APPLICATION OF SPECTRUM EXPRESSION AND INDEPENDENT COMPONENT ANALYSIS IN COMPOSITION ANALYSIS OF MIXED PIGMENT

1GONGMING WANG, 2ZHIYONH LIU
1 Department of Information Art and Design, Academy of Art and Design, Tsinghua University, Beijing, China
2 Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
E-mail: 1 gongmingwang@126.com, 2 zyliu@ict.ac.cn

ABSTRACT

Reflectance spectroscopy (RS) is suitable for the composition analysis of mixed pigment, because the corresponding basic pigment can be determined by calculating the spectrum similarity. However, the spectrum similarity between known pigment and mixed pigment may be not the one between known pigment and the most similar basic pigment, and the known pigment with the largest spectrum similarity may not be the highest proportion of mixed pigment, due to the possible relationship among the basic pigments. Thus, it is easy to affect the accuracy of composition analysis. In this study, a composition analysis method of mixed pigment based on spectrum expression and independent component analysis (ICA) is proposed. First of all, the spectral information of mixed pigment is obtained with spectrometer, and is expressed as the mixed signal. After that, the independent signal (spectrum of basic pigment) is generated with ICA. Then, the type of basic pigment is determined by calculating the spectrum similarity between basic pigment and known pigment. Finally, the percentage of basic pigments is obtained by solving Kubelka-Munk equation system. To validate this method, the 4, 5, 6 simulated spectrums were mixed from four spectrums of Munsell color card, and the experiments were carried out in two aspects. On the one hand, the composition analysis of three groups of original simulated spectrums was executed. On the other hand, each group was mixed with one disturbance spectrum in order to test the anti-interference performance of this method. The separated SK curve is very similar to the original SK curve. The average similarity is 96.29%, and the maximum one can reach to 98.37%. The calculated percentage of basic pigment closes to the original one. It can be seen that this method is suitable for composition analysis of mixed pigment.

Keywords: Color, Composition Analysis, Spectrum, Similarity, Independent Component Analysis, Munsell

1. INTRODUCTION

The composition analysis of mixed pigment can determine the type and percentage of basic pigment that constitutes the mixed pigment, which is widely used for criminal identification [1], cultural relic identification [2], synthetic dye [3] and other fields. Reflectance spectroscopy (RS) [4] is suitable for composition analysis of mixed pigment. First, the type of basic pigment is determined by comparing the spectrum similarities between known pigments and mixed pigments [5]. Second, the percentage is calculated by solving Kubelka-Munk equation system. This method is simple and efficient, and has a certain application value. However, the spectrum similarity between known pigment and mixed pigment may be not the one between known pigment and the most similar basic pigment, and the known pigment with the largest spectrum similarity may not be the highest proportion of mixed pigment, due to the possible relationship among the basic pigments. Thus, it is easy to affect the accuracy of composition analysis. In addition, the experimental result is only effective for several kinds of pigments [6], does not have universal applicability.

It can be seen that the accuracy of type judgment of basic pigment is the main factor that affects the accuracy of RS. According to the Kubelka-Munk theory, the spectrum of mixed pigment is the linear combination of spectrums of basic pigments. Thus, the calculated spectrum similarity is the synthesis of spectrum similarities between known pigment and all basic pigments, not the one between known pigment and a basic pigment. In this case, even if the known pigment is one of the basic pigments, the
calculated spectrum similarity doesn't reach to 1. In addition, the dimension of spectral curve is often hundreds of thousands, and has the problem of "dimension disaster"[7]. In essence, all known pigments have no homologous with the mixed pigment. Thus, the difference of all spectrum similarities is a little [8], and it isn't convenient to determine the type of basic pigment accurately. The above disadvantages in similarity calculation are the root cause of the inaccuracy of type of basic pigment.

The process of mixing basic pigments can be regarded as the linear combination of random signals by taking spectrum as random signal. Independent component analysis (ICA) [9] is an effective tool for separation of linear mixed signal, and the independent signal can be deduced from the mixed signal. In this study, a composition analysis method of mixed pigment based on spectrum expression and ICA is proposed. First, the color of mixed pigment is expressed as reflection spectrum, in order to solve the "metamerism" problem. Second, spectrum vector is regarded as the mixed signal, and the independent signal can be deduced with ICA. Third, the independent signal is taken as the spectrum of basic pigment, and the type of basic pigment is determined by comparing the spectrum similarities between basic pigments and known pigments. Finally, the percentage of basic pigment is obtained by solving Kubelka-Munk equation system. To validate this method, the simulated spectrum was generated from Munsell color system, and the experiments of composition analysis of mixed pigment were carried out under the circumstance of normality and disturbance. The experiment results show that this method can accurately get the type and percentage of basic pigment, and is suitable for the composition analysis of mixed pigment.

The remainder of the paper is organized as follows: the advantage of spectrum expression, Kubelka-Munk theory and ICA are introduced in Section 2; the proposed composition analysis method of mixed pigment is described in Section 3; the composition analysis experiments are carried out in Section 4; and the conclusion and future work are shown in Section 5.

2. RELATED WORKS

2.1 Spectrum Expression of Mixed Pigment

Spectrum is the essential factor to determine surface color of object, and it has no difference along with the change of environment and observer. Thus, it can ensure the accuracy of color quantization and analysis. The Kubelka-Munk theory is useful for the composition analysis of mixed pigment, because it reveals the relationship between mixed pigment spectrum and basic pigment spectrum.

2.1.1 Spectrum expression of pigment

The tristimulus values representation is widely used for color representation in color engineering. Its theoretical basis is the three-dimensional (3D) color theory by Grassmann who is the founder of linear algebra in 1853[10]. That is to say, three variables are sufficient to define any color. Its physiological basis is the perception-response difference between three cones of human eye to color component [11]. The variable \((X, Y, Z)\) is used to describe color [12]. And the component \(X\), \(Y\) and \(Z\) are used to describe the perception intensities of three cones. The calculate method is shown in equation 1.

\[
\begin{bmatrix}
X \\
y \\
Z
\end{bmatrix}
= k \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} S_n(\lambda) \cdot \rho(\lambda) \cdot \tilde{x}(\lambda) d\lambda
\]

where \(\lambda_{\text{min}}\) and \(\lambda_{\text{max}}\) are the minimum and maximum spectrums of visible light, \(S_n(\lambda)\) is the spectral power distribution of light source, \(\rho(\lambda)\) is the spectral reflectance, \(\tilde{x}(\lambda)\), \(\tilde{y}(\lambda)\) and \(\tilde{z}(\lambda)\) are the standard colorimetric tristimulus values from observer with normal color recognition. According to the tristimulus values theory, all kinds of 3D color representations are proposed, such as XYZ, LAB, CMY and RGB.

The "metamerism" problem exists in the tristimulus values representation. That is to say, the tristimulus values of two different colors may be equal after integrated by equation 1. Thus, they may be seen as the same color. And the accuracy of copy and analysis of artwork would be affected by color aberration from this phenomenon. Generally speaking, the root reason is that the integral operation result in formula 1 masks the spectral reflectance difference between different colors. In addition, because the resolution of cones is limited, only seven colors in the different intervals of visible light can be distinguished. Therefore, the accuracies of \(\tilde{x}(\lambda)\), \(\tilde{y}(\lambda)\) and \(\tilde{z}(\lambda)\) in formula 1 are affected.

It is reported that the surface color of object is determined by reflectance spectrum on the surface of object. The reflection spectrum is the inherent physical property and independent on the environment and observer [13]. Thus, it can describe the color characteristics of the pigment.
itself, and avoid the uncertainty of type judgment of pigment derived from "metamerism" problem.

From the formula 1, it can be seen that the type judgment of pigment is mainly determined by the following three factors: pigment, environment and observer. The change of any factors would influence the result of tristimulus value representation [14], which affects the accuracy of color quantization and analysis. However, spectrum expression does not have this problem.

To sum up the above statements, the accuracy of color representation is the premise of composition analysis of mixed pigment, and the accuracy of spectrum expression is better than the one of tristimulus values representation. Thus, spectrum expression is suitable for the composition analysis of mixed pigment.

2.1.2 Kubelka-Munk theory

When light shines on the surface of opaque medium, besides a few is reflected, most of light enters into the medium, and the color is generated by absorption and scattering. The relationship between spectral reflectance and absorption/scattering is deduced by Kubelka and Munk [15].

![Figure 1: Incident and reflected light in any position](image)

**Figure 1: Incident and reflected light in any position**

![Figure 2: Light reflection effect from infinite lamella](image)

**Figure 2: Light reflection effect from infinite lamella**

Vertical downward is the positive direction, and incident light is along with vertical direction. The bottom of medium is the origin of coordinates x = 0, and the value of x is gradually increasing from bottom to top. In figure 1, the incident and reflected light in any place are i and j. The incident light i include the original incident light, the reflected light of pigment particles on medium interface, and the incident light reflected on external surface. Similarly, the reflected light j include the reflected light of pigment particles on medium interface, and the reflected light on bottom layer. The medium with thickness of D is divided into a series of successive infinite lamellas. Absorption, reflection and transmission would be taken place when i and j go through any infinite lamella. In figure 2, the thickness, absorption coefficient and scattering coefficient of any lamella are dx, K and S. The variation of i and j is as follows.

\[
\begin{align*}
\frac{dj}{dx} &= -Kjdx - Sjdx + Sidx \\
\frac{di}{dx} &= -Kidx - Sidx + Sidx
\end{align*}
\]

(2)

where \(dj\) include scattering Sidx of i, scattering -Sjdx of j and absorption -Kjdx of medium to j. The reason of negative sign in -Sjdx and -Kjdx is that the positive direction of x is vertical downward. The interpretation of \(di\) is similar. The solving process of this equation is as follows.

The formula 2 can be written

\[
\frac{dj}{dx} = S \cdot (i - aj) \quad \text{and} \quad \frac{di}{dx} = S \cdot (ai - j)
\]

(3)

where \(a = I + K/S\). Using derivative formula of quotient of two functions, it concludes

\[
\frac{d(j/i)}{dx} = S \cdot \left(\frac{j^2}{i^2} - 2a \frac{j}{i} + 1\right)
\]

(4)

Let spectral reflectance of medium with thickness of x is \(r = j/i\), we have

\[
\frac{dr}{dx} = S(r^2 - 2ar + l)
\]

(5)

Using integration, this is

\[
\int_{r_0}^{r} \frac{dr}{r^2 - 2ar + l} = S \int_{x_0}^{x} dx
\]

(6)

Let spectral reflectance of mediums with thickness of 0 and D are \(R_{0}\) and \(R\). The formula 6 can be written

\[
\int_{r_0}^{r} \frac{dr}{r^2 - 2ar + l} = \frac{1}{2b} \ln \left(\frac{(R - a - b)(R_e - a + b)}{(R - a + b)(R_e - a - b)}\right)
\]

\[
= S \int_{x_0}^{x} dx = SD
\]

(7)

where \(b = \sqrt{a^2 - l}\). It concludes

\[
R = \frac{1 - R_e (a - b \coth(bSD))}{a + b \coth(bSD) - R_e}
\]

(8)

where \(\coth(bSD)\) is the coth of bSD. When \(D \to \infty\), the formula 8 can be simplified

\[
R = 1 + K/S - \left[\left(K/S\right)^2 + 2(K/S)\right]^{\frac{1}{2}}
\]

(9)

It concludes

\[
K/S = \frac{(1 - R)^2}{2R}
\]

(10)

The formula 10 is named as the simplified Kubelka-Munk formula. The absorption/scattering
ratio of mixed pigment can be regarded as the linear combination of the ones of all basic pigments. That is

\[
\frac{(K_i/S_{mix})}{(K_i/S_{n})} = c_i \frac{(K_i/S_{n})}{(K_i/S_{n})} + \ldots + c_i \frac{(K_i/S_{n})}{(K_i/S_{n})}
\]

where \((K_i/S_{mix})\) and \((K_i/S_{n})\) are the absorption/scattering ratio of mixed pigment and substrate, \((K_i/S)\), and \(c_i\) are absorption/scattering ratio and percentage of \(i\)-th pigment, \(i = 1, 2, \ldots, n\).

2.2. Independent Component Analysis

2.2.1 Blind source separation

When the parameters of source and transmission channel are unknown, blind source separation (BSS) [16] can restore the independent component from observed signal, according to the statistical characteristic of input source signal. It is widely used in speech enhancement [17], communication system [18], biological medicine [19], seismic survey [20], information security [21] and other fields. Especially, the successful application in wireless communication and biological medicine has promoted its rapid development.

\[
\text{Figure 3: Basic mathematical model of BSS}
\]

The basic mathematical model of BSS is shown in figure 3. \(m\) mixed signals (or are called observed signals) \(x(t)\) are generated from \(n\) independent original signals \(s_i(t), i = 1, 2, \ldots, n, j = 1, 2, \ldots, m\).

\[
x(t) = \sum_{i=1}^{n} a_{ij} s_i(t)
\]

where \(a_{ij}\) is the mixing coefficient. Let original signal vector is \(S(t) = [s_1(t), s_2(t), \ldots, s_n(t)]\)\(^T\), observed signal vector is \(X(t) = [x_1(t), x_2(t), \ldots, x_m(t)]\)\(^T\), mixing matrix is \(A = [a_{ij}]_{n \times m}\). Then, the formula 12 can be written

\[
X(t) = A \cdot S(t)
\]

The result of BSS is a \(n \times m\) separative matrix \(W\). The multiplication \(Y(t)\) of \(W\) and \(X(t)\) should be close to \(S(t)\)

\[
Y(t) = W \cdot X(t)
\]

Not all the mixed signals can be separated using BSS, the separable observed signals must meet the following conditions.

1. The number of observed signals isn’t less than the one of original signals, e.g. \(m \geq n\).
2. The mixing matrix \(A\) is full rank.

All components \(s_i(t)\) of original signal \(S(t)\) are independent each other, and only up to one component obeys gaussian distribution.

If the above requirements are satisfied, \(Y(t)\) would be an estimation of \(S(t)\).

In practice, BSS algorithm have two uncertainty: order uncertainty and amplitude uncertainty. The former refers that the order of separated signal is different from the one of original signal. The later refers that the amplitude of separated signal is different from the one of original signal. However, the independent characteristic of signal couldn’t be affected, because the information of signal is determined by waveform. Therefore, the above requirements are still met, and the effect of BSS algorithm is all the same[22].

2.2.2. FastICA algorithm

BSS algorithms are divided into three kinds: BSS algorithm based on ICA, BSS algorithm based on second order statistics and BSS algorithm based on sparsity of original signal. Among these methods, BSS algorithm based on ICA (denoted as ICA for convenience) is widely used. There are a variety of ICA algorithms, such as ICA algorithm based on maximum likelihood criterion, ICA algorithm based on minimum mutual information, ICA algorithm based on information maximization and fast fixed-point ICA algorithm (FastICA). Among them, FastICA is applied widely[21], because it has the characteristics of iterative stability, fast convergence and separative matrix orthogonality.

\[
\text{Figure 4: Implementation process of FastICA algorithm}
\]

The implementation process of FastICA mainly includes two phases: whitening and separation, which is shown in figure 4. The operation sequence of whitening is: centering data -> singular value decomposition (SVD) -> generating whitening matrix. In this phase, the observed signal \(X(t)\) is...
converted into the orthonormal signal \( Z(t) \), in order to ensure the independence of signal. The main operation in the phase of separation is constructing the separative matrix. The main strategies include tightening approximation method and equilibrium approximation method. The former is also called one-dimension algorithm, and only one independent component is chosen and updated at a time. The latter is also called multi-dimension algorithm, and all independent components are updated at the same time.

2.2.3. Composition analysis of mixed pigment

Spectrum is the accurate expression of pigment, and the spectrum of mixed pigment is the linear combination of spectrums of basic pigments. First, the spectrum is taken as the random signal. Second, based on the spectrum database of known pigments, the type and percentage of basic pigment are calculated by integrating with ICA, spectrum similarity, Kubelka-Munk theory, etc. Finally, the traceability of mixed pigment is achieved.

3. METHODOLOGY

3.1. Composition Analysis Method of Mixed Pigment

The \( K/S \) of \( n \) basic pigments are \( SP_1, SP_2, \ldots, SP_n \), where \( SP = \{SP_1, SP_2, \ldots, SP_n\} \), \( i = 1,2,\ldots,n \). The \( K/S \) of \( m \) mixed pigments are \( DP_1, DP_2, \ldots, DP_m \), where \( DP = \{DP_1, DP_2, \ldots, DP_m\} \) and \( j = 1,2,\ldots,m \). The mixing process is expressed as

\[
DP = \sum_{i=1}^{m} c_{ij} * SP_i,
\]

where \( c_{ij} \) is the mixing coefficient, \( p = 1,2,\ldots,m \), \( q = 1,2,\ldots,n \), and \( '*' \) represents vector-scalar multiplication. With vector representation, \( SP = [SP_1, SP_2, \ldots, SP_n] \), \( DP = [DP_1, DP_2, \ldots, DP_m] \) and \( C = [c_{ij}] \), the above mixing process can be written \( DP = C * SP \).

If \( K/S \) of basic pigment and mixed pigment meet the separable condition of mixed signal, \( DP \) of mixed pigment can be proceed by FastICA algorithm. The output is \( \hat{SP} = [\hat{SP}_1, \hat{SP}_2, \ldots, \hat{SP}_n] \), where \( \hat{SP} \) is an approximate representation of \( SP \). The types of \( n \) basic pigments corresponding to \( \hat{SP} \) are determined by calculating the spectrum similarities between \( \hat{SP} \) and all known pigments. The reflection spectrums of \( n \) basic pigments are \( SP_1, SP_2, \ldots, SP_n \). The vector representation is \( SP = [SP_1, SP_2, \ldots, SP_n] \), where \( SP \) is an accurate representation of \( SP \). That is to say \( DP = C * SP \).

One thing to note here is that \( K/S \) of actual basic pigments may exist correlation, although the one in our experiment are independent each other. Therefore, the actual data must be decorrelated. The main decorrelation methods include Karhunen-Loeve Transform (KLT) [23], decorrelation-stretch Transform [24], wavelet transform [25], etc. Sometimes, orthogonalization is required, and the main methods include gram-schmidt transform [26], householder transformation [27], etc.

3.2. Effect Evaluation

First of all, \( SP \) and \( SP' \) are normalized. After that, the effect evaluation is carried out with the following methods. In these formulas, \( i \) and \( j \) are serial number of pigments, their ranges are \( i = 1,2,\ldots,n \) and \( j = 1,2,\ldots,n \). \( SP \) is the actual data, and \( SP' \) is the separated result.

1. Relative mean square error (RMSE)

The RMSE of \( i \)-th pigment is shown in formula 15.

\[
RMSE_i = \sqrt{\frac{1}{N} \sum_{n=1}^{N} \left( SP_i - SP'_i \right)^2} \quad \text{(15)}
\]

2. Signal to Noise Ratio (SNR)

Let \( SP \) is signal, and \( SP - SP' \) is noise. Then, the SNR of \( i \)-th pigment is shown in formula 16.

\[
SNR = 10 \log_{10} \left( \frac{\sum_{i=1}^{n} (SP_i^2)}{\sum_{i=1}^{n} (SP_i - SP'_i)^2} \right) \quad \text{(16)}
\]

3. Similarity

The similarity between \( SP_i \) and \( SP'_i \) is shown in formula 17.

\[
\xi_{ij} = \frac{\sum_{i=1}^{n} SP_i SP'_i}{\sqrt{\sum_{i=1}^{n} SP_i^2} \sum_{i=1}^{n} SP'_i^2} \quad \text{(17)}
\]

4. EXPERIMENTS AND RESULTS

4.1. Material

Munsell color system is proposed by American chromatist Munsell. And it has become one of the recognized color systems after repeated revisions by American National Standards Institute (ANSI) and Optical Society of America (OSA). For this method, three elements (hue, brightness and saturation) of all sorts of color can be described with the Munsell color cube as shown in figure 5.
According to brightness and reflectivity, Munsell color system is divided into two kinds: full gloss and semi-gloss. The full gloss that includes 1600 spectrums of Munsell color card was chosen in our experiment. The reflection spectrum database was constructed by downloading the full gloss spectral curves from Website of Spectral Color Research group (http://www.uef.fi/spectral/home). The simulated spectrum was generated by mixing spectrum of this database. And the experiments of composition analysis of mixed pigment were carried out under the circumstance of normality and disturbance.

4.2. Composition Analysis Experiment under Normal

The $K/S$ curves of four chosen Munsell color cards are shown in figure 6. In every subfigure, x-coordinate represents wavelength whose range is from 400nm to 780nm, and y-coordinate is $K/S$. The subfigure’s name is the color numbers in Munsell color system. For convenience of description, these curves are called original $K/S$ curves.

Three groups of mixed $K/S$ curves were processed with FastICA algorithm. The number of decomposed $K/S$ curves in each group is four. These curves are called separated $K/S$ curves. The correspondence between separated $K/S$ curves and original $K/S$ curves was determined by comparing spectrum similarities between them. In order to improve the matching effect, all original $K/S$ curves and separated $K/S$ curves were normalized. The matching result is shown in figure 8. In every subfigure, original $K/S$ curves are shown with

$$M_1 = \begin{bmatrix} 0.23 & 0.10 & 0.48 & 0.46 \\ 0.10 & 0.23 & 0.29 & 0.20 \\ 0.33 & 0.45 & 0.03 & 0.05 \\ 0.34 & 0.22 & 0.20 & 0.29 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 0.15 & 0.42 & 0.31 & 0.20 & 0.25 \\ 0.30 & 0.10 & 0.01 & 0.22 & 0.05 \\ 0.22 & 0.06 & 0.43 & 0.27 & 0.48 \\ 0.33 & 0.42 & 0.25 & 0.31 & 0.22 \end{bmatrix}$$

$$M_3 = \begin{bmatrix} 0.38 & 0.07 & 0.29 & 0.30 & 0.18 & 0.19 \\ 0.17 & 0.51 & 0.17 & 0.33 & 0.51 & 0.33 \\ 0.37 & 0.03 & 0.15 & 0.12 & 0.14 & 0.07 \\ 0.08 & 0.39 & 0.39 & 0.25 & 0.17 & 0.41 \end{bmatrix}$$

Three groups of mixed $K/S$ curves were processed with FastICA algorithm. The number of decomposed $K/S$ curves in each group is four. These curves are called separated $K/S$ curves. The correspondence between separated $K/S$ curves and original $K/S$ curves was determined by comparing spectrum similarities between them. In order to improve the matching effect, all original $K/S$ curves and separated $K/S$ curves were normalized. The matching result is shown in figure 8. In every subfigure, original $K/S$ curves are shown with
solid line, and three corresponding separated K/S curves are shown with scatter diagrams that are consisted by cross, "X" mark and right-pointing triangle respectively.

![Figure 9: Average separated K/S](image)

In order to facilitate evaluation, the separated K/S curves of figure 8 were averaged. In figure 9, curves \( T_1, T_2, T_3 \) and \( T_4 \) are average separated K/S curves of figure 8.a, 8.b, 8.c and 8.d, and they are corresponding to figure 6.a, 6.b, 6.c and 6.d respectively. According to criterions in section 3.2, \( SP_1 \) and \( SP_2 \) were generated from curve data of figure 6 and figure 9, and RMSE, SNR and similarity were calculated, as shown in table 1. In this table, all RMSE are less than 10\(^{-5}\). It indicates that the separated K/S curve of this method is close to the actual one.

**Table 1: Effect evaluation of separated K/S curves**

<table>
<thead>
<tr>
<th>Criterion Separated K/S curve</th>
<th>RMSE</th>
<th>SNR</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curve ( T_1 )</td>
<td>0.000000433</td>
<td>0.9028</td>
<td>0.9594</td>
</tr>
<tr>
<td>Curve ( T_2 )</td>
<td>0.000000955</td>
<td>0.9185</td>
<td>0.9397</td>
</tr>
<tr>
<td>Curve ( T_3 )</td>
<td>0.000000609</td>
<td>0.9886</td>
<td>0.9857</td>
</tr>
<tr>
<td>Curve ( T_4 )</td>
<td>0.000000633</td>
<td>0.9796</td>
<td>0.9509</td>
</tr>
</tbody>
</table>

The percentage of basic pigment in each group was calculated. According to method in section 3.1, \( DP \) of group 1, 2 and 3 were generated from curve data of figure 7.a, 7.b and 7.c, \( SP \) was generated from curve data of original K/S curves in figure 8, and percentage was calculated with \( C = DP \times (SP)^{\frac{1}{2}} \), as shown in table 2. In this table, the calculated percentage is similar with formula 18, 19 and 20. The results demonstrate that this method can get accurate percentage.

**4.3. Composition Analysis Experiment under Disturb**

In order to test the anti-interference performance of this method, each group of mixed K/S curves was mixed with one different K/S curve, and they are called new mixed K/S curves. In figure 10, curves \( R_1, R_2, \) and \( R_3 \) were added into group 1, 2 and 3 respectively. These additional K/S curves are called original disturbance K/S curves.

**Mixed K/S curve**

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Curve 1</th>
<th>Curve 2</th>
<th>Curve 3</th>
<th>Curve 4</th>
<th>Curve 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0 GY 7 / 8</td>
<td>0.23</td>
<td>0.10</td>
<td>0.49</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>10.0 PB 5 / 2</td>
<td>0.23</td>
<td>0.29</td>
<td>0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.0 PP 9 / 2</td>
<td>0.33</td>
<td>0.02</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral 0.75</td>
<td>0.34</td>
<td>0.22</td>
<td>0.20</td>
<td>0.30</td>
<td></td>
</tr>
</tbody>
</table>

**Mixed K/S curve**

<table>
<thead>
<tr>
<th>Group 2</th>
<th>Curve 1</th>
<th>Curve 2</th>
<th>Curve 3</th>
<th>Curve 4</th>
<th>Curve 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0 GY 7 / 8</td>
<td>0.15</td>
<td>0.42</td>
<td>0.31</td>
<td>0.19</td>
<td>0.24</td>
</tr>
<tr>
<td>10.0 PB 5 / 2</td>
<td>0.30</td>
<td>0.10</td>
<td>0.01</td>
<td>0.22</td>
<td>0.06</td>
</tr>
<tr>
<td>10.0 PP 9 / 2</td>
<td>0.21</td>
<td>0.16</td>
<td>0.28</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Neutral 0.75</td>
<td>0.33</td>
<td>0.42</td>
<td>0.25</td>
<td>0.31</td>
<td>0.22</td>
</tr>
</tbody>
</table>

**Mixed K/S curve**

<table>
<thead>
<tr>
<th>Group 3</th>
<th>Curve 1</th>
<th>Curve 2</th>
<th>Curve 3</th>
<th>Curve 4</th>
<th>Curve 5</th>
<th>Curve 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.0 GY 7 / 8</td>
<td>0.38</td>
<td>0.08</td>
<td>0.29</td>
<td>0.30</td>
<td>0.17</td>
<td>0.19</td>
</tr>
<tr>
<td>10.0 PB 5 / 2</td>
<td>0.17</td>
<td>0.50</td>
<td>0.17</td>
<td>0.34</td>
<td>0.51</td>
<td>0.33</td>
</tr>
<tr>
<td>10.0 PP 9 / 2</td>
<td>0.37</td>
<td>0.03</td>
<td>0.15</td>
<td>0.12</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>Neutral 0.75</td>
<td>0.08</td>
<td>0.39</td>
<td>0.39</td>
<td>0.24</td>
<td>0.18</td>
<td>0.41</td>
</tr>
</tbody>
</table>

The new mixed K/S curves were processed with FastICA algorithm, and the number of decomposed K/S curves in each group is five. The five K/S curves in each group can be divided into two kinds by comparing similarities between them and original K/S curves. One is four separated original K/S curves that correspond to the original K/S curves of figure 6. The other is one separated disturbance K/S curve that corresponds to the disturbance K/S curve of figure 10.

In the following, the five mixed K/S curves by adding disturbance K/S curve \( R \) into mixed K/S curves of figure 7.a is taken as an example for describing the classification method. The five
separated K/S curves of FastICA algorithm are shown in figure 11, and the similarity matrix between them and original K/S curves is shown in table 3. In this table, the maximum absolute values of similarity in row 1, 2, 3 and 4 are more than 0.929, but the one in row 5 is less than 0.24. Thus, it can conclude that curves $F_1$, $F_2$, $F_3$ and $F_4$ are separated original K/S curves, and curve $F_5$ is a separated disturbance K/S curve.

Table 3: Similarity and correspondence between the separated K/S curve and the original K/S curve

<table>
<thead>
<tr>
<th>Original K/S curve</th>
<th>Separated K/S curve</th>
<th>10.0 GY / 8</th>
<th>10.0 PB / 5</th>
<th>10.0 PP / 2</th>
<th>Neutral</th>
<th>Correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curve F_1</td>
<td>0.2753</td>
<td>0.0326</td>
<td>0.9825</td>
<td>0.1505</td>
<td>10.0 PP</td>
<td>9 / 2</td>
</tr>
<tr>
<td>Curve F_2</td>
<td>0.0376</td>
<td>0.9307</td>
<td>0.0266</td>
<td>-0.0738</td>
<td>10.0 PB</td>
<td>5 / 2</td>
</tr>
<tr>
<td>Curve F_3</td>
<td>0.9294</td>
<td>-0.1952</td>
<td>-0.1745</td>
<td>-0.2383</td>
<td>10.0 GY</td>
<td>7 / 8</td>
</tr>
<tr>
<td>Curve F_4</td>
<td>-0.0520</td>
<td>0.2294</td>
<td>0.0599</td>
<td>0.9524</td>
<td>Neutral</td>
<td>0.75</td>
</tr>
<tr>
<td>Curve F_5</td>
<td>0.2374</td>
<td>0.2049</td>
<td>0.0054</td>
<td>0.0893</td>
<td>disturbance</td>
<td></td>
</tr>
</tbody>
</table>

According to the above steps, the separated K/S curves of other groups were generated, as shown in figure 12 and 13. The separated original K/S curves corresponding to figure 6.a, 6.b, 6.c and 6.d are shown in figure 12.a, 12.b, 12.c and 12.d, and each subfigure have three K/S curves derived from three groups. The separated disturbance K/S curves corresponding to figure 6.a, 6.b and 6.c are shown in figure 13.

The separated original K/S curves in figure 12 were averaged, as shown in figure 14. In this figure, curves $G_1$, $G_2$, $G_3$ and $G_4$ correspond to figure 6.a, 6.b, 6.c and 6.d. According to this correspondence and the criterions in section 3.2, the evaluation results were calculated, as shown in table 4. In this table, all RMSE are less than $10^{-5}$. The results show the robust of this method under disturbance.

Table 4: Effect evaluation of separated original K/S curves

<table>
<thead>
<tr>
<th>Criterion</th>
<th>RMSE</th>
<th>SNR</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curve G_1</td>
<td>0.00000360</td>
<td>11.3021</td>
<td>0.9630</td>
</tr>
<tr>
<td>Curve G_2</td>
<td>0.00000152</td>
<td>13.1710</td>
<td>0.9759</td>
</tr>
<tr>
<td>Curve G_3</td>
<td>0.00000797</td>
<td>14.6047</td>
<td>0.9827</td>
</tr>
<tr>
<td>Curve G_4</td>
<td>0.000070725</td>
<td>9.7845</td>
<td>0.9475</td>
</tr>
</tbody>
</table>

Before calculating percentage, the original disturbance K/S curve must be deleted from the new mixed K/S curves. For each separated disturbance K/S curves of figure 13, the spectrum similarities between it and original disturbance K/S curves of figure 10 are more than the one between them and mixed K/S curves of figure 7. Therefore, the new mixed K/S curve with the largest spectrum similarity is the original disturbance K/S curve. After deleting original disturbance K/S curves, the mixed K/S curves of figure 7 are restored, and the percentage was calculated according to the method in section 3.1, as shown in table 5. The result of this table is similar with formula 18, 19 and 20. These data demonstrate that this method can get accurate percentage.
Table 5: Calculated percentage of each group

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Curve1</th>
<th>Curve2</th>
<th>Curve3</th>
<th>Curve4</th>
<th>Curve5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral 0.75</td>
<td></td>
<td>0.34</td>
<td>0.21</td>
<td>0.20</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>10.0 PP 9/2</td>
<td></td>
<td>0.33</td>
<td>0.45</td>
<td>0.03</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>10.0 PB 5/2</td>
<td></td>
<td>0.10</td>
<td>0.24</td>
<td>0.29</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>10.0 GY 7/8</td>
<td></td>
<td>0.23</td>
<td>0.10</td>
<td>0.48</td>
<td>0.46</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSION

The composition analysis of mixed pigment has extensive application space. RS is a commonly used method in this field, but the accuracy of type judgment of basic pigment would be affected to some extent. In our study, the mixed pigments are expressed as spectrums, and the types and percentages of basic pigments are determined by integrating with ICA, spectrum similarity, Kubelka-Munk theory, etc. To validate this method, the experiment with simulated spectrum from Munsell color system was carried out. The results demonstrate that the type and percentage of basic pigment can be determined accurately under the circumstance of normality and disturbance, which meets the traceability of mixed pigment.

In this paper, the composition analysis method of mixed pigment based on spectrum expression and ICA is proposed and proved, but it is only the beginning. In practice, a lot of elements should be considered. In order to improve its practicability, decorrelation and orthogonalization would be integrated with our study.

ACKNOWLEDGMENT

The authors are grateful to Yingqing Xu for his discussion and suggestion about background knowledge, validity, experiment, etc. This work was supported by National Basic Research Program of China (Grant No. 2012CB725300), National Natural Science Foundation of China (Grant No. 61373072) and China Postdoctoral Science Foundation (Grant No. 2013M530035).

REFERENCES:


