Research Article

Evaluating the Performance of Polynomial Regression Method with Different Parameters during Color Characterization

Bangyong Sun, 1,2 Han Liu, 2 Shisheng Zhou, 1 and Wenli Li 1

1 School of Printing and Packing, Xi’an University of Technology, Xi’an 710048, China
2 School of Automation and Information Engineering, Xi’an University of Technology, Xi’an 710048, China

Correspondence should be addressed to Bangyong Sun; sunbangyong@xaut.edu.cn

Received 22 April 2014; Accepted 17 June 2014; Published 3 July 2014

Academic Editor: Yoshinori Hayafuji

The polynomial regression method is employed to calculate the relationship of device color space and CIE color space for color characterization, and the performance of different expressions with specific parameters is evaluated. Firstly, the polynomial equation for color conversion is established and the computation of polynomial coefficients is analyzed. And then different forms of polynomial equations are used to calculate the RGB and CMYK’s CIE color values, while the corresponding color errors are compared. At last, an optimal polynomial expression is obtained by analyzing several related parameters during color conversion, including polynomial numbers, the degree of polynomial terms, the selection of CIE visual spaces, and the linearization.

1. Introduction

As color electronics often have different imaging characteristics, such as the imaging mechanism, color space, apparatus capability, and material peculiarity [1], the color images always look different in some detail when they are output by different devices. For example, the same image displayed on monitor looks brighter and more colorful than that printed out on paper by printer. Even the same image displayed on two different monitors sometimes produces different visual effects. Thus, for color signal processing system with several color devices, in order to maintain the color consistence of color images, the precision of color signal transmission between different devices must be high enough. Now, for most of the color signal processing systems, the device-connection space is often used, such as the CIE color spaces CIEXYZ and CIELAB [2]. If the CIELAB space, for example, is chosen as the standard connection space, the color transmission process can be divided into two parts, device-to-CIELAB and CIELAB-to-device. Therefore, the color signal processing precision is highly depending on the color conversion algorithms between device colors and CIELAB colors.

There are many models which can be used for converting color signals, such as Neugebauer model, neural network, interpolation method, and polynomial regression. The regression model is widely used in color signal processing systems [3], since it can produce the high accuracy by using less sample data and also it can be used in both the device-to-CIELAB and the CIELAB-to-device directions. However, there are still some problems unresolved for this model during processing color signals; for example,

(1) the calculation precision of this model is highly dependent on the number and the degree of polynomial terms [4], so it is important to obtain the optimal polynomial expressions for specific signal processing process;

(2) the selection of device-connection space, such as CIEXYZ and CIELAB spaces, may have some influence on the signal processing precision [5, 6], so it still needs to be analyzed and tested for polynomial regression models;

(3) in some cases, the RGB signals are linearized before processing with CIE colors, but in other cases linearization is not added. Hence, for both the RGB and the CMYK signals, the effect of linearization processing should be analyzed and tested [7, 8], which may reveal whether or not it should be added for specific color devices and polynomials.
In this paper, these issues above are analyzed and tested in corresponding experiments. The different polynomials expressions, the different device-connection color spaces, and influence of linearization for signal processing are all tested on RGB and CMYK devices. At last, for the specific RGB and CMYK color signal processing systems, the optimal parameters are obtained with detailed analysis.

2. Polynomial Regression Model for Color Signal Processing

Polynomial regression is a form of linear regression in which the relationship between the independent variable $x$ and the dependent variable $y$ is modeled as an $n$th degree polynomial. Meanwhile, polynomial regression fits a nonlinear relationship between the value of $x$ and the corresponding conditional mean of $y$ values denoted by $E(y \mid x)$ and has been used to describe nonlinear phenomena, such as the growth rate of tissues [9], the distribution of carbon isotopes in lake sediments [10], and the progression of disease epidemics [11].

Within the polynomial regression model, if the dependent variable $y$ and multiple independent variables $x_1, x_2, \ldots, x_M$ have the linear relationship and there are $N$ groups of sample data:

$$(x_{i1}, x_{i2}, \ldots, x_{iM}, y_i), \quad t = 1, 2, \ldots, N,$$  \hspace{1cm} (1)

then the relationship between them can be described as [12]

$$
\begin{align*}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_M x_{1M} + \epsilon_1, \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_M x_{2M} + \epsilon_2, \\
&\vdots \\
y_N &= \beta_0 + \beta_1 x_{N1} + \beta_2 x_{N2} + \cdots + \beta_M x_{NM} + \epsilon_N,
\end{align*}
$$  \hspace{1cm} (2)

where $\beta_0, \beta_1, \ldots, \beta_M$ are the coefficients to be determined and $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ are independent random variables.

The system of expressions above can be represented using the matrix

$$Y = X\beta + \varepsilon,$$  \hspace{1cm} (3)

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1M} \\ 1 & x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_M \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix}.$$  \hspace{1cm} (4)

If $b_0, b_1, \ldots, b_M$ are the estimated values by least squares methods for parameter $\beta$, then the regression equation is

$$\hat{y} = b_0 + b_1 x_1 + \cdots + b_M x_M.$$  \hspace{1cm} (5)

From the principle of least squares [13–15], the coefficients of $b_0, b_1, \ldots, b_M$ should obtain the minimal residuals square sum for all the measured value $y_i$ and regression value $\hat{y}_i$:

$$Q = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2.$$  \hspace{1cm} (6)

By using least squares method, the coefficients $b$ can be resolved as follows:

$$b = (X^T X)^{-1} X^T Y.$$  \hspace{1cm} (7)

In addition, the polynomial regression method can also be used to describe nonlinear problems, in which the dependent variable $y$ is modeled as an $n$th degree polynomial of independent variables, so this model can be rightly used in color signal processing systems. Taking the RGB monitor as an example, with the CIEXYZ device-connection space, the relationship between RGB and CIEXYZ can be expressed as

$$X = \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \beta_{ij} R^i G^j B^k,$$

$$Y = \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \beta_{ij} R^i G^j B^k,$$

$$Z = \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \beta_{ij} R^i G^j B^k,$$

where $\beta$ is the polynomial coefficients, $n$ is the degree of polynomial, and $i + j + k \leq n$, and the expression above can also be represented using matrix

$$[X \ Y \ Z]^T = B_{3\times L} \times \rho_L,$$  \hspace{1cm} (9)

where the $B_{3\times L}$ is the coefficients matrix and $\rho_L$ is the matrix of polynoms, while $L$ represents the number of polynomials.

Thus, when $n = 1, L = 4$, the first-degree polynomial matrix $\rho_4$ is shown as follows:

$$\rho_4 = \begin{bmatrix} R & G & B & 1 \end{bmatrix}^T,$$  \hspace{1cm} (10)

when $n = 2, L = 10$, the second-degree polynomial matrix $\rho_{10}$ should be

$$\rho_{10} = \begin{bmatrix} R^2 & R G & R B & G^2 & B^2 & \rho_4 \end{bmatrix}^T,$$  \hspace{1cm} (11)

when $n = 3, L = 20$, the third-degree polynomial matrix $\rho_{20}$ is

$$\rho_{20} = \begin{bmatrix} R^3 & R^2 G & R^2 B & R G^2 & R B^2 & G^3 & B^3 & \rho_{10} \end{bmatrix}^T,$$  \hspace{1cm} (12)

and when $n = 4, L = 35$, the fourth-degree polynomial matrix $\rho_{35}$ is

$$\rho_{35} = \begin{bmatrix} R^4 G^4 & R^4 B^4 & R^3 G^2 B^2 & R^2 G^3 B^3 & R G^4 B^2 & G^4 B^3 & B^4 \rho_{20} \end{bmatrix}^T.$$  \hspace{1cm} (13)
There are also some other forms of polynomials used in color signal processing, and the polynomial matrices are shown as follows:

\[
\begin{align*}
\rho_3 &= [R \ G \ B]^T, \\
\rho_6 &= [RG \ RB \ GB \ \rho_3]^T, \\
\rho_8 &= [RGB \ \rho_6 \ 1]^T, \\
\rho_{14} &= [R^3 \ G^3 \ B^3 \ RGB \ \rho_{10}]^T.
\end{align*}
\]

When the coefficients \( n \) and \( L \) are defined in color signal processing, with the sample color data which consist of device colors and CIELAB colors, the coefficient matrix \( B_{3\times L} \) can be resolved using least square method.

3. Study on the Key Parameters during Color Signal Processing

To determine the key parameters within the polynomial regression model, a color signal processing system with several color devices is introduced in the experiment. As the additive primary color is \( RGD \) and subtractive primary color is CMYK, an \( RGD \) monitor and CMYK printer are chosen as the typical testing color devices. Within the color signal processing system, the device-connection space is CIEXYZ or CIELAB, so the color processing is mainly based on four color spaces.

For the purpose of obtaining the relationship between the device colors and the device-connection colors, the training sample data should be gathered in advance. For the IBM monitor, the three primary channels Red, Green, and Blue are all divided evenly into 9 parts, and each value of \( R, G, \) and \( B \) colors ranges within \([0 \ 32 \ 64 \ 96 \ 128 \ 160 \ 192 \ 224 \ 255]\). When all these \( 9^3 = 729 \) patches are displayed on monitor, the corresponding CIEXYZ and CIELAB colors are measured with Spectrophotometer X-Rite528, and all these CMY and corresponding CIE colors form the training sample data. To verify the accuracy of polynomial regression model, the testing sample data should also be collected. Similar to the training sample data, the testing data consists of \( 7^3 = 349 \) color patches with the single channel ranging within \([16 \ 48 \ 80 \ 112 \ 144 \ 176 \ 208 \ 240]\).

For the CMYK Epson 9880 printer, the single channel is divided into 11 parts with the interval 10, so every color channel ranges within \([0 \ 10 \ 20 \ 30 \ 40 \ 50 \ 60 \ 70 \ 80 \ 90 \ 100]\). Because the subtractive primary colors are Cyan, Magenta, and Yellow (CMY) and the color of Black can be seen as the replacement of a certain amount of CMY, in experiment the device color CMYK is treated as CMY. Thus, when all the \( 11^3 = 1331 \) CMY color patches are printed out, the corresponding CIEXYZ and CIELAB colors are measured with Spectrophotometer X-Rite 528, and all these CMY and corresponding CIE colors form the training sample data. In addition, the testing data consists of \( 6^3 = 216 \) patches with the single channel ranging within \([5 \ 25 \ 45 \ 65 \ 85 \ 95]\).

In general, the regression errors for the training sample reduce data as the number of polynomial terms or the degree of polynomials increases. However, for the color data of entire range, the regression errors will increase when the number of polynomial terms exceeds a certain value. Therefore, it is highly important and necessary to find the optimal polynomial expressions producing least errors, especially determined by the number of polynomial terms or the degree of polynomials. In experiment, to evaluate the different polynomial expressions employed in color signal processing, the error computation is defined by the color difference CIE76 formula [16]:

\[
\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2},
\]

where \((L_1^*, a_1^*, b_1^*)\) is the regression color and \((L_2^*, a_2^*, b_2^*)\) is the measured color.

3.1. Number and the Degree of Polynomial Terms. To find the appropriate polynomial expressions, the \( RGD \) monitor is tested and the signal processing errors with different polynomial expressions are compared. Firstly, the training sample data is used to obtain the polynomial coefficients between \( RGD \) and CIELAB signals; then for all the \( RGD \) colors, the CIELAB values can be simulated by using the obtained coefficients; secondly, in order to analyze the regression precision, the measured CIELAB color values from the training data are used to compute the errors of different polynomial expressions; at last the errors are represented as color differences between measured and simulated CIELAB values as below.

From the above result, it can be seen that the regression precision obviously improves as the number of polynomial terms increases, but when the number of terms reaches a certain value the mean error becomes small enough. For example, for the first-degree polynomial \( \rho_1 \), the average error is \( 10.9106 \Delta E \) which exceeds the reproduction error threshold [17, 18], and the maximal error is \( 37.5866 \Delta E \) which is a visually unacceptable error. Additionally its standard deviation and variance are \( 5.4642 \Delta E \) and \( 29.8576 \Delta E \), respectively, which indicate that the distribution of errors is unsatisfactory.

In general, the regression precision can be evaluated mainly from the average error for different polynomials shown in Figure 1. The regression errors for polynomials \( \rho_4, \rho_6, \rho_8, \rho_{14}, \rho_{20}, \) and \( \rho_{10} \) are all exceeding \( 5 \Delta E \) units, while for the other polynomials \( \rho_{14}, \rho_{20}, \) and \( \rho_{10} \) their color differences are all acceptable. The figure shows that polynomial \( \rho_{20} \) is the most accurate, but its precision is very close to polynomial \( \rho_{20} \). In addition, too many terms of the polynomial may increase the difficulty of coefficients-solving process [14], so the polynomial \( \rho_{20} \) should be most suitable for \( RGD \) color signal processing.

Using the regression coefficients from the training data, the relationship between \( RGD \) and CIELAB signals can be described. To verify the precision of different polynomials for the whole range of \( RGD \) signals, the testing sample data should be used. For the testing data which is outside the range of training data, the \( RGD \)'s corresponding simulation CIELAB values can be computed using the relationship...
polynomials $\rho_2, \rho_10, \rho_14, \rho_{20}$, and $\rho_{35}$ are only tested for CMYK signals. With the obtained polynomial coefficients solved with training data, for the 216 testing patches of testing sample data, the color differences are shown in Table 3.

It can be seen that, for the CMYK devices, most of the polynomials perform well with the average error below 3\(\Delta E\). This is mainly because the color-rendering properties of CMYK printers surpass those of the RGB monitors, such as the regularity of color gamut and color consistence [1, 19]. On the whole the fourth-degree polynomial $\rho_{35}$ with the largest number of terms has the smallest color difference and its error distribution is also ideal. The performance of the third-degree polynomial $\rho_{20}$ is close to the $\rho_{35}$, which indicates that the preferred polynomials for CMYK and CIELAB signal processing should be the $\rho_{35}$ or $\rho_{20}$.

### 3.2. Selection of CIE Color Spaces during Color Signal Processing

During the signal processing for different color devices, the use of two CIE color spaces, CIEXYZ and CIELAB spaces, often brings in different color errors. Hence, it is important to find the appropriate CIE color space for the specified color devices and polynomials.

In the experiment, the CIEXYZ and CIELAB color spaces are tested, respectively, on RGB monitors and on CMYK printers to compare their color errors. For the IBM monitor, when the CIEXYZ is selected as device-connection space, the errors of different polynomials are shown in Table 4. Corresponding to the color errors listed in Table 2 where the CIELAB is the device-connection space, the mean color differences corresponding to these two CIE spaces are compared in Figure 2.

As the figure shows, for the RGB monitors, the polynomials perform somewhat differently with the CIEXYZ and CIELAB color spaces. For the low-degree polynomials, the color error of signal processing with CIEXYZ space is greater than with CIELAB space, and along the increasing of polynomial terms, the influence of device-connection color space becomes very little. So when the polynomials of less

obtained above, and the errors can also be calculated by comparing the measured values. For the different polynomials and the testing data, the errors are shown below.

From Table 2, for the RGB and CIELAB color signals, the signal processing precision of different polynomials for testing data is similar to the training data shown in Table 1, and the acceptable forms of polynomials are $\rho_{14}, \rho_{20}$, and $\rho_{35}$, respectively.

For the purpose of testing the different polynomials' performance for CMYK signals, the training sample data of EPSON9880 printer are used to obtain the relationship between CMYK and CIELAB. Because the errors are too large for polynomials with few terms such as $\rho_4$ and $\rho_6$,
than 10 terms are used in RGB signal processing, the CIELAB color space is recommended as the device-connection space, while for the case when the polynomial terms are between 10 and 20, the CIEXYZ space is suggested.

To test the influence of CIE color space for CMYK devices, the errors of CIEXYZ and CIELAB color spaces are also compared on Epson9889 printer. The CMYK signal processing errors with CIELAB space have been listed in Table 3; for the polynomials of 8 to 35 terms, the color errors with CIEXYZ space are recorded in Table 5.

It can be seen that, for the second-degree polynomials $\rho_{10}$ and $\rho_{14}$, third-degree polynomial $\rho_{20}$, and fourth-degree polynomial $\rho_{35}$, all the color errors are acceptable. Taking account of the precision and computing efficiency, the third-degree polynomial $\rho_{20}$ is the most suitable model for CMYK devices using CIEXYZ color space. In addition, similar to the comparison of RGB signal processing with different device-connection spaces, the CMYK signal processing using CIELAB and CIEXYZ spaces is compared in Figure 3.

It can be seen that, within the CMYK signal processing based on polynomial regression method, the precision of color conversion using CIELAB color space is higher than that of using CIEXYZ space for a majority of the second, third, and fourth degree of polynomials. Therefore, in CMYK signal processing, the CIELAB color space is preferred as the device-connection space for the polynomial regression model.

### Table 5: The color differences of CMYK printer for different polynomials using CIEXYZ space.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Mean $\Delta E$</th>
<th>Min $\Delta E$</th>
<th>Max $\Delta E$</th>
<th>Std. $\Delta E$/var. $\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>11.9339</td>
<td>0.9928</td>
<td>37.2081</td>
<td>7.6683/58.8035</td>
</tr>
<tr>
<td>10</td>
<td>4.6879</td>
<td>0.5701</td>
<td>19.8019</td>
<td>3.5922/12.9040</td>
</tr>
<tr>
<td>14</td>
<td>4.3475</td>
<td>0.5852</td>
<td>20.1294</td>
<td>3.5023/12.2659</td>
</tr>
<tr>
<td>20</td>
<td>1.5809</td>
<td>0.3099</td>
<td>4.0116</td>
<td>0.6906/0.4769</td>
</tr>
<tr>
<td>35</td>
<td>1.2455</td>
<td>0.2666</td>
<td>3.0209</td>
<td>0.4968/0.2468</td>
</tr>
</tbody>
</table>

### Table 6: The RGB color difference of different polynomials with linearization applied.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>Mean $\Delta E$</th>
<th>Min $\Delta E$</th>
<th>Max $\Delta E$</th>
<th>Std. $\Delta E$/var. $\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8.0337</td>
<td>0.8814</td>
<td>18.4480</td>
<td>3.0490/9.2965</td>
</tr>
<tr>
<td>10</td>
<td>4.5930</td>
<td>0.4682</td>
<td>17.8630</td>
<td>2.6520/7.0334</td>
</tr>
<tr>
<td>14</td>
<td>3.9435</td>
<td>0.3297</td>
<td>18.0798</td>
<td>2.6400/6.9698</td>
</tr>
<tr>
<td>20</td>
<td>2.7392</td>
<td>0.4485</td>
<td>14.3771</td>
<td>1.8530/3.4335</td>
</tr>
<tr>
<td>35</td>
<td>2.3601</td>
<td>0.2703</td>
<td>9.6480</td>
<td>1.3645/1.8619</td>
</tr>
</tbody>
</table>

3.3. Linearization during Color Signal Processing. For the RGB devices, the linearization is often applied to calibrate the color device’s gray balance [20], in which the RGB signals are firstly converted into lightness signals before color conversion process. In the experiment, the linearization is described as follows:

$$ R_l = f_R(R), $$
$$ G_l = f_G(G), $$
$$ B_l = f_B(B), $$

where $R_l$, $G_l$, $B_l$, respectively, are the lightness signals, respectively, and $f_R$, $f_G$, $f_B$ are the linearization functions which are described as follows:

$$ f_C(C) = \alpha_0 + \alpha_1 C + \alpha_2 C^2 + \alpha_3 C^3, $$

where $C$ stands for one of the colors $R$, $G$, or $B$ and $\alpha_i$ $(i = 0, 1, 2, 3)$ are the coefficients.

Within the RGB color signals processing, the device colors RGB are firstly linearized into the lightness signals $(RGB)_l$, then the relationship between $(RGB)_l$ and CIELAB is obtained by polynomial regression model, and at last the color error is calculated by using the measured colors within the testing data. For the IBM monitor in the experiment, the color differences of different linearized polynomials are shown in Table 6.
4. Conclusions

In this paper, the polynomial regression model is used for RGB and CMYK color signal processing. For the purpose of improving color signal processing precision, the item number and degree of the polynomials are tested. By comparing the color errors within the color signal processing, the appropriate polynomial expressions for RGB and CMYK color devices are obtained. In addition, the parameters of device-connection color space and linearization are tested. In general, the RGB and CMYK color signal processing by employing the polynomial regression method can be concluded as follows.

(1) During the RGB and CMYK color signal processing with polynomials regression, it is advised to use the third-degree polynomials $\rho_{20}$ or fourth-degree polynomials $\rho_{35}$. Taking into account the coefficient solving process and the color errors in experiment, the third-degree $\rho_{20}$ is the most appropriate model.

(2) When the CIE color space is used as the device-connection space, for the RGB devices, the signal processing precision is higher with CIEXYZ space than with CIELAB space for polynomials including 10 to 20 terms, while in other cases the CIELAB is more precise. For CMYK devices, in most cases the CIELAB color space performs better than the CIEXYZ space.

(3) When the linearization is added in the color signal processing, the precision improves somewhat for parts of polynomials for the RGB devices, while for the CMYK devices, the addition of linearization reduces the signal processing accuracy instead, so the linearization is not recommended to be used within the CMYK signal processing.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
Acknowledgments

This work is supported by the National Science Foundation of China (no. 61174101), Doctor Foundation of Xi’an University of Technology (no. 104-211302), and "13115" Creative Foundation of Science and Technology (no. 2009GDGC-06), Shaanxi province of China.

References
