

3 种主流分色算法的误差分析

周倩, 刘真, 朱明

(南京林业大学江苏省制浆造纸科学与技术重点实验室, 南京 210037)

摘要: 在分析多项式回归、神经网络、修正的纽介堡方程这 3 种主流分色算法基本原理的基础上, 对这 3 种类型算法分别采用新的实现方法进行设计, 即: 多项式回归算法的建模, 通过多元回归分析求解出多项式的最小二乘估计值进行; 神经网络算法的建模, 采用基于径向基的 RBF 神经网络; 纽介堡方程的建模, 采用迭代法直接求解的方式。最后采用 Matlab 编程实现, 并实验比较了 3 种算法的分色精度、稳定性等性能。研究结果表明: 多项式回归算法稳定性和精度都最佳且平均色差小, RBF 算法色块建模速度快, 转换效果较好。最后分析了分色结果, 提出了算法改进方向。

关键词: 分色算法; Lab 色空间; CMYK 色空间; 误差分析

中图分类号: TS807; TS801.3 **文献标识码:** A **文章编号:** 1001-3563(2011)07-0107-05

Errors Analysis of Three Major Color Separation Algorithms

ZHOU Qian, LIU Zhen, ZHU Ming

(Jiangsu Provincial Key Lab. of Pulp and Paper Science and Technology, Nanjing Forestry University, Nanjing 210037, China)

Abstract: The fundamental of three major color separation algorithms was analyzed, of which were Neugebauer equations, polynomials, and neural networks. The algorithms were designed with new methods: polynomials were solved by multiple regression analysis; neural networks were modeled by RBF; Neugebauer equation by iteration method. They were programmed by Matlab and experiments were designed to compare the accuracy and stability of the three algorithms. The results showed that polynomial regression algorithm's stability and accuracy were the best and has the smallest average color difference; RBF algorithm was the fastest. Improved method was proposed by analyzing separation results.

Key words: color separation algorithm; Lab color space; CMYK color space; error analysis

分色技术是色彩复制技术实现的前提条件。在颜色理论上, 从 Lab 色空间向 CMYK 色空间转换的实质就是分色的过程^[1]。分色算法的优劣直接影响印刷输出的质量, 目前主流的分色算法包括修正的纽介堡方程、多项式回归、神经网络、多光谱、查找表等方法。文献[2]中多项式回归算法采用将非线性多项式转化为线性求解的方法, 该方法不能保证在整个色域的转换精度一致; 文献[3]采用 BP 神经网络模型建立分色模型, 该种网络模型训练时间较长, 泛化能力较弱; 文献[4]提出的纽介堡方程修正方法, 采用加密算法对纽介堡方程进行修正, 修正后出现分色后颜色并集的现象, 一定程度上影响了分色精度。笔者

针对这 3 种类型算法, 在分析各算法原理的基础上, 对这 3 种类型算法分别, 采用新的实现方法进行设计, 即: 多项式回归算法的建模通过多元回归分析求解出多项式的最小二乘估计值进行; 神经网络算法的建模, 采用训练速度和泛化能力较高的基于径向基的 RBF 神经网络; 纽介堡方程的建模采用迭代法直接求解的方式。最后采用 Matlab 编程实现, 并且设计实验比较 3 种算法的分色精度、稳定性等分色性能。

1 主流分色算法及建模

选取德国 FOGRA ECI2002 标准色靶测量所得

收稿日期: 2011-01-10

作者简介: 周倩(1986—), 女, 江苏江都人, 南京林业大学硕士生, 主攻印刷色彩管理。

的数据作为建模和检测数据。

1.1 多项式回归算法

多项式回归^[2]是研究一个应变量和一个或多个自变量之间关系的一种数学方法。回归算法的多项式分为线性和非线性,由于色空间的转换是非线性的,在建立分色模型时选择非线性回归方程。在使用多项式回归算法构建分色模型时,以 C,M,Y,K 的网点面积率为应变变量, L,a,b 值为自变量,根据三刺激值的可加性原理,对源色空间和目标色空间间建立多项式回归关系,见式(1)^[5]。理论上多项式的项数越多,精确度会越高,但精确度提高的幅度随着项数的增加逐渐变小,过高的项数可能引起震荡。一般在科学技术领域内,式(1)的二次多项式逼近精度已经足够。

$$\begin{aligned}
 C &= nb_0 + \sum_1^n Lb_1 + \sum_1^n ab_2 + \sum_1^n bb_3 + \\
 &\sum_1^n L^2b_4 + \sum_1^n a^2b_5 + \sum_1^n a^2b_6 + \sum_1^n Lab_7 + \\
 &\sum_1^n Lbb_8 + \sum_1^n abb_9 \\
 M &= nb_0 + \sum_1^n Lb_1 + \sum_1^n ab_2 + \sum_1^n bb_3 + \\
 &\sum_1^n L^2b_4 + \sum_1^n a^2b_5 + \sum_1^n b^2b_6 + \sum_1^n Lab_7 + \\
 &\sum_1^n Lbb_8 + \sum_1^n abb_9 \\
 Y &= nb_0 + \sum_1^n Lb_1 + \sum_1^n ab_2 + \sum_1^n bb_3 + \\
 &\sum_1^n L^2b_4 + \sum_1^n a^2b_5 + \sum_1^n b^2b_6 + \sum_1^n Lab_7 + \\
 &\sum_1^n Lbb_8 + \sum_1^n abb_9 \\
 K &= nb_0 + \sum_1^n Lb_1 + \sum_1^n ab_2 + \sum_1^n bb_3 + \\
 &\sum_1^n L^2b_4 + \sum_1^n a^2b_5 + \sum_1^n b^2b_6 + \sum_1^n Lab_7 + \\
 &\sum_1^n Lbb_8 + \sum_1^n abb_9
 \end{aligned}
 \quad (1)$$

以 C 通道为例,通过变量转换将非线性多项式转换为多元线性回归方程,见式(2),简化多项式的求解。

$$\hat{c} = nb_0 + z_1b_1 + z_2b_2 + z_3b_3 + z_4b_4 + z_5b_5 + z_6b_6 + z_7b_7 + z_8b_8 + z_9b_9 \quad (2)$$

其中, $\hat{c}$ 为 C 通道网点面积率的回归估计值。通过多元线性回归分析求出 $b_0, b_1, b_2 \dots b_9$ 的最小二乘估计值 $\beta_0, \beta_1, \beta_2 \dots \beta_9$,最终得到色度值向网点面积率值的数学转换模型,建立起源色空间到目标色空间的转换关系。

$$C = \beta_0 + \beta_1L + \beta_2a + \beta_3b + \beta_4La + \beta_5Lb + \beta_6ab + \beta_7L^2 + \beta_8a^2 + \beta_9b^2 \quad (3)$$

1.2 神经网络算法

径向基函数神经网络(RBF)是一种高效的前馈式神经网络,具有较强的非线性拟合能力,收敛速度快,拓扑结构紧凑。主要用于函数逼近、数据压缩等方面^[3,6],近年来也被应用于颜色空间的转换。RBF 神经网络结构模型见图 1。

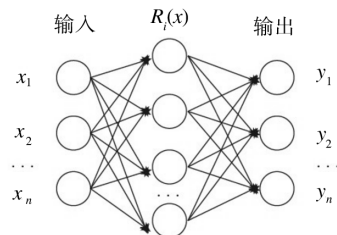


图 1 RBF 神经网络的结构

Fig. 1 Structure of RBF neural network

RBF 神经网络由 3 层组成,包括输入层、隐含层和输出层。在使用 BP 神经网络算法构建分色模型时,输入层和输出层为源颜色空间和目标颜色空间数据,通过样本数据建立的神经网络模型则可以用于色空间转换。RBF 神经网络的调用格式为^[7-8]:

$$\text{net} = \text{newrbe}(T, P, \text{Spread}); y = \text{sim}(\text{net}, Tts)$$

其中: T 为源数据矩阵; P 为目标数据矩阵; Tts 为检验模型精度的数据矩阵; Spread 为径向基函数的分布密度, Spread 值直接影响 RBF 网络的预测性能, Spread 值越大,仿真数据与测试数据的逼近程度越好,但网络的逼近误差随之增大。

选取 ECI2002 色靶中 1000 组数据作为建模样本,485 组作为检验模型精度的检测样本。建模样本和检测样本的色块根据各个分色网点面积率均匀选取。由于输出数据 $CMYK$ 值在 $[0, 1]$ 之间,而 Lab 值变化范围 $[-128, 127]$,远离该区间,且数据幅值大小不一,这样会影响训练效果,增加训练时间。在训练之前,需要归一化训练数据,将 Lab 值根据式(4)归一化到 $[0, 1]$,最后再根据式(5)对仿真数据 $CMYK$ 值进行反归一化处理^[9]。把归一化后的数据作为目标数据进行训练。

$$X'_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (4)$$

$$X = X'_i(X_{\max} - X_{\min}) + X_{\min} \quad (5)$$

其中: X_i 是需要归一化的数据; X'_i 是归一化后的数据; $X_{\max}$ 和 $X_{\min}$ 分别是数据集中的最大和最小数据。

Spread 值对 RBF 神经网络的预测性能影响很

大,借助均方误差函数 MSE 来确定 Spread 值。Spread 分别取 1,3,6,8,12 时的 MSE 值,见表 1。

表 1 网络训练均方差随 Spread 值的变化
Tab.1 Change of Spread values
with training mean square deviation

Spread	MSE
12	0.042 022
8	0.020 184
6	0.010 878
3	0.010 471
1	0.010 432

当 Spread 值为 6,3,1 时,仿真数据与测试数据的逼近效果较好,并且误差逼近 0.01,而 Spread 值越大函数越平滑,因此网络的 Spread 值选为 6,将 1000 组建模样本代入网络建立模型。

1.3 修正的纽介堡方程算法

纽介堡方程是根据 Grsassmann 颜色定律和印刷网点模型建立的印刷品呈色方程。纽介堡方程虽然原理清晰,但直接使用时由于网点扩大等因素,其分色精度较低,因此,国内外很多学者对方程在消除网点扩大方面进行了修正^[8]。修正纽介堡算法为:

$$\left. \begin{aligned} (X/X_0)^{1/n_x} &= \sum_{i=1}^{16} f_i U_i \\ (Y/Y_0)^{1/n_y} &= \sum_{i=1}^{16} f_i V_i \\ (Z/Z_0)^{1/n_z} &= \sum_{i=1}^{16} f_i W_i \end{aligned} \right\} \quad (6)$$

$$\Delta x_1 = \frac{(F_2 F_{32} - F_3 F_{22})(F_{13} F_{22} - F_{23} F_{12}) - (F_1 F_{22} - F_2 F_{12})(F_{23} F_{32} - F_{33} F_{22})}{(F_{11} F_{22} - F_{21} F_{12})(F_{23} F_{32} - F_{33} F_{22}) - (F_{21} F_{32} - F_{31} F_{22})(F_{13} F_{22} - F_{23} F_{12})}$$

$$\Delta x_2 = \frac{(F_2 F_{33} - F_3 F_{23})(F_{11} F_{23} - F_{21} F_{13}) - (F_1 F_{23} - F_2 F_{13})(F_{21} F_{33} - F_{31} F_{23})}{(F_{12} F_{23} - F_{22} F_{13})(F_{21} F_{33} - F_{31} F_{23}) - (F_{22} F_{33} - F_{32} F_{23})(F_{11} F_{23} - F_{21} F_{13})}$$

$$\Delta x_3 = \frac{(F_2 F_{31} - F_3 F_{21})(F_{12} F_{21} - F_{22} F_{11}) - (F_1 F_{21} - F_2 F_{11})(F_{22} F_{31} - F_{21} F_{32})}{(F_{13} F_{21} - F_{23} F_{11})(F_{22} F_{31} - F_{21} F_{32}) - (F_{23} F_{31} - F_{33} F_{21})(F_{12} F_{21} - F_{22} F_{11})} \quad (9)$$

这里设 x_1, x_2, x_3 的初始值 $x_1(0), x_2(0), x_3(0)$ 为 0.618,如果 F_i 函数第一次计算结果在 x_1, x_2, x_3 时的取值满足不等式(10),其中 ϵ 取任意极小正数,这里使 $\epsilon = 0.000\ 1$,那么此时的 x_1, x_2, x_3 值即为该采样点的分色值;如果不满足,那么使用式(9)中计算得到的 $\Delta x_1, \Delta x_2, \Delta x_3$ 来修正 x_1, x_2, x_3 的初始值,得到第一次迭代值 $x_1(1), x_2(1), x_3(1)$,如式(11)。

$$\sqrt{F_1(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k+1)}) + F_2(x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k+1)}) + F_3(\dots)} \leq \epsilon \quad (12)$$

此时的 $x_1(k+1), x_2(k+1), x_3(k+1)$ 就是方程

其中: X_0, Y_0, Z_0 为白场的三刺激值。 $U_i = (X_i/X_0)^{1/n_x}, V_i = (Y_i/Y_0)^{1/n_y}, W_i = (Z_i/Z_0)^{1/n_z}$, 分别为各个色元相对三刺激值的 $1/n_x, 1/n_y, 1/n_z$ 次方。

将纽介堡方程看作一个非线性方程组,根据一元函数的 Taylor 公式和一元非线性方程牛顿迭代法之间的关系,利用多元函数的 Taylor 公式推导出 3 元非线性方程组的牛顿迭代法,在此基础上利用 Matlab 编程完成方程组的求解。首先将纽介堡方程组进行变形,变形后的方程组形式为:

$$\begin{cases} F_1(x_1, x_2, x_3) = \sum_{i=1}^8 f_i X_i^{1/2.7} - X^{1/2.7} = 0 \\ F_2(x_1, x_2, x_3) = \sum_{i=1}^8 f_i Y_i^{1/2.65} - Y^{1/2.65} = 0 \\ F_3(x_1, x_2, x_3) = \sum_{i=1}^8 f_i Z_i^{1/2.5} - Z^{1/2.5} = 0 \end{cases} \quad (7)$$

其中: x_1, x_2, x_3 分别为 CMY 通道的网点面积率,即分色值。将上式按一阶泰勒公式展开,并简化为线性表示:

$$\begin{cases} F_1 + F_{11} \Delta x_1 + F_{12} \Delta x_2 + F_{13} \Delta x_3 = 0 \\ F_2 + F_{21} \Delta x_1 + F_{22} \Delta x_2 + F_{23} \Delta x_3 = 0 \\ F_3 + F_{31} \Delta x_1 + F_{32} \Delta x_2 + F_{33} \Delta x_3 = 0 \end{cases} \quad (8)$$

其中: F_1, F_2, F_3 为 F_i 函数第一次计算结果在 x_1, x_2, x_3 时的取值; $\Delta x_1, \Delta x_2, \Delta x_3$ 为第一次迭代求出的网点面积率值的修正量。以 $\Delta x_1, \Delta x_2, \Delta x_3$ 为未知量,通过求解线性方程组可得:

$$\sqrt{F_1^2 + F_2^2 + F_3^2} \leq \epsilon \quad (10)$$

$$\begin{aligned} x_1^{(1)} &= x_1^{(0)} + \Delta x_1, x_2^{(1)} = x_2^{(0)} + \Delta x_2, x_3^{(1)} = \\ &= x_3^{(0)} + \Delta x_3 \end{aligned} \quad (11)$$

设定停止迭代的条件为:第 $k+1$ 次迭代时函数 F_i 满足:

组的解,即分色结果。

2 误差分析

2.1 绝对误差

使用 3 种分色算法建模后,根据各个分色网点面积率均匀选取 ECI2002 色靶中 100 个色块作为检测样本,分别使用 3 种分色算法分色,将分色结果与标准值比较,计算其绝对误差,绘制各个分色的直方图,见图 2,直方图表明:神经网络和多项式回归 2 种分色算法绝对误差在 0.02 之内,CMYK 4 种分色较集中,神经网络算法 C 和 K 误差集中在 $[0,0.02]$ 之间,M 和 Y 的误差集中在 $[0,0.02]$ 和 $[0.04,0.1]$,相对 C 和 K 的分色效果要差;多项式回归算法中 M 的绝对误差在 $[0,0.02]$ 区间的占 16%,优于其他 3 个分色。

3 种分色算法 C,M,Y,K 网点面积率计算值与标准值之间的最小误差、最大误差、平均误差和标准差见表 2。

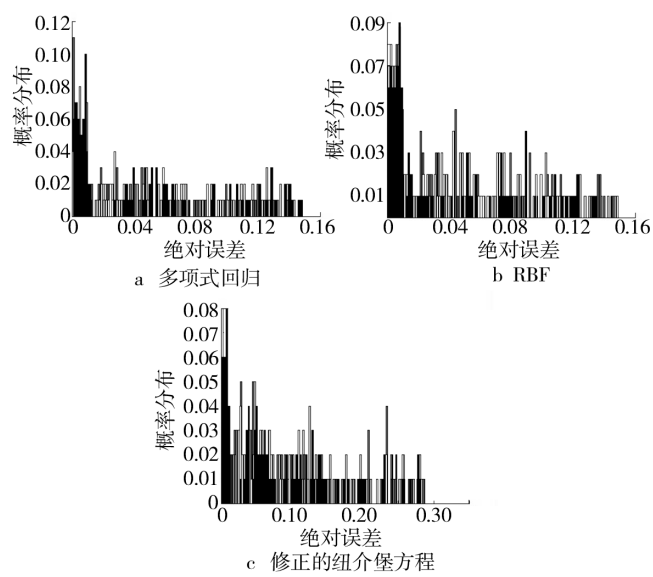


图 2 3 种算法的各分色绝对误差直方图
Fig. 2 Histogram of color separation absolute error of the three algorithms

表 2 3 种算法误差分析
Tab.2 Error analysis of the three algorithms

分色算法	平均误差 %				最小误差 %				最大误差 %			
	C	M	Y	K	C	M	Y	K	C	M	Y	K
多项式回归	4.6	4.3	4.5	4.8	0.7	1.0	0.2	1.9	14.8	14.8	14.7	14.2
RBF	9.2	4.6	4.5	4.0	1.5	0.4	0.2	1.4	14.9	14.5	14.0	14.9
修正组介堡方程	15.4	9.9	7.2	11.6	4.5	2.8	1.1	1.4	27.5	28.8	25.7	26.1

其中平均误差反映了分色算法的精度,RBF 和多项式回归算法的平均误差相近,都在精细印刷品网点增大值的范围内,修正的组介堡方程平均误差最大即精度最差,多项式回归算法的精度最高。

2.2 稳定性分析

标准差可以反映整个样本变量的分散程度。样本标准差小,说明样本变量的分布比较密集在平均数附近,否则,表明样本的分布比较离散。模型检测数据的标准差可以反映模型的稳定性,3 种分色模型分色 CMYK 网点面积率的标准差的比较见图 3。

综合分析图 3,可明显看出修正的组介堡方程稳定性最差,多项式回归的稳定性为 3 者中最佳;3 种分色模型中,CMYK 四色稳定性相比,Y 的稳定性最优,C 最不稳定。

2.3 色差分析

在 Photoshop 拾色器中将标准及使用分色算法得出的 C,M,Y,K 值转换为 L^*,a^*,b^* 值,计算两者

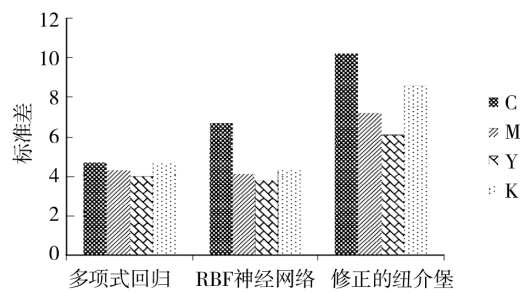


图 3 3 种分色算法的标准差
Fig. 3 Standard deviation of the three algorithms

色差,结果见表 3。

表 3 3 种分色算法的色差分布情况
Tab.3 Color difference distribution of the three algorithms

分色算法	0~3	3~6	6~10	10~15	15~22	>22	平均色差
多项式回归	43	27	17	6	4	3	3.58
RBF	37	34	15	8	4	2	4.26
修正组介堡方程	29	25	22	13	7	4	6.23

可以明显看出,RBF 色差小于 6 的色块数量为 3 者中最多,占 71%,多项式回归算法色差小于 3 的色块数量较多。修正的纽介堡方程在色差小于 10 的 3 个区间分布较均匀,整体色差较大。3 种分色算法都有色差大于 15 的超大色差色块,多项式回归算法和修正的纽介堡方程集中在暗调区域,与之相反,RBF 在亮调集中。综合比较 RBF 算法的转换效果最好。

3 结论

使用 MATLAB 工具对这 3 种分色算法的进行建模和实现,并且从稳定性、分色精度和色差这 3 个方面对算法进行了实验测试,测试结果表明:多项式回归算法的稳定性和精度最高,RBF 转换效果最佳,修正的纽介堡方程为 3 者中最差。这主要是因为:纽介堡方程是基于印刷网点呈色原理粗略构建而成,其本身存在精度不足的缺陷,即使经过指数修正后,相对于基于样本点测量的经验方法,前者在分色精度上还是略微不足。多项式回归算法稳定性和精度都较高,在暗调区域拟合效果较差,需要进行分段拟合进行优化;RBF 算法建模速度为 3 者中 fastest,结合绝对误差和色差分析,其稳定性和精度中等,但在亮调区域分色效果较差,通过适当增加亮调区域建模样本,可以提高精度和稳定性。

(上接第 92 页)

密度逐渐减小,完全叠印时密度最小。网点面积越大,因排列状态不同造成的色彩差异越大,尤其是网点完全重叠与网点并列产生的色差最大, ΔE_{ab} 值可达到 35.0,因此实际印刷工艺中要尽量避免网点完全叠印情况的发生。研究还发现,不同网点排列情况产生的色密度及光谱差异还与油墨颜色有关,当 C 与 M 网点混合时,因排列方式不同造成的色差总体来说小于其它 2 组墨色,但色差也远远超出了色差为 6 的可容许范围^[7]。小于 25% 的三原色网点,全叠与另外 2 种排列情况比较时色差较大,而并列状态与网点叠印面积小于 50% 的状态,其颜色差异小于 3,具有极佳的呈色一致性。至于大于 50% 网点的排列方式与呈色效果的差别,将在今后的研究中进一步探讨。

由以上分析及研究结论可以推论,网点不同排列状态中,点子并列时密度最大,色彩最饱满;网点完全重叠对色彩的削减最为严重,呈色均匀性及色彩饱和

参考文献:

- [1] 李瑞娟. RGB 到 CIEXYZ 颜色空间转换的研究[J]. 包装工程, 2009, 30(3): 79—81.
- [2] 周瑞. 彩色数码相机色彩空间转换模型的研究[D]. 西安: 西安理工大学, 2007.
- [3] 熊康鹏. 分色方法的建模研究与误差分析[D]. 西安: 西安理工大学, 2009.
- [4] SI Li-li, LIU Zhen, LIU Hao-xue. Researches on Neugebauer Equation Correction [C]. The 6th International Conference on Imaging Science and Hardcopy ICISH' 2008, 2008.
- [5] 孙静. 基于数码打样的 CMYK 与 $L^* a^* b^*$ 颜色空间转换方法的研究[D]. 西安: 西安理工大学, 2009.
- [6] SON Chang-hwan, PARK Hyung-min. Improved Color Separation Based on Dot-Visibility Modeling and Color Mixing Rule for Six-Color Printers[J]. Imaging Science and Technology, 2011, 55(1): 16
- [7] TOMINAGA Sho-ji. Color Conversion Using Neural Networks [C]. Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III, 1998.
- [8] 飞思科技产品研发中心. 神经网络理论与 MATLAB7 实现[M]. 北京: 电子工业出版社, 2005.
- [9] FUSELIER E J. Error and Stability Estimates for Surface-divergence Free RBF Interplants on the Sphere[J]. Mathematics of Computation 2009, 78: 2157—2186.

度最差;网点叠加面积越多越不利于色彩再现,因此在工艺上应尽量避免或减少网点交叉重叠。

参考文献:

- [1] 戴春. 彩色印刷的成色法[J]. 今日印刷, 2000(11): 11.
- [2] 王琪, 刘真. 数字加网中的网点保真技术研究[J]. 包装工程, 2010, 31(13): 101—102.
- [3] 刘浩学. 印刷色彩学[M]. 北京: 中国轻工业出版社, 2008.
- [4] 杜功顺. 印刷色彩学[M]. 北京: 印刷工业出版社, 2003.
- [5] LEE C, LEE E, et al. Color Space Conversion Via Gamut-based Color Samples of Printer[J]. Journal of Imaging Science and Technology, 2001, 45(5): 427—435.
- [6] 徐艳芳, 刘文耀. 彩色打印机色空间变换的色度密度方法[J]. 光学精密工程, 2004, 12(3): 265—269.
- [7] ISO 12647-2, Graphic Technology-Process Control for the Manufacture of Half-tone Color Separations, Proof and Production Prints[S].