanjing Forestry University, Nanjing, Jiangsu 210037, China)  
(<sup>2</sup> College of Light Industry and Food Sciences, South China University of Technology, Guangzhou, Guangdong 510640, China)

**Abstract** A method of basis colorant selection based on gamut analysis and cluster analysis is proposed. Conventional 4-color gamut and target hifi-color gamut are analyzed to pick out the featured color sets where the colorants exist, and the principal component analysis is performed for the spectral reflectance of these colors to get the numbers of basis colorant  $m$ . With the cluster number  $m$ , spectral reflectance of the featured colors in the set is divided by clustering method, and the average spectral reflectance of the featured colors is calculated.  $m$  spectral reflectance among the  $m$  cluster which are most nearest to the average spectral reflectance is selected as the spectral reflectance of the basis colorant. The results show that the proposed method can estimate spectral reflectance of new basis colorant accurately, which realises the maximum reproduction of the target high fidelity color gamut.

**Key words** visual optics; vision and color; multi-color reproduction; basis colorant selection; gamut analysis

**OCIS codes** 330.1730; 100.2000; 110.1758

## 1 引 言

基色色料估计是多基色高保真复制的关键技术之一。现有的基色色料估计方法主要有两种, 即基于光谱复制模型的基色估计和基于色域最大化的基色估计。基于光谱复制模型的基色估计流程分为 3

个步骤: 1) 基于 Kubelka-Munk 的线性混合颜色空间变换; 2) 光谱主成分分析; 3) 条件旋转变换<sup>[1~3]</sup>。此外, 也有研究者<sup>[4]</sup>提出了使用旋转主成分分析方法进行基色色料估计。另一种基色估计算法是在确定基色个数的情况下利用计算机仿真的方法在基色

收稿日期: 2011-10-09; 收到修改稿日期: 2012-02-09

基金项目: 国家自然科学基金(60972134)、浙江省自然科学基金(Y1100771)、上海科委科研项目(09220502700)和江苏省高校研究生科研创新计划(CX09B\_180Z)资助课题。

作者简介: 李金城(1980—), 男, 博士研究生, 主要从事色彩再现理论与技术等方面的研究。

E-mail: lijincheng@hdu.edu.cn

导师简介: 刘 真(1953—), 女, 教授, 博士生导师, 主要从事印刷光学工程、色彩再现理论与应用等方面的研究。

E-mail: lunaprint@163.com(通信联系人)

油墨库中筛选能够实现最大色域再现的基色色料,代表性的研究如文献[5]所述。第一种基色估计算法仅仅是方法反演,并不能直接用于基色筛选;第二种方法须预先确定基色个数。在实际工业应用中,业界主要通过增加在4色(CMYK)上增加新基色完成多色高保真复制。因此,研究在传统4色基础上增加其他基色的多色高保真复制具有更高的理论意义和实际应用价值。

本文提出的估计算法以传统4色色域为研究基础,通过比较传统4色和目标高保真色域,结合遥感科学中端元的提取算法,综合运用了色域分析、主成分分析和聚类分析等多种技术方法估计出基色的光谱反射率。

## 2 算法的提出

在遥感科学中,端元是指标准地物,例如植被、矿物等,遥感图像是端元光谱以不同丰度混合的结果。提取端元光谱是遥感影像分类、识别的前提和关键技术。在印刷复制中,基色油墨光谱就是混合光谱中的纯光谱,类似于遥感图像中的端元。成熟的端元提取算法分为两类,第一类假设图像中存在仅含每类地物的纯像元,通过各种算法可以从图像本身提取出合适的像元作为端元,该类算法比较成熟,例如纯净像元指数(PPI),迭代误差分析(IEA)等;第二类,当图像中不存在某类地物的纯像元时,从图像上提取出最接近于该地物光谱的像元光谱作为端元<sup>[6]</sup>。在印刷复制的图像中一定存在基色,因此可以借鉴第一类成熟的端元算法进行基色的光谱估计。高晓惠<sup>[7]</sup>提出了先利用基于空间特征的光谱分类算法对高光谱图像进行分类,从所有类别的标准光谱中提取纯光谱的端元提取算法,该算法的基本思路适用于基色的筛选。

综合考虑端元提取算法在基色光谱筛选方面的适用性以及印刷光谱分类方面的研究成果,提出的基色筛选算法如下:

1) 对传统4色和目标高保真色域进行分析,提取目标高保真色域中位于4色色域外的特征点集合 $\Phi$ ;

2) 对 $\Phi$ 中光谱反射率进行主成分分析,确定基色个数 $m$ ;

3) 以基色个数 $m$ 为分类数,对特征点集合的光谱反射率进行聚类分析,得到 $m$ 个光谱分类;

4) 计算 $\Phi$ 中光谱反射率的平均值,在每个光谱分类中找到距离平均光谱欧氏距离最远的光谱作为

基色光谱,筛选得到 $m$ 个基色光谱;

5) 在基色库中选择与 $m$ 个基色光谱相似度最大的光谱作为最终选择的基色光谱。

## 3 色域分析

现有基色色料估计方法主要采用对目标色域内所有特征色的光谱反射率进行主成分分析,获取能重建所有光谱反射率的基色光谱。研究表明,受到主成分个数的限制,其对大量光谱数据的预测效果偏低<sup>[8]</sup>,因此应设法减少计算的颜色数。多基色复制所要达到目的是复制出传统4色无法复制的色彩,即复制4色色域外的颜色。为了有效减少计算颜色数量,可以先比较目标高保真色域和4色色域,筛选出4色色域外的高保真特征色,只用这些特征色进行分析得到新基色的光谱。本文所指特征色是在色彩空间中对色彩转换起基础和控制作用的颜色,即能够以最少的色彩数目和相应算法就能够完整、准确地描述整个色彩空间的颜色<sup>[9]</sup>,具体来讲特征色就是标准色靶中的颜色。

为此,需要构建4色色域并找到色域外的特征色。目前通过离散的特征色点集(离散特征色的色彩属性可以从色彩标靶中获得)构建色域是色域描述的主要方法之一,色域边界可以描述为一个凸壳。色域及边界的描述有多种方法,其中利用Delaunay三角网格剖分,将离散点集进行四面体网格分割能够实现色域的快速确定。采用以下方法构建4色色域。

1) 选择所有的4色特征点的色度值(CIE  $L^* a^* b^*$ ,  $L^*$ 为明度值, $a^*$ 为红绿, $b^*$ 为黄绿,CIE为国际照明委员会)作为样本(数据从标准色靶中获得);

2) 利用Delaunay四面体网格将所有特征点进行分区,构建出若干个四面体。

4色色域外的点不能存在于已经构建的任何四面体中,因此只需要判断目标高保真色域特征点是否包含在某四面体内,就能找到所有4色色域外的特征色。

## 4 主成分分析方法

主成分分析是一种降维方法,目前已经广泛应用于印刷色彩光谱重建、基于光谱的色彩复制中,证明了该方法对于印刷色彩光谱估计的有效性<sup>[10,11]</sup>。 $P$ 个基色的线性组合可以近似描述整个特征色 $\Phi_\lambda$ :

$$\Phi_\lambda = \sum_{i=1}^P b_i e_{\lambda,i}, \quad (1)$$

式中  $e_{\lambda,i}$  为第  $i$  个特征向量,  $b_i$  为重建特征光谱的相应系数。主成分分析法的主要目的是降低变量的个数进而简化源数据, 因此一般很少直接应用所有  $P$  个主成分, 而是选取前  $m$  ( $m < P$ ) 个主成分来表示源数据。通常用主成分累计贡献率(CVC) 来表示前  $m$  个主成分的重要程度, 以确定所需要的主成分个数。前  $m$  个主成分累计贡献率的定义为

$$f_{\text{CVC}} = \frac{\sum_{i=1}^m \lambda_i}{\sum_{j=1}^p \lambda_j}, \quad (2)$$

式中  $\lambda_i$  代表第  $i$  个主成分的得分。

## 5 聚类分析方法

聚类分析法是一种常见的多元统计方法, 此方法依据集合中样本间的某种相似属性, 将原始数据分为若干子集。2004年, Mohammadi 等<sup>[12]</sup> 在色卡设计的研究中通过计算光谱数据间的相关距离, 进而对颜色集中的颜色进行分类。以任意两条光谱曲线为例  $R_r = [R_{r1}, R_{r2}, \dots, R_{rp}]$ ,  $R_s = [R_{s1}, R_{s2}, \dots, R_{sp}]$ , 令  $\bar{R}_r = \frac{1}{n} \sum_p R_{rp}$ ,  $\bar{R}_s = \frac{1}{n} \sum_p R_{sp}$ ,  $p$  为波长范围 400 ~ 700 nm, 则这两个光谱之间的相关距离为<sup>[14]</sup>

$$D(R_r, R_s) = 1 - \frac{(R_r - \bar{R}_r)(R_s - \bar{R}_s)}{[(R_r - \bar{R}_r)(R_r - \bar{R}_r)']^{1/2} [(R_s - \bar{R}_s)(R_s - \bar{R}_s)']^{1/2}}. \quad (3)$$

## 6 实验验证与分析

### 6.1 实验数据的获取

选择的 4 色色域特征点为在 ISO 12647 国际标准规定印刷条件下得到的(印刷承印物为 175 g 铜版纸, 加网线数为 175 线每英寸) 国际色彩联盟 (ICC) 特征文件中提取, 数据包含 1617 个特征色及其  $L^* a^* b^*$  数据。6 色高保真特征色 (CMYK) 数据为现有 Profile Maker 色彩管理软件生成的 6 色 ICC 特性文件“ArtPaper Test Hi-fi. icc”<sup>[13]</sup> 中提取, 6 色特征色共有 2000 个, 包含  $L^* a^* b^*$  值及 380 ~ 730 nm, 波长间隔为 10 nm 的光谱反射率。

### 6.2 4 色色域外特征色的筛选结果

选择 Matlab 函数 Delaunay 对 4 色色靶 1617 个颜色点构建四面体, 共得到 9683 个四面体, tsearchn 函数搜索 6 色特征点是否在 4 色色域内, 最终得到 454 个在 4 色色域外的特征点, 将这些点绘制在 CIE  $a^* b^*$  色相面上如下图所示, 454 个特征色的光谱反射率如图 2 所示。

### 6.3 主成分分析结果

将特征色的光谱反射率进行主成分分析, 并计算前  $p$  个主成分的累积贡献率。调用 Matlab 工具箱中的 princomp 函数进行主成分分析。对 454 个特征色对应的光谱数据进行主成分分析, 前两个主成分的累积贡献率达到 94.68%, 满足贡献率达到 90% 以上的原则, 可以看出特征色集合主要由两个主要基色构成, 因此本文的基色个数为 2。需要说明的是如果要复制所有特征色或者目标高保真色域, 仅仅选择两个基色是远远不够的, 由于本文的研

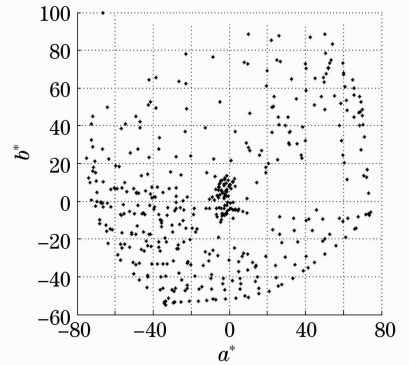


图 1 6 色色域中位于 4 色色域外所有色彩的  $a^* b^*$  坐标  
Fig. 1  $a^* b^*$  coordinate of all points out of the traditional CMYK gamut in 6c gamut

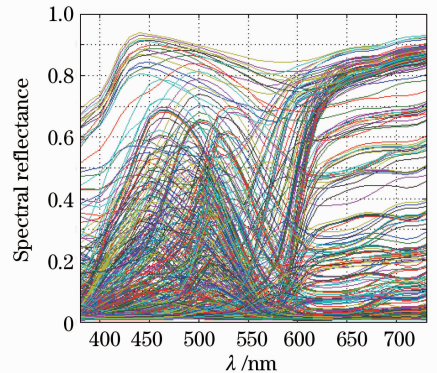


图 2 454 个特征色的光谱反射率

Fig. 2 Spectral reflectance of 454 color

究是在传统 4 色上选择新增的基色, 4 色仍然会参与最终色域的复制, 而 4 色色域外的特征色只需要两个主成分便能实现 94.68% 的重现, 因此选择基色个数为 2 是合理的。

### 6.4 聚类分析结果

$K$  均值聚类是预先指定分类数的聚类方法,在图像分割、分色中有较好表现,满足本研究中已经确定分类数的情况。使用修改了 Matlab 聚类函数 kmeans 中的距离表达法,使用(3)式作为距离度量,对特征色的光谱反射率数据进行聚类分析。指定聚

类个数为 2,分类结果如图 3 所示。接着,计算特征色集合的平均光谱,在得到的 2 个光谱分类中找到和平均光谱欧氏距离最大的光谱,得到的两个光谱如下图 4 所示。实际上,图 4 中的两个光谱就是目标高保真色域中的两个基色光谱,证明了算法提取基色光谱是准确的。

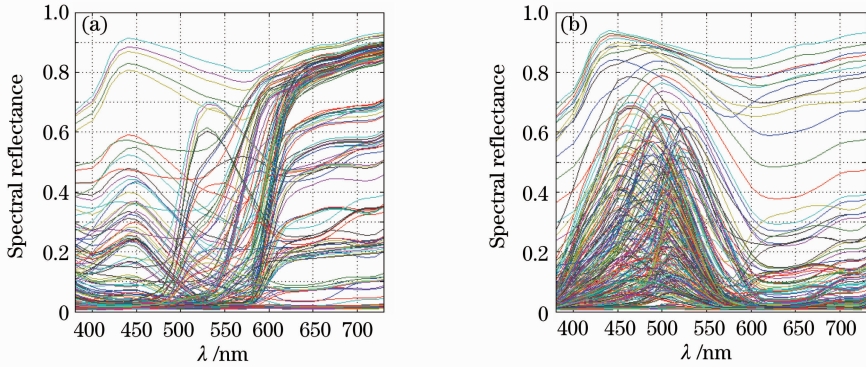


图 3 聚类分析后得到的两类光谱反射率曲线。(a)第一类;(b)第二类

Fig. 3 Spectral reflectance of the two clusters by cluster analysis. (a) Cluster I ; (b) cluster II

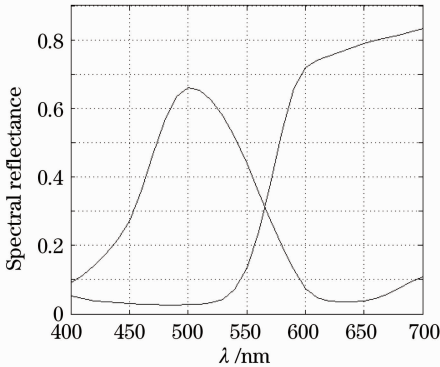


图 4 筛选得到的两个光谱反射率曲线

Fig. 4 Two selected spectral reflectance curves

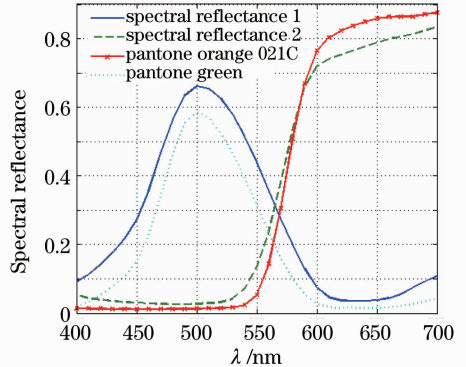


图 5 筛选得到的两个光谱反射率曲线与最终确定的基色光谱反射率的比较

Fig. 5 Comparison of the two selected spectral reflectance curves with the final basis colorant

### 6.5 实际基色确定

Pantone 专色是业界通用的颜色样本,在学术研究和工业应用方面都发挥着重要的作用。本文选择 Pantone 铜版纸专色样本作为基色库,以(3)式搜索基色库中与得到的两条光谱反射率曲线最接近的两个专色作为最终的基色。最终筛选得到的两个基色为 Pantone Orange021C 和 Pantone Green,如图 5 所示。

问题,考虑到实验条件的限制,采用  $n$  值( $n$  值取典型的 1.4)修正的光谱组介堡方程进行色域预测并使用 Matlab 进行计算机仿真的方式考察色域再现的情况。其中组介堡基色由 K/S 预测模型计算。将 CMYKOG 空间分为 YOMK, OMCK, MCGK, CGYK, CMYK5 个子空间,每个基色划分为 0~100%,间隔 10% 共 11 个等级,首先对每个子空间进行色域再现计算基色不同网点百分比组合得到的  $L^* a^* b^*$  值,最后将 5 个子空间进行合并,色域再现的结果如图 6 所示。图 6(a)为目标色域在仿真色域外特征点的  $L^* a^* b^*$  色域分布图。图 6(b)为目标色域在仿真色域外特征点(共 232 个)的  $a^* b^*$  色域平面图。

### 6.6 色域仿真

综上所述,最终确定参与目标高保真色域再现的基色个数为 6,分别为 C(Pantone Process C)、M(Pantone Process M)、Y(Pantone Process Y)、K(Pantone Process K)、O(Pantone Orange 021C)、G(Pantone Green)。由于实际高保真印刷涉及到加网方式、分色方法、油墨印刷适性和印刷色序等工艺



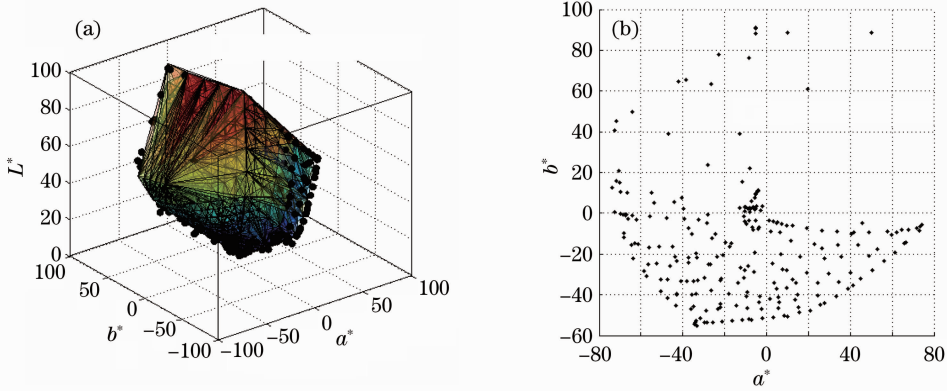


图6 (a) CMYKOG 色域与目标高保真色域的比较;(b)位于 CMYKOG 色域外的色彩

Fig.6 (a) Gamut comparison between CMYKOG and the target hi-fi gamut in  $L^*a^*b^*$  color space; (b) points out of CMYKOG gamut

## 6.7 结果分析

使用传统4色和从现有基色库中筛选出的基色进行色彩复制的结果可以看出,454个4色色域外的特征色中存在232个特征点仍然在色域外,从图6(b)可以看出新选择的基色O和G最大程度地弥补了4色色域在O和G区域的不足,复制出与目标6色接近的色域。

另外,仍然存在较少量饱和的黄绿、青绿、青色、品色,少量低饱和度品色和青绿色仍然在色域外,主要存在两方面的原因:1)因为本文选取的目标高保真6色使用了更饱和的C、M、Y色,使用传统4色和新基色并不能改善C、M、Y区域的复制。2)由于现有基色库的限制,确定的基色与筛选出的基色光谱只能接近,而不是一致。其中1)是主要原因。

## 7 结论

特别需要强调的是,运用本文方法筛选的基色光谱就是目标高保真复制中所使用的O和G光谱,证明了借鉴纯净端元提取的方法筛选基色光谱是完全可行的。在仿真中,虽然未完全实现目标6色的复制,主要原因是目标高保真色域并不使用传统4基色。由于能准确筛选出基色光谱,该方法不但可以应用于基于色域的基色筛选,还可以应用于图像的基色筛选。

## 参考文献

- 1 T. Kohler, R. S. Berns. Reducing metamerism and increasing gamut using five or more colored inks[C]. IS&T's 3rd Technical Symposium on Press, Proofing & Printing, 1993. 24~28
- 2 Di-Yuan Tzeng, Roy S. Berns. Spectral-based ink selection for multiple-ink printing I. Colorant estimation of original objects [C]. Proceedings of the 6<sup>th</sup> IS &T/S ID Color Imaging Conference, 1998. 1062111

- 3 Di-Yuan Tzeng, Roy S. Berns. Spectral-based ink selection for multiple-ink printing II. Optimal ink selection[C]. Proceedings of the 7<sup>th</sup> IS &T/ S ID Color Imaging Conference, 1999. 1822187
- 4 Xu Faqiang, Wan Xiaoxia, Zhu Yuanhong. Color component prediction based on rotated principal component analysis [J]. *Optics and Precision Engineering*, 2008, **16**(3): 518~523
- 许法强, 万晓霞, 朱元泓. 基于旋转主成分分析的颜色组分预测 [J]. *光学精密工程*, 2008, **16**(3): 518~523
- 5 J. Viggiano, W. J. Hoagland. Colorant selection for six-color lithographic printing[C]. Proceedings of the IST/SID 1998 Color Imaging Conference, 1998, 112~115
- 6 Li Shanshan. An endmember progressive extraction algorithm for hyperspectral remote sensing image [J]. *J. Remote Sensing*, 2009, **13**(2): 269~275
- 李姗姗. 高光谱遥感图像的端元递进提取算法 [J]. *遥感学报*, 2009, **13**(2): 269~275
- 7 Gao Xiaohui. Research on endmember extraction algorithm based on spectral classification [J]. *Spectroscopy and Spectral Analysis*, 2011, **31**(7): 1995~1998
- 高晓惠. 基于光谱分类的端元提取算法研究 [J]. *光谱学与光谱分析*, 2011, **31**(7): 1995~1998
- 8 Liu Qiang. Comparative study of metrics for spectral color grouping [J]. *China Printing And Packaging Study*, 2010, **11**(2): 104~107
- 刘强. 不同光谱颜色分组度量的比较研究 [J]. *中国印刷与包装研究*, 2010, **11**(2): 104~107
- 9 Wang Qiang. The Study on Of Color Management Mechanism of Spatial Information [D]. Wuhan: Wuhan University, 2005. 55~56
- 王强. 空间信息的色彩管理机制研究[D]. 武汉: 武汉大学, 2005. 55~56
- 10 M. J. Vrhel, R. Gershon, L. S. Iwan. Measurement and analysis of object reflectance spectra [J]. *COLOR Research and Application*, 1994, **19**(1): 4~9
- 11 B. Bastani, B. Funt, J. Dicarlo. Spectral reproduction: how many primaries are needed? [C]. Proceeding of NIP23, 23rd International Conference on Digital Printing Technologies and Digital Fabrication, Alaska, 2007, 410~413
- 12 M. Mohammadi, M. Nezamabadi. Spectral imaging target development based on hierarchical cluster analysis [C]. Proceedings of the 12th Color Imaging Conference, 2004, 59~64
- 13 Zhu Ming, Liu Zhen, Chen Guangxue. Research on six-color separation model based on subarea neugebauer equations[J]. *Acta Optica Sinica*, 2011, **31**(7): 0733001
- 朱明, 刘真, 陈广学. 基于分区组介堡方程的6色印刷分色模型研究 [J]. *光学学报*, 2011, **31**(7): 0733001