White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/176915/

Version: Accepted Version

Article:
Liang, J, Zhu, Q, Liu, Q et al. (2022) Optimal selection of representative samples for efficient digital camera-based spectra recovery. Color Research & Application, 47 (1). pp. 107-120. ISSN: 0361-2317

https://doi.org/10.1002/col.22718

© 2021 Wiley Periodicals LLC. This is the peer reviewed version of the following article: Liang, J, Zhu, Q, Liu, Q et al. (1 more author) (2022) Optimal selection of representative samples for efficient digital camera-based spectra recovery. Color Research & Application, 47 (1). pp. 107-120. ISSN 0361-2317, which has been published in final form at https://doi.org/10.1002/col.22718. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions. This article may not be enhanced, enriched or otherwise transformed into a derivative work, without express permission from Wiley or by statutory rights under applicable legislation. Copyright notices must not be removed, obscured or modified. The article must be linked to Wiley's version of record on Wiley Online Library and any embedding, framing or otherwise making available the article or pages thereof by third parties from platforms, services and websites other than Wiley Online Library must be prohibited. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Optimal selection of representative samples for efficient digital camera-based spectra recovery

Jinxing Liang[1,4], Qiang Zhu[1,4], Qiang Liu[2], Kaida Xiao[3]

[1]School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan, Hubei, China

[2]School of Printing and Packaging, Wuhan University, Wuhan, Hubei, China

[3]School of Design, University of Leeds, Leeds, UK

[4]Engineering Research Center of Hubei Province for Clothing Information,Wuhan, Hubei, China

Correspondence

Qiang Liu, School of Printing and Packaging, Wuhan University, Wuhan, 430079, Hubei, China

Email: liuqiang@whu.edu.cn

Abstract: For digital camera-based spectra recovery, the spectral reflectance of the object being imaged always needs to be accurately recovered using training samples from available database. Considering the heavy workload when using all samples in database as training samples in practice, a new representative samples selection method is proposed for efficient digital camera-based spectra recovery based on single RGB image. The representative simulation system is firstly constructed through correlation analysis of spectra recovery results of different systems, and based on the representative simulation system, a few number of representative samples are selected from the database based on minimum of the defined simulate spectra recovery error. The effectiveness of the proposed method is evaluated and compared with existing method. As the results show, the proposed method outperform the existing methods, and the robustness of the selected representative samples is consistent with the database in practical applications.

Key words: multispectral imaging; spectra recovery; digital camera; representative samples selection

1. INTRODUCTION

The spectral reflectance of the object is regarded as the fingerprint of the color, and at the same time it also indicate the intrinsic physical and chemical properties of the object, therefore, the surface spectral of object plays an important role in many application fields such as cultural heritages protection, computer vision, agriculture, biomedical and so on.[1-4] Digital camera-based spectra recovery based on single RGB image, as one of the important way for object spectra acquisition, was extensively researched and applied recently because of the low price of the equipment, the higher spatial resolution and the flexible application of the system.

For digital camera-based spectra recovery, the spectral reflectance of the object being imaged always needs to be accurately recovered using training samples from available database. In different application fields such as cultural heritages protection, computer vision, printing, textile and so on, the specific sample databases including a large number of samples that cover the range of commonly used colors are constructed for different purposes. These sample database also can be used for spectra recovery. However, previous studies shows that for spectra recovery, these databases that containing lots of samples have significant sample redundancy,[5] and there is no need to use all samples in database as training samples, otherwise, it will cause heavy workload and inconvenience for sample capturing and processing, especially in outdoor applications. Thus, the optimal selection of representative samples from available database has been an important aspect for spectra recovery.

Hardeberg proposed an representative samples selection method for camera spectral sensitivity functions estimation based on minimum conditional number,[6] Mohammadi *et al* proposed the spectra recovery target development method based on hierarchical cluster analysis,[7] Cheung *et al* proposed four optimal samples selection rules based on the criterion that the subsequent sample to be selected into the representative sample subset should be as different as possible from those already selected,[8] Shen *et al* proposed the eigenvector-based and virtual-imaging-based method for representative samples selection.[9] These methods either follow the criterion of maximizing the difference in spectral or chromaticity between the selected samples, or make the spectral equivalence of the selected samples to the sample database. Although the selected result is stability, the representative samples selected by these methods are not optimal for specific system applications. Recently, Eckhard et al proposed a representative samples selection method based on recursive top-down algorithm for printed ink spectra recovery for quality assessment in in-line print inspection system,[10] however, the selected samples of are directed to the specific system, the stability and versatility to other different type of systems cannot be guaranteed.

Inspired by existing researches, this study presents a new representative samples selection method for efficient digital camera-based spectra recovery based on single RGB image. The first step of the method is to construct a representative simulation system. Then, with the constructed representative simulation system, the representative samples are selected one-by-one from the database based on minimum spectra recovery error of the selected samples to the database. The spectra recovery error is defined by multiplying the root-mean-square error (RMSE) and goodness-of-fit coefficient (GFC) error,[11] which will make sure that the most spectral representative samples can be selected. The novelty of the proposed method is that the common features of digital camera-based spectra recovery imaging system based on single RGB image is analyzed and incorporated into the samples selection process, which make sure the selected samples are superiority to the samples selected by exiting methods. Moreover, the robust of the selected representative samples of the proposed method in practical applications is also validated through simulation and practice experiment.

2. THEORY AND METHODS

2.1 Fundamentals of digital camera-based spectra recovery

Assuming a linear opto-electronic transfer function in the camera,[12] the camera response formed in a pixel of an image can be formulated as equation (1):

$$d_i = \int_{\Phi} l(\lambda) s_i(\lambda) r(\lambda) d\lambda + e \qquad (1)$$

where the camera response d_i is related to the channel i of a pixel in the image, λ is the wavelength, $l(\lambda)$ is the relative spectral power distribution of the illuminant, $s_i(\lambda)$ is the spectral sensitivity functions of the ith channel of the camera, $r(\lambda)$ is the spectral reflectance of a surface point in the scene to be imaged, Φ is the spectra integrated range of the camera imaging system, e is the system additive imaging noise and often simulated with the Gaussian white noise in different researches.[13-15] In practice, we can formulate a discrete version of equation (1) as:

$$\mathbf{d} = \mathbf{M}\mathbf{r} + \mathbf{e} \qquad (2)$$

where $\mathbf{d}$ is the response vector of a pixel in the image, $\mathbf{M}$ is the overall spectral sensitivity function matrix of the camera imaging system including the product of the matrix form of $l(\lambda)$ and $s_i(\lambda)$, and $\mathbf{r}$ denotes the spectral reflectance vector of a surface point in the scene to be imaged, and $\mathbf{e}$ is the noise vector. For digital camera-based spectra recovery, the first step is to construct the spectra recovery matrix based on the camera imaging model. Many methods have been proposed to calculate the spectra recovery matrix, such as pseudoinverse, PCA, Wiener estimation, compressive sensing and so on.[16-20] In addition, deep learning based methods, such CNN, U-net and GAN-based methods, are also introduced into spectra recovery in recent years with the fast development of computer visual technology.[21-23] In this study, taking the commonly used pseudoinverse method as example, the spectra recovery matrix is calculated as in equation (3):

$$\mathbf{Q} = \mathbf{R}_{\text{train}} \mathbf{D}_{\text{train}}^{+} \qquad (3)$$

where $\mathbf{Q}$ is the spectra recovery matrix, $\mathbf{R}_{\text{train}}$ and $\mathbf{D}_{\text{train}}$ are the spectral reflectance and camera response matrix of training set, and superscript + is the pseudoinverse operator. With the established spectra recovery matrix $\mathbf{Q}$, the spectral reflectance of any pixel in the image can be recovered as shown in equation (4).

$$\mathbf{r}_{\text{test}} = \mathbf{Q}\mathbf{d}_{\text{test}} \qquad (4)$$

where $\mathbf{d}_{\text{test}}$ is the camera response vector of a pixel in the image, $\mathbf{r}_{\text{test}}$ is the corresponding recovered spectral reflectance vector. The current research of digital camera-based spectra recovery are all based on the linear camera imaging model and spectra recovery theory, and the typical and commonly used regularized polynomial model-based method for spectra recovery from single RGB image is adopted in this study, the implementation details of the method can refer to reference 19.

2.2 Theory basis to construct the representative simulation system

Different from the type of multi-channel spectra recovery imaging systems where both the number of channels and the spectral sensitivity functions of them are often different from each other,[24-28] the digital camera-based spectra recovery system based on single RGB image have common features, they are all have three imaging channels (Red, Green and Blue) to simulate the color perception of human visual system. The research carried by Jiang *et al* showed that the spectral sensitivity functions of different brands of digital cameras all have similar distribution shape,[29] and it is accurate

enough to estimate the spectral sensitivity functions of a digital camera using only the first two dimension of eigenvectors of spectral sensitivity functions database. Research conclusions of Jiang *et al* provide the basis for the hypothesis of this research.

The assumption of construct a representative simulation spectra recovery system based on single RGB image is reasonable, and it is not difficult to establish such a representative simulation system. In this study, we use average results of a spectral sensitivity functions database developed by Jiang *et al* to construct the representative simulation system.[29] Details of the representative simulation system are illustrated in the follow sections.

2.3 Proposed method for representative samples selection

Once the representative simulation system is established, the representative samples are selected one-by-one from the sample database based on minimum simulate spectra recovery error of the selected samples to the database. From selection of the first representative sample, the corresponding spectra recovery matrix $\mathbf{Q}_i$ for each of the samples in database $\mathbf{\Theta}$ is calculated as in equation (5):

$$\mathbf{Q}_i = \mathbf{r}_i \cdot RPM(\mathbf{d}_i) \qquad (5)$$

where $RPM(*)$ represents the regularized polynomial model-based spectra recovery method that proposed in reference 19, $\mathbf{r}_i$ is the spectral reflectance vector of the ith sample in database, $\mathbf{d}_i$ is the corresponding simulated camera response vector of the ith sample under representative simulation system with noise added, $\mathbf{Q}_i$ is the corresponding spectra recovery matrix. Then, the spectral reflectance of the database $\mathbf{\Theta}$ is estimated using $\mathbf{Q}_i$, as illustrated in equation (6).

$$\mathbf{R}_{\Theta,i} = \mathbf{Q}_i \cdot \mathbf{D}_\Theta \qquad (6)$$

where $\mathbf{D}_\Theta$ is the simulation camera response matrix of the database under the representative simulation system, $\mathbf{R}_{\Theta,i}$ is the recovered spectral reflectance matrix of database. In the next step, the average RMSE and GFC between the recovered and ground-truth spectral reflectance of database is calculated as shown in equation (7) and equation (8), and the total spectra recovery error is calculated by multiplying the RMSE and 1-GFC, as shown in equation (9).

$$\mathrm{RMSE}_i = E\left\{\left\|\mathbf{R}_{\Theta,i} - \mathbf{R}_\Theta\right\|\right\} \qquad (7)$$

$$\mathrm{GFC}_i = F\left\langle \mathbf{R}_{\Theta,i} \middle| \mathbf{R}_\Theta \right\rangle \qquad (8)$$

$$\mathrm{TOTAL}_i = \mathrm{RMSE}_i \times (1 - \mathrm{GFC}_i) \qquad (9)$$

where $\mathbf{R}_\Theta$ is the measured ground-truth spectral reflectance matrix of the database, $E\{\|*\|\}$ and $F<*>$ are the functions to calculate the average RMSE and GFC of spectral reflectance, details to calculate RMSE and GFC between any two spectral reflectance can refer to equation (13) and (14), TOTAL_i is the corresponding spectral reflectance recovered error of using the ith sample in database as training sample. And as indicated in equation (10), the sample with the smallest spectra recovery error will be selected as the first representative sample of the database.

$$s_1 = \arg\min_{\mathbf{r}_i \in \mathbf{\Theta}} \mathrm{TOTAL}_i \qquad (10)$$

where s_1 represent the first selected representative sample from the database, and the selected representative samples subset $\mathbf{\Omega}_1$ get the first sample as shown in equation (11).

$$\mathbf{\Omega}_1 = \{s_1\} \qquad (11)$$

Using the same rules as select the first representative sample, when selecting the remaining 2nd to *k*th representative samples, it just need to repeat the steps as shown in equation (12), where *j* indicates the *j*th sample in the database but excluding the selected samples. Until the TOTAL_j becomes converge, the selection of the representative samples are completed and the final representative samples subset $\mathbf{\Omega}_k$ is acquired.

$$\begin{cases} \mathbf{Q}_j = \left\{ \mathbf{\Omega}_{k-1} \cup \mathbf{r}_j \right\} \cdot RPM\left(\mathbf{d}_{\left\{ \mathbf{\Omega}_{k-1} \cup \mathbf{r}_j \right\}} \right) \\ \mathbf{R}_{\Theta,j} = \mathbf{Q}_j \cdot \mathbf{D}_\Theta \\ \mathrm{TOTAL}_j = \mathrm{RMSE}_i \times (1 - \mathrm{GFC}_i) \\ s_k = \arg\min_{\mathbf{r}_j \in \mathbf{\Theta}} \mathrm{TOTAL}_j \\ \mathbf{\Omega}_k = \mathbf{\Omega}_{k-1} \cup \left\{ s_k \right\} \end{cases} \tag{12}$$

3. EXPERIMENTAL

The experiment is carried out sequentially in two sections. The first section is designed to demonstrates the feasibility to construct a representative simulation system for spectra recovery based on single RGB image. In the second section, the effectiveness of the proposed representative samples selection method is verified and compared with the existing commonly used methods.

3.1 Representative simulation system construction

To prove that it is reasonable to establish a representative simulation system for digital camera-based spectra recovery based on single RGB image, and at the same time to construct the representative simulation system, we randomly select four spectral sensitivity functions of digital cameras in daily use and commercially available from Jiang's database to construct four simulation systems.[29] The distribution of spectral sensitivity functions are plotted in Figure 1. It can be seen from Figure 1 that even though there are some distinguishes in detail between these functions, they all show similar curve shape in general. These simulation systems will be used to test the correlation relationship between their spectra recovery results, which could help us to explore whether the difference of simulation systems have significant influence on spectra recovery results.

In addition, in order to perform correlation analysis of spectra recovery results between simulation systems and real system, we used a Nikon D7200 digital camera to construct the real system. The spectral sensitivity functions that estimated using Jiang's method[30] is plotted in Figure 2. The correlation between simulation systems, and between simulation systems and the real system together will help us to explore whether it is really reasonable to establish a representative simulation system for digital camera-based spectra recovery based on single RGB image.

For a more comprehensive correlation analysis of spectra recovery results between simulation and real systems, we used four color sample sets including X-rite ColorChecker SG chart (SGchart), Dunhuang pigment chart (DHPchart),[31] textile samples (Textile)[32] and the printed skin color chart (Skinchart) to construct 16 groups of spectra recovery test pairs. The spectral reflectance of textile samples were measured by Konica Minolta CM-3600A spectrophotometer, and the

spectral reflectance of other three sample set were measured using the X-rite i1-pro spectrophotometer. With the measured spectral reflectance, the colorimetric values were calculated under the CIE specifications.[33]

The CIE D65 standard illuminant is used as the light source for simulation systems, and one percent of Gaussian white noise is added to the simulated system as the additive imaging noise. For the real system, because of the limitation of experimental settings, the fluorescent lamp panel with the correlated color temperature (CCT) of ~6000K is used as the light source for uniformity illumination. The difference of the illumination conditions may have some influence on the calculation of evaluation metrics, but according to the previous studies[34,35] and the experiment results in this study, the difference of the illumination may have not significant influence on the experiment results of our research. In addition, to make sure the image has sufficient dynamic range and the response of white color sample with no truncate, the aperture, ISO and shutter speed of the digital RGB camera were set to f5.6, 200 and 1/8 second, respectively.

The spectra recovery error of all 16 test pairs are calculated for both simulation and real system. Method to calculate the RMSE and GFC are showed in equation (13) and equation (14) respectively, where $\mathbf{r}_1$ is the ground-truth spectral reflectance of the testing sample, $\mathbf{r}_2$ is the corresponding recovered spectral reflectance, superscript T is the transpose operator, N is the sampling number of spectral in visible spectrum.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}(\mathbf{r}_1 - \mathbf{r}_2)^T(\mathbf{r}_1 - \mathbf{r}_2)} \qquad (13)$$

$$\mathrm{GFC} = \frac{\mathbf{r}_1^T \mathbf{r}_2}{\sqrt{\mathbf{r}_1^T \mathbf{r}_1}\sqrt{\mathbf{r}_2^T \mathbf{r}_2}} \qquad (14)$$

With the support of the experiment results in this section, a representative simulation system will be finally constructed by the using the CIED65 standard illuminant, the regularized polynomial model-based spectra recovery method,[19] the average data of a spectral sensitivity functions database, and the one percent of Gaussian white additive system noise. Details to construct the representative simulation system will be described in section 4.1.1.

3.2 Representative samples selection

In this section, we used a mineral pigment sample database that containing a total of 784 samples and the Nikon D7200 digital camera to test the effectiveness of the proposed representative samples selection method. The database with different particle size of rock pigment is acquired from Nakagawa Gofun, it contains almost all kinds of the rock pigments used to painting the East ancient murals. The spectral reflectance data of the samples in database was measured with X-rite i1-pro spectrophotometer. With the constructed representative simulation system, the representative samples of the pigment sample database are selected using the proposed method.

Once the representative samples are selected, its practical effectiveness as training samples for spectra recovery is tested and compared with existing commonly used methods. For the practical testing, the digital images of the pigment sample database is captured by the Nikon D7200 digital camera. The mean raw responses of each color patch of its central part with area of 35×35 pixels is extracted for spectra recovery test. To evaluate the spectra recovery accuracy, except for the metric of RMSE and GFC, the average CIELAB color difference (ΔE_{ab}) that calculated under three

illuminants of CIEA, CIED50 and CIED65 was also used to evaluate the spectra recovery accuracy in this section. Method to calculate CIELAB color difference is shown in equation (15), where (L_1, a_1, b_1) and (L_2, a_2, b_2) are the colorimetric values of a sample corresponding to the ground-truth and recovered spectral reflectance.

$$\Delta E_{ab} = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \qquad (15)$$

4. RESULTS AND DICUSSION

4.1 Results

4.1.1 Results of representative simulation system construction

The spectra recovery RMSE and GFC of 16 test pairs under four simulation systems and the real system are summarized in Table 1 and Table 2 respectively. It can be seen from Table 1 and Table 2 that the spectra recovery RMSE and GFC of 16 test pairs under four simulation systems are very similar, and there is no big difference of RMSE and GFC for any of the test pair under four simulation systems.

In addition, in order to make a more comprehensive analysis of the experimental results, correlation analysis is performed to the spectra recovery error of RMSE and GFC between simulation systems, and between simulations and the real system. The Pearson correlation coefficients (with p-values of zero) were calculated and summarized in Table 3 and Table 4. We can see from Table 3 that correlation coefficients of spectra recovery error of RMSE between the simulation systems are all greater than 0.99, and the correlation coefficients of spectra recovery RMSE between simulation systems and the real system are all greater than 0.98. From Table 4, we can see that the correlation coefficients of spectra recovery error of GFC between the simulation systems are all greater than 0.99, and the correlation coefficients of spectra recovery GFC between simulation systems and the real system are all greater than 0.91.

The results in Table 3 and Table 4 indicated that there are very strong correlations of spectra recovery results between simulation systems, and between simulation systems and the real system. Furthermore, the one-way ANOVA and Tukey's post-hoc tests were performed to the spectra recovery error of all the five systems with an acceptable confidence level of 0.05. The statistical analysis results showed that there is no means significant difference between the spectra recovery error of any of tested systems.

The similar spectra recovery results in Table 1 and Table 2, together with the strong correlations of spectra recovery error in Table 3 and Table 4 indicated that different simulation systems tested in this research almost have no effect on spectra recovery accuracy. Furthermore, the similar spectra recovery results and the strong correlations of spectra recovery error between the simulation systems and the real system also illustrated that it is feasible to perform representative sample selection based on a simulation system. Therefore, a representative digital camera sensitivity functions is calculated by averaging the spectral sensitivity functions database constructed by Jiang *et al*,[29] and the final representative simulation system is constructed using the representative spectral sensitivity functions, the regularized polynomial model-based spectra recovery method[15], the CIED65 standard illuminant, and the additive system noise as described above.

4.1.2 Results of representative samples selection

With the constructed representative simulation system, the proposed representative samples selection method was tested under the above experimental conditions. It should be noted that since the digital camera Nikon D7200 used in this study is not included in the cameras that used to construct the spectral sensitivity function database in Jiang's research, therefore, the testing of proposed method using Nikon D7200 digital camera can be considered as random verification. Figure 3 shows the real spectra recovery error distribution using different number of selected samples to recover the database. The dashed line in Figure 3 indicates the spectra recovery error when using the database itself as training samples, they are 3.15%, 0.9841% and 1.65 for spectral RMSE, spectral GFC and color difference ΔE_{ab}, respectively.

It is easy to infer from Figure 3(a) that with the number of selected representative samples increase, the TOTAL error defined by RMSE×(1-GFC) is decreased rapidly, and when the selected samples reaches to about 60, the TOTAL error tends to stable. From Figure 3(b) and Figure 3(d), we can see the same changing trend for spectra recovery error of RMSE and for color difference ΔE_{ab}, and when the selected samples reaches to about 60 the spectra recovery error will basically coincides with using the database as training samples. In addition, from Figure 3(c) we can see the values of GFC increase rapidly with the number of selected representative samples increase, after the selected samples reaches to about 40, the spectra recovery accuracy is better than the sample database, and with the continues increase of the selected representative samples the GFC is also tends to stable.

The changing trend of the spectra recovery result in Figure 3 is similar to the previous findings of Mohammadi and Shen *et al.*[7,9] To further verify the performance of the proposed method, in the conditions of select different number of samples including 10, 30, 50, 70, 90, and 110, the proposed method is compared with some of the existing commonly used methods including Hardeberg method, Mohammadi method, Cheung method and Shen method.[6-9] The spectra recovery error of different method are plotted in Figure 4, where 'Total' represents the spectra recovery results that using the database as training samples.

It can be seen from Figure 4 that in terms of RMSE and GFC, the proposed method is apparently outperformed the existing methods in selecting different number of samples, and in terms of ΔE_{ab}, except the group of 30 selected samples, the proposed method is still outperformed the existing methods. During the tested methods, the method proposed by Hardeberg *et al*[6] shows relative lower accuracy overall, the proposed method in this study shows the best accuracy, and the performance of the rest methods are between the Hardeberg method and the proposed method.

Therefore, for the application of digital camera-based spectra recovery based on single RGB image, the proposed representative samples selection method is superior than the existing commonly used methods in general. We believe that the superiority of the proposed method is to benefit from that the representative samples are selected with consideration of the features of the spectra recovery system.

4.1 Discussions

Although the reasonable of the hypothesis on the proposed method as well as its effectiveness in selecting representative samples have been proved, however, there are still two issues need to be further discussed.

4.1.1 Discussion on representative samples number determination

The first issue of the proposed method is how to determine the appropriate number of representative samples, or another way of expression, when the selection of the representative samples should be stopped? As we reported in Figure 3, with the number of selected samples increase, the spectra recovery error of TOTAL, RMSE and color difference ΔE_{ab} are decreased rapidly and tends to stable, and the evaluation metric of GFC showed similar changing trend of the accuracy. Moreover, the spectra recovery error of TOTAL and RMSE showed very similar changing trend. Based on the research experience on spectra recovery, we define to use RMSE as the metric to decide whether to stop sample selection, and stop at the spectra recovery RMSE intersection as indicated in Figure 3(b).

There are three reasons to use RMSE as metric to decide whether to stop the sample selection. The first one is that, to compare with the calculation of color difference, the calculation of RMSE is not affected by factors of light source and observer factors. We know that for the same pair of spectral data, different color differences may be calculated under different light sources and observer conditions, but the RMSE between the them is always fixed. The second reason is that it is not easy to decide when to stop the selection based on GFC, and the TOTAL error showed very similar changing trend to RMSE. The third reason is that the proposed method is tend to select the spectral representative samples for digital camera-based spectra recovery.

In addition, with the proposed representative samples selection method in this paper, the empirical test found that for a database containing a large number of samples, with the number of selected samples increases, the spectra recovery RMSE of using the selected samples and the database itself as training to recover the database will always have a intersection. And after the intersection, with the continues increase of the selected samples, the spectra recovery accuracy of selected samples is no longer significantly improved.

4.1.2 Discussion on robustness of the representative samples

The second issue about the proposed method is the robustness of the selected representative samples compared to the database in practical applications. As the selection of representative samples is based on minimizing the spectra recovery error of the selected samples to the database, therefore, there is no external testing samples during the sample selection progress, and we actually don't know whether the selected optimal samples has the same robustness as the database in recovering other different sample sets.

To explore the robustness of the selected representative samples, based on the constructed representative simulation system, 60 selected representative samples and the database were used as training samples to recover the four color sample sets listed in Table 1. The spectra recovery RMSE and GFC are summarized in Table 5 and Table 6 respectively, where ‘Diff’ is the difference of spectral recovery result between 60 selected representative samples and database, ‘Ave’ represents the average of the test results.

As is reported in Table 5, the maximum spectra recovery RMSE difference between 60 selected representative samples and the sample database is 0.29(%) and the minimum difference is -0.12(%), and the average difference is 0.14(%) for all color sample sets. As is reported in Table 6, the maximum spectra recovery GFC between 60 selected representative samples and the sample database is 0.18(%) and the minimum difference is -0.28(%), and the average difference is

-0.03(%) for all color sample sets. Moreover, the one-way ANOVA and Tukey's post-hoc tests were also performed to the results in Table 5 and Table 6 with an acceptable confidence level of 0.05, and the statistical analysis showed that there is no significant difference between them. It can be seen from the robustness test results in Table 5 and Table 6 that the spectra recovery error of 60 selected representative samples and the database have no significant difference, which proved that 60 selected representative samples have same robustness as the database.

Moreover, the 60 selected representative samples and the database were also used as training samples to recover a simulated mural paintings (seen as Figure 5(a)) that we built in previous studies.[31] The simulated mural was painted using mineral pigments and the painting techniques of the ancient Chinese Dunhuang murals with the help of Dunhuang Academy China. Nineteen points are marked for intuitively observed and compared of the spectra recovery results. Figure 5(b) and Figure 5(c) shows the RMSE map (together with the mean RMSE results at the top of the figure) of the simulated murals that recovered by the 60 selected representative samples and the database.

It can be seen from Figure 5(b) and Figure 5(c) that even there are only few areas was not accurately recovered by the 60 selected representative samples and the database, the total mean RMSE of 3.25% and 3.01% for the 60 selected representative samples and the database illustrated that the simulated mural paintings is well recovered, and there is still no significant difference between their spectra recovery results. The GFC as well as color difference ΔE_{ab} are also reported in Table 7 together with RMSE. It is indicated that there is no significant difference of these metrics between 60 selected representative samples and the database.

To more specifically compare the spectra recovery results between the 60 selected representative samples and the database, the recovered spectral reflectance of nineteen marked points in Figure 5(a) are plotted in Figure 6. It is obviously that the nineteen recovered spectral reflectance by the 60 selected representative samples and the database are not significantly different from each other in general, and they are both consistent with the groundtruth. These results further proved the robustness as well as the effectiveness of the selected representative samples in practical applications. However, there still some points are deviate from the groundtruth in some wavelength range (such as point 6, 14 and 18), this may caused by the spectra recovery method used in this research, or in another case, the database do not include the corresponding training samples can represent the spectral reflectance of the pigments in these points. We will explore the specific reasons on this issue in future research.

5. CONCLUSION

This paper presents a novel method for representative samples selection from a big sample database for spectral characterization of digital camera-based spectra recovery system. The proposed method can help the users to select a very small number of representative samples (less than one tenth in this study) from a large number of samples under the premise of ensuring the spectra recovery accuracy and the robustness of application. This is useful for developing portable training sample chart, which will reduce the workload and improve the work efficiency for digital camera-based spectra recovery. To be specific, the successful of the proposed method is relies on the reasonable hypothesis that it is make sense to construct a simulation digital camera-based spectra recovery system

that can represents the real system. Based on the assumption, a representative simulation system is firstly constructed, and based on the constructed simulation system, the representative samples are selected one-by-one from the sample database with the minimum of the spectra recovery error.

The proposed representative samples selection method was comprehensively tested and discussed through simulation and practical experiment. The outperformance of the proposed method was verified, and the rules to select appropriate number of representative samples were defined. The same effectiveness and robustness of the selected optimal samples as the database encourage us to develop portable sample chart for digital camera-based spectra recovery using in different application fields. In the future research, more tests of the proposed method in different application fields will be carried out. In summary, the results achieved in this study show that the proposed method is a suitable for digital camera-based spectra recovery based on single RGB image, and can help the user to improve the work efficiency and flexibility through developing the portables charts.

ACKNOWLEDGMENTS

This study was supported by the Hubei Province Natural Science Foundation, China(2020CFB386); Team Plan of Scientific and Technological Innovation of Outstanding Youth in Universities of Hubei Province, China (T201807), and National Natural Science Foundation of China (61575174).

CONFLICT OF INTEREST

The author declares no potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the first author, Jinxing Liang, upon reasonable request.

REFERENCES

1. Liang H. Advances in multispectral and hyperspectral imaging for archaeology and art conservation. *Appl. Phys. A* 2012;106(2): 309-323.
2. Amiri, MM, Garcia-Nieto S, Morilla S, MD Fairchild. Spectral Reflectance Reconstruction Using Fuzzy Logic System Training: Color Science Application. *Sensors* 2020;20(17):4726.
3. Raju VB, Sazonov E. Detection of Oil-Containing Dressing on Salad Leaves Using Multispectral Imaging. *IEEE Access* 2020;8: 86196-86206.
4. Kim S, Kim J, Hwang M, Kim M, Youn S, Jang JE, Je M, Dong HL, Lee B, Farkas DL. Smartphone-based multispectral imaging and machine-learning based analysis for discrimination between seborrheic dermatitis and psoriasis on the scalp. *Biomed. Opt. Express* 2019;10(2): 879-891.
5. Kohonen O, Parkkinen J, Jääskeläinen T. Databases for spectral color science. *Color Res. Appl.* 2006;31(5): 381-390.
6. Hardeberg JY. Acquisition and reproduction of color images: colorimetric and multispectral approaches. Universal-Publishers, 2001.
7. Mohammadi M, Nezamabadi M, Berns RS, Taplin, LA. Spectral imaging target development based on hierarchical cluster analysis. In: Color and Imaging Conference. Society for Imaging Science and Technology 2004;2004(1): 59-64.
8. Cheung V, Westland S. Methods for optimal color selection. *J. Imaing Sci. Technol.* 2006;50(5): 481-488.

9. Shen HL, Zhang HG, Xin JH, Shao SJ. Optimal selection of representative colors for spectral reflectance reconstruction in a multispectral imaging system. *Appl. Optics* 2008;47(13): 2494-2502.
10. Eckhard T, Valero EM, Hernández-Andrés J, Schnitzlein M. Adaptive global training set selection for spectral estimation of printed inks using reflectance modeling. *Appl. Optics* 2014;53(4): 709-719.
11. Hernández-Andrés J, Romero J, García-Beltrán A, Nieves JL. Testing linear models on spectral daylight measurements. *Appl. Optics* 1998;37(6): 971-977.
12. Hardeberg JY, Schmitt FJM, Brettel H. Multispectral color image capture using a liquid crystal tunable filter. *Opt. Eng.* 2002;41(10): 2532-2548.
13. Shimano N. Recovery of spectral reflectances of objects being imaged without prior knowledge. *IEEE T. Image Process.* 2006;15(7): 1848-1856.
14. Peyvandi S, Amirshahi SH, Hernandez-Andres J, Nieves JL, Romero J. Generalized inverse-approach model for spectral-signal recovery. *IEEE T. Image Process.* 2012;22(2): 501-510.
15. Li SX. Superiority of optimal broadband filter sets under lower noise levels in multispectral color imaging. *Color Res. Appl.* 2021;46(4): 783-790.
16. Cao B, Liao N, Cheng H. Spectral reflectance reconstruction from RGB images based on weighting smaller color difference group. *Color Res. Appl.* 2017;42(3): 327-332.
17. Heikkinen V. Spectral reflectance estimation using Gaussian processes and combination kernels. *IEEE T. Image Process.* 2018;27(7): 3358-3373.
18. Wu G. Reflectance spectra recovery from a single RGB image by adaptive compressive sensing. *Laser Phys. Lett.* 2019;16(8): 085208.
19. Connah DR, Hardeberg JY. Spectral recovery using polynomial models. In: Color Imaging X: Processing, Hardcopy, and Applications. International Society for Optics and Photonics, 2005;5667: 65-75.
20. Shimano N, Terai K, Hironaga M. Recovery of spectral reflectances of objects being imaged by multispectral cameras. *J. Opt. Soc. Am. A* 2007;24(10): 3211-3219.
21. Lin YT, Finlayson GD. Physically Plausible Spectral Reconstruction. *Sensors* 2020;20(21): 6399.
22. Arad B, Ben-Shahar O, Timofte R. Ntire 2018 challenge on spectral reconstruction from rgb images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018: 929-938.
23. Arad B, Timofte R, Ben-Shahar O, et al. Ntire 2020 challenge on spectral reconstruction from an rgb image[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020: 446-447.
24. Valero EM, Nieves JL, Nascimento SM, Amano K, Foster DH. Recovering spectral data from natural scenes with an RGB digital camera and colored filters. *Color Res. Appl.* 2007;32(5): 352-360.
25. Pelagotti A, Del Mastio A, De Rosa A, Piva A. Multispectral imaging of paintings. *IEEE Signal Proc. Mag.* 2008;25(4): 27-36.
26. Berns RS, Byrns S, Casadio F, Fiedler I, Gallagher C, Imai FH, Newman A, Taplin LA. Rejuvenating the color palette of Georges Seurat's A Sunday on La Grande Jatte-1884: A simulation. *Color Res. Appl.* 2006;31(4): 278-293.
27. Hirai K, Tanimoto T, Yamamoto K, Tominaga S. An LED-based spectral imaging system for surface reflectance and normal estimation. In: International Conference on Signal-Image Technology & Internet-Based Systems. IEEE, 2013: 441-447.
28. Liang J, Wan X, Liu Q, Li C, Li J. Research on filter selection method for broadband spectral imaging system based on ancient murals. *Color Res. Appl.* 2016;41(6): 585-595.
29. Jiang J, Liu D, Gu J, Süsstrunk S. What is the space of spectral sensitivity functions for digital color cameras?. In: IEEE Workshop on Applications of Computer Vision 2013: 168-179.
30. Jiang J, Gu J. Recovering spectral reflectance under commonly available lighting conditions[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2012: 1-8.
31. Liang J, Wan X. Prototype of a pigments color chart for the digital conservation of ancient murals. *J. Electron. Imaging* 2017;26(2): 023013.
32. Liang J, Xiao K, Pointer MR, Wan X, Li C. Spectra estimation from raw camera responses based on adaptive local-weighted linear regression. *Opt. Express* 2019;27(4): 5165-5180.
33. Colorimetry: understanding the CIE system. John Wiley & Sons, 2007.
34. Vilaseca M, Pujol J, Arjona M. Illuminant influence on the reconstruction of near-infrared spectra. *J. Imaing Sci. Technol.* 2004; 48(2): 111-119.
35. Sávoli Z, Kránicz B, Horváth A. Spectral reconstruction in case of different illuminants. *Annals of the Faculty of Engineering Hunedoara*. 2016;14(1): 45.

TABEL 1 The spectra recovery RMSE of 16 test pairs under four simulation systems and the real system

(%)			Systems				
			Simulation				Real
No.	**Training**	**Testing**	**Canon 1D MarkIII**	**Nikon D5100**	**Pentax K5**	**Sony Nex5N**	**Nikon D7200**
1	SGchart	SGchart	1.85	1.87	1.79	1.96	1.81
2	SGchart	DHPchart	3.28	3.40	3.25	3.37	5.35
3	SGchart	Textile	6.40	6.42	6.29	6.45	7.24
4	SGchart	Skinchart	4.19	3.91	3.80	4.05	4.04
5	DHPchart	SGchart	8.25	12.36	8.20	9.12	14.25
6	DHPchart	DHPchart	2.26	2.26	2.03	2.74	1.79
7	DHPchart	Textile	8.98	10.46	8.60	8.84	11.49
8	DHPchart	Skinchart	5.23	5.38	4.87	3.41	4.96
9	Textile	SGchart	12.89	13.02	12.76	12.94	11.45
10	Textile	DHPchart	8.39	8.35	8.16	8.42	7.70
11	Textile	Textile	2.80	2.83	2.73	2.86	3.17
12	Textile	Skinchart	7.56	7.77	7.78	7.66	7.48
13	Skinchart	SGchart	345.46	304.10	287.92	314.51	565.30
14	Skinchart	DHPchart	354.65	310.96	302.27	342.96	515.12
15	Skinchart	Textile	190.57	168.41	149.46	174.78	326.42
16	Skinchart	Skinchart	0.75	0.74	0.71	0.81	0.80

TABEL 2 The spectra recovery GFC of 16 test pairs under four simulation systems and the real system

(%)			Systems				
			Simulation				Real
No.	Training	Testing	Canon 1D MarkIII	Nikon D5100	Pentax K5	Sony Nex5N	Nikon D7200
1	SGchart	SGchart	99.33	99.32	99.38	99.18	99.59
2	SGchart	DHPchart	98.97	98.94	99.01	98.86	98.86
3	SGchart	Textile	93.63	93.54	93.81	93.51	93.62
4	SGchart	Skinchart	99.17	99.26	99.29	99.23	99.27
5	DHPchart	SGchart	95.86	92.69	95.89	95.5	90.38
6	DHPchart	DHPchart	99.29	99.18	99.23	98.83	99.64
7	DHPchart	Textile	88.05	85.13	87.9	87.52	86.29
8	DHPchart	Skinchart	98.79	98.69	98.91	99.4	99.18
9	Textile	SGchart	92.62	92.52	92.76	92.58	90.78
10	Textile	DHPchart	94.78	94.88	95.07	94.83	95.15
11	Textile	Textile	98.29	98.28	98.35	98.25	98.3
12	Textile	Skinchart	97.53	97.53	97.5	97.51	97.02
13	Skinchart	SGchart	61.35	62.57	61.57	61.82	76.74
14	Skinchart	DHPchart	72.75	74.41	73.44	73.24	76.29
15	Skinchart	Textile	67.12	67.19	67.71	67.95	70.5
16	Skinchart	Skinchart	99.96	99.95	99.95	99.95	99.99

TABEL 3 Correlation coefficients of spectra recovery RMSE of different systems in Table 1

		Simulation systems				Real system
		Canon 1D MarkIII	Nikon D5100	Pentax K5	Sony Nex5N	Nikon D7200
Simulation systems	**Canon 1D MarkIII**	-	-	-	-	-
	Nikon D5100	0.9999	-	-	-	-
	Pentax K5	0.9991	0.9989	-	-	-
	Sony Nex5N	0.9989	0.9987	0.9993	-	-
Real system	**Nikon D7200**	0.9943	0.9946	0.9897	0.9886	-

TABEL 4 Correlation coefficients of spectra recovery GFC of different systems in Table 2

		Simulation systems				Real system
		Canon 1D MarkIII	**Nikon D5100**	**Pentax K5**	**Sony Nex5N**	**Nikon D7200**
Simulation systems	**Canon 1D MarkIII**	-	-	-	-	-
	Nikon D5100	0.9911	-	-	-	-
	Pentax K5	0.9998	0.9921	-	-	-
	Sony Nex5N	0.9994	0.9931	0.9996	-	-
Real system	**Nikon D7200**	0.9127	0.9379	0.9106	0.9149	-

TABLE 5 The spectra recovery RMSE of 60 selected representative samples and the database as training samples to recover four color sample sets

(%)	**Testing**				
Training	**SGchart**	**DHPchart**	**Textile**	**Skinchart**	**Ave.**
60 selected samples	5.08	4.05	4.65	3.67	4.36
sample database	4.79	3.97	4.37	3.79	4.23
Diff.	0.29	0.08	0.28	-0.12	0.14

TABLE 6 The spectra recovery GFC of 60 selected representative samples and the database as training samples to recover four color sample sets

(%)	**Testing**				
Training	**SGchart**	**DHPchart**	**Textile**	**Skinchart**	**Ave.**
60 selected samples	97.84	98.16	96.89	99.32	98.05
sample database	97.66	98.23	97.17	99.27	98.08
Diff.	0.18	-0.07	-0.28	0.05	-0.03

TABLE 7 The spectra recovery results of 60 selected representative samples and the database as training samples to recover the simulated mural paintings

	RMSE(%)	**GFC(%)**	**ΔE_{ab}**
60 selected samples	3.25	98.16	2.84
sample database	3.01	98.64	2.58

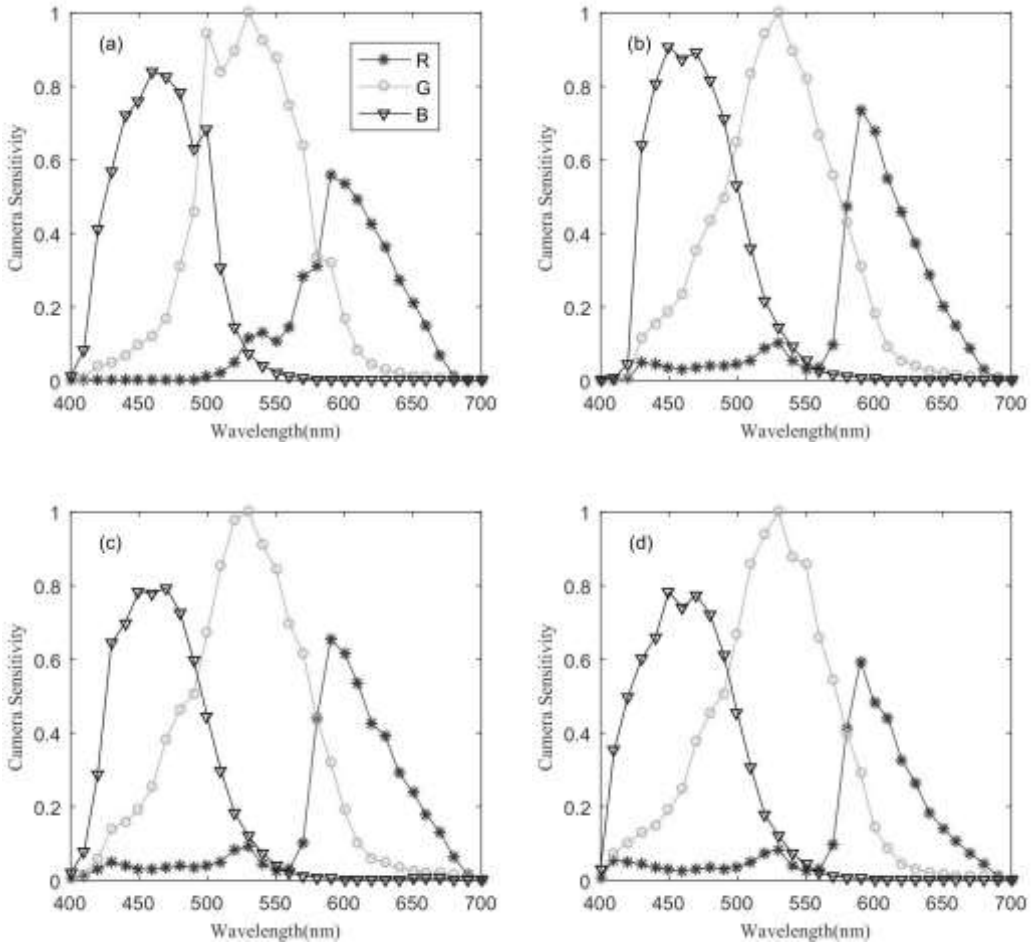

FIGURE 1 The distribution of spectral sensitivity functions of four selected digital cameras: (a) Canon 1D Mark III, (b) Nikon D5100, (c) Pentax K5, (d) Sony Nex5N

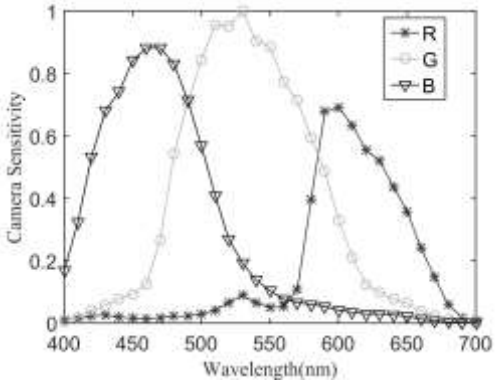

FIGURE 2 The distribution of estimated spectral sensitivity functions of Nikon D7200

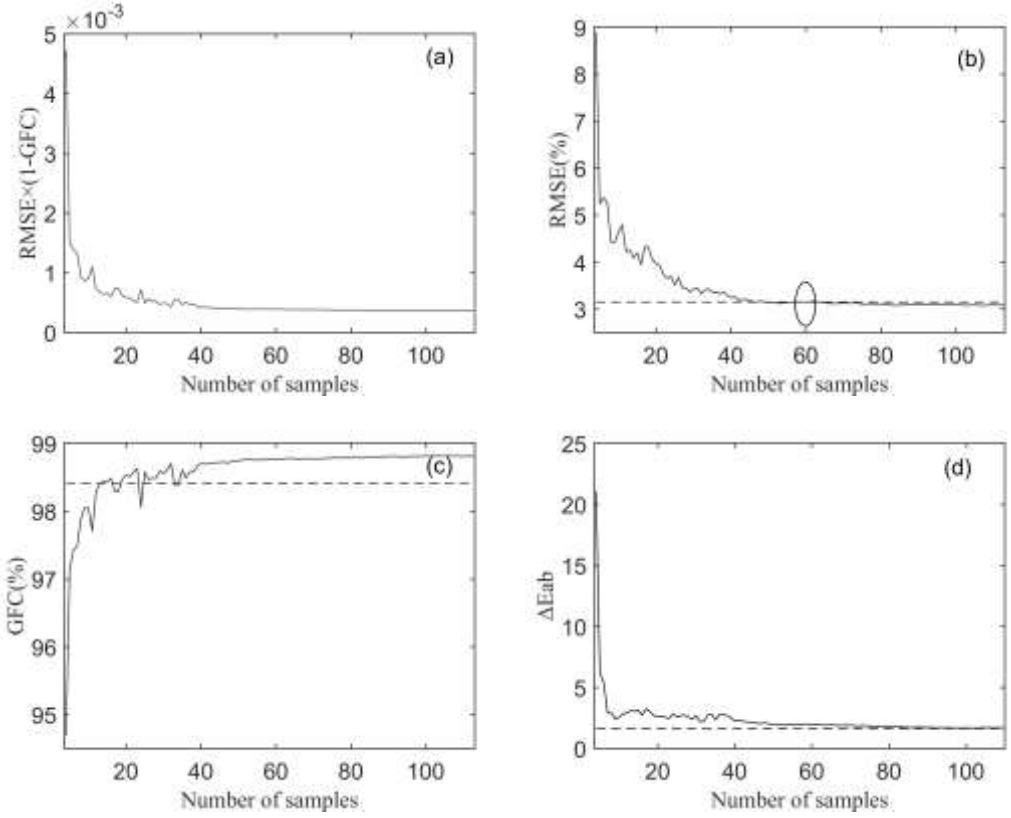

FIGURE 3 The real spectra recovery error distribution using different number of selected samples to recover the database: (a) the distribution of TOTAL error, (b) the distribution of RMSE, (c) the distribution of GFC, (d) the distribution of ΔE_{ab}

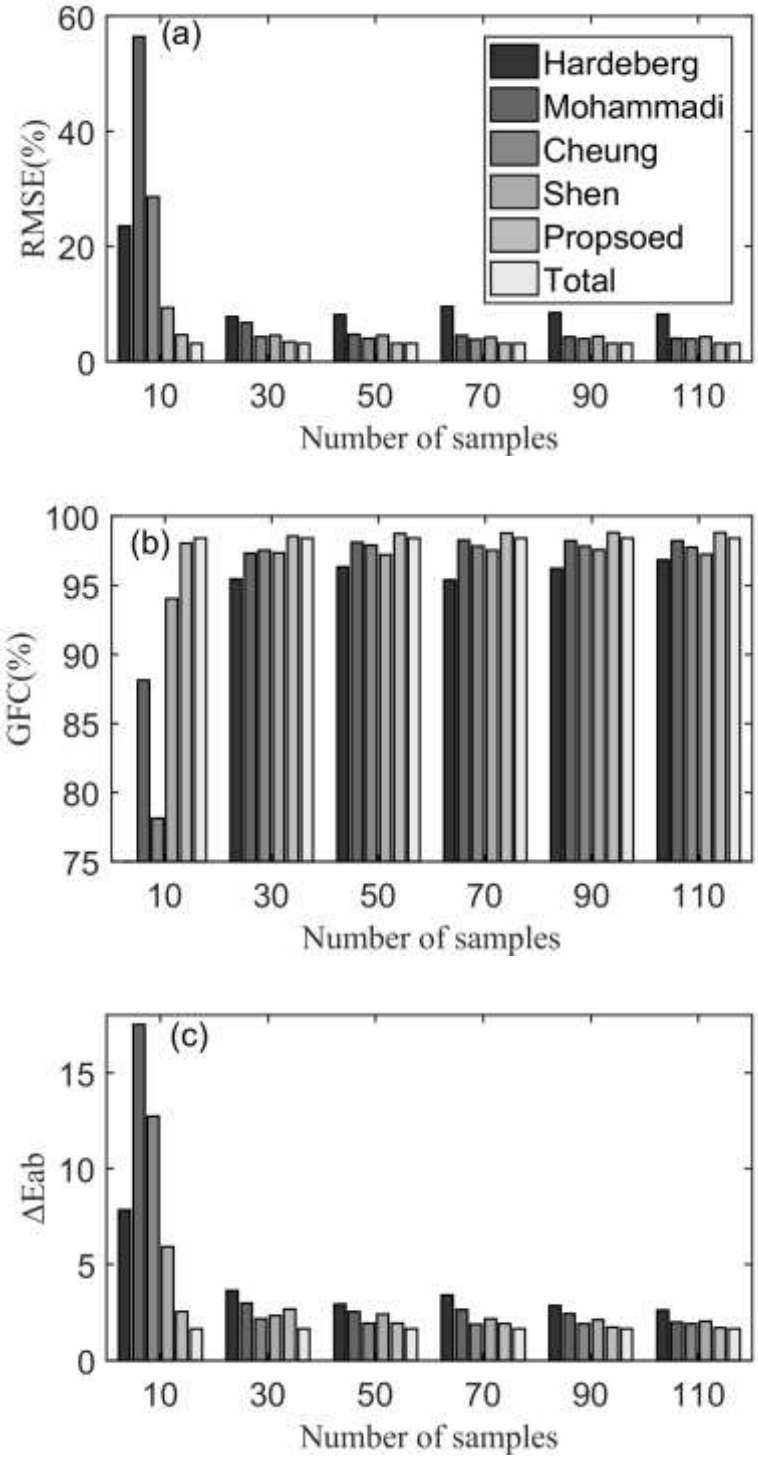

FIGURE 4 The comparison of the real spectra recovery error of different methods in selecting different number of samples: (a) the comparison of RMSE, (b) the comparison of RMSE, (c) the comparison of ΔE_{ab}

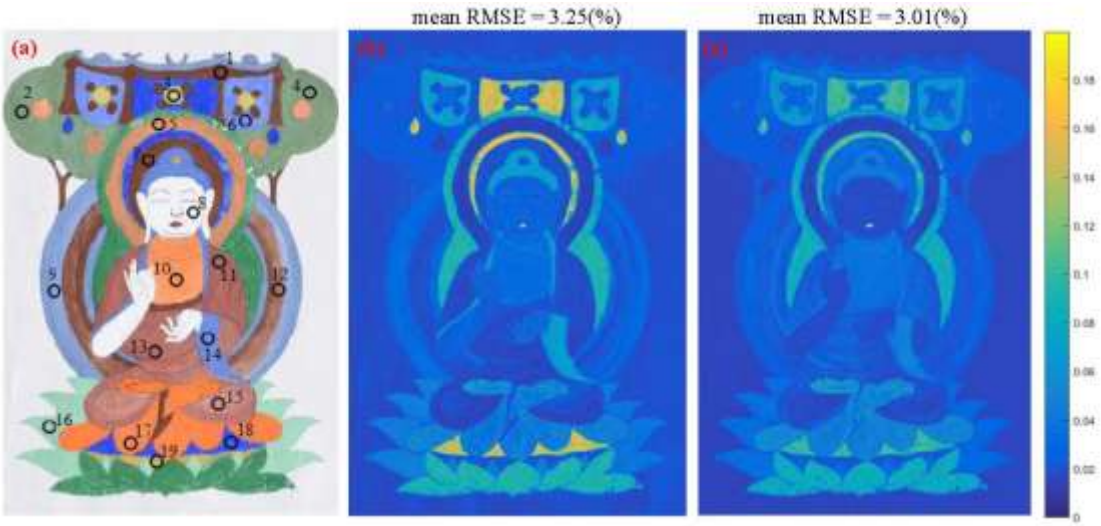

FIGURE 5 The spectra recovery RMSE map and mean RMSE of the simulated murals recovered by the 60 selected representative samples and the database: (a) color rendering of the simulated murals with nineteen marked points, (b) RMSE map recovered by the 60 selected representative samples, (c) RMSE map recovered by the database

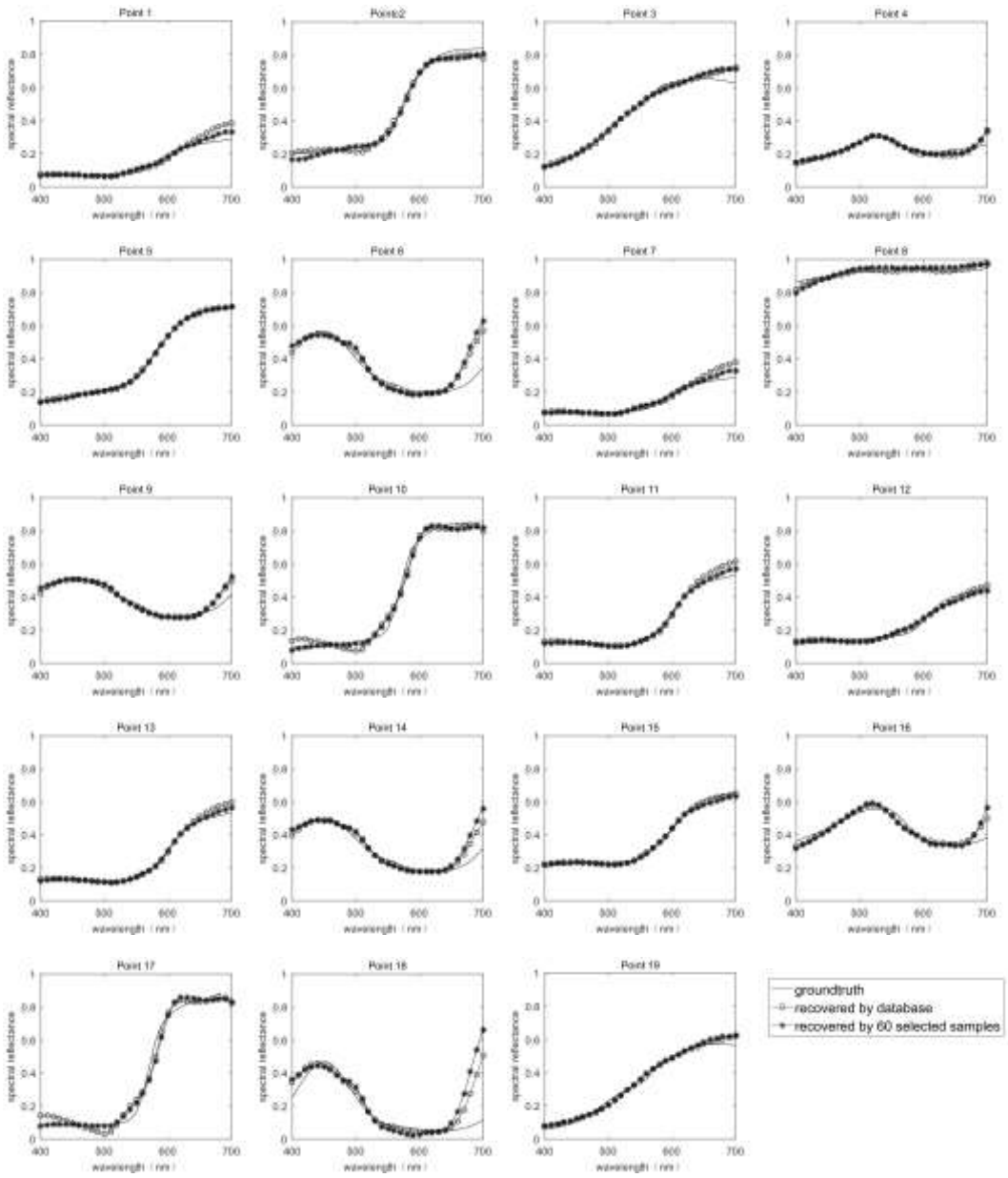

FIGURE 6 The recovered spectral reflectance of nineteen marked points in Figure 5(a) by the 60 selected representative samples and the database

AUTHOR BIOGRAPHIES

Jinxing Liang is currently an academic lecture in Wuhan Textile University. He received a PhD in color and imaging science at the Wuhan University (China) in 2019. His research interests include color science, color management, multispectral imaging and image processing.

Qiang Zhu is currently an academic lecture in Wuhan Textile University. He received a PhD in information technology at the Huazhong Normal University (China) in 2019. His research interests include machine learning and graph neural networks.

Qiang Liu received a PhD degree in computer science from Wuhan University, Wuhan, in 2013. He is an associate professor at the Wuhan University. His research interests include colour quality evaluation for lighting and spectral colour management.

Kaida Xiao is an associate professor in the Colour and Imaging Science in the School of Design, the University of Leeds, United Kingdom. His research interest are related to 3D colour image reproduction, 3D colour printing, 3D printing facial prostheses, medical image capture and analysis, colour appearance modeling and image quality enhancement.