Adaptive global training set selection for spectral estimation of printed inks using reflectance modeling

Timo Eckhard,[1,*] Eva M. Valero,[1] Javier Hernández-Andrés,[1] and Markus Schnitzlein[2]

[1]Optics Department, University of Granada, ES-18071 Granada, Spain

[2]Chromasens GmbH, Max-Stromeyer-Straße 116, D-78467 Konstanz, Germany

*Corresponding author: timo.eckhard@gmx.com

The performance of learning-based spectral estimation is greatly influenced by the set of training samples selected to create the reconstruction model. Training sample selection schemes can be categorized into global and local approaches. Most of the previously proposed global training schemes aim to reduce the number of training samples, or a selection of representative samples, to maintain the generality of the training dataset. This work relates to printed ink reflectance estimation for quality assessment in in-line print inspection. We propose what we believe is a novel global training scheme that models a large population of realistic printable ink reflectances. Based on this dataset, we used a recursive top-down algorithm to reject clusters of training samples that do not enhance the performance of a linear least-square regression (pseudoinverse-based estimation) process. A set of experiments with real camera response data of a 12-channel multispectral camera system illustrate the advantages of this selection scheme over some other state-of-the-art algorithms. For our data, our method of global training sample selection outperforms other methods in terms of estimation quality and, more importantly, can quickly handle large datasets. Furthermore, we show that reflectance modeling is a reasonable, convenient tool to generate large training sets for print inspection applications. © 2014 Optical Society of America

OCIS codes: (150.1708) Color inspection; (110.4234) Multispectral and hyperspectral imaging; (150.3040) Industrial inspection; (100.3190) Inverse problems; (120.5800) Scanners.

http://dx.doi.org/10.1364/AO.53.000709

1. Introduction

Since being introduced to the field of spectral reflectance estimation, learning-based approaches for spectral estimation have gathered a considerable amount of attention and interest in the scientific community. The approaches have been applied in many practical applications, such as spectral estimation of art painting [1] and spectral estimation of human iris spectral reflectances [2]. Learning-based approaches are popular because they are highly adaptive to specific application domains. A main issue in those approaches is the training process, which is used to adapt the system to the statistical structure of the data. It is well known that an adequate selection of training samples is crucial for quality estimations.

Usually, spectral estimation is carried out using low-dimensional sensor response data from a camera as input for the estimation algorithm. In this context, the purpose of training is to establish the relation between camera responses and spectral reflectance data of a given image scene. Unknown reflectances

can then be estimated on a pixel-by-pixel basis from camera responses by using the previously learned relation of the data.

A number of different approaches for training sample selection have been proposed and can be categorized according to several criteria:

- **Global versus local**. In a *global* scheme, the estimation is computed using the same transformation for all test samples, while in the *local* approach the transformation used is different for each specific sample, and therefore is adaptive to the sample features.
- **Bottom-up versus top-down**. A *bottom-up* selection scheme starts with an empty set and successively adds samples to the set. *Top-down* approaches, in contrast, start with a full set and reject samples consecutively until the final training set is obtained.
- **General purpose versus application dependent selection**. Most of the state-of-the-art selection schemes tackle the training sample selection from a rather general perspective, meaning that the selection is not optimized for a specific application but rather to a specific objective, such as to select most distinct colors from a set of available samples. Those methods can therefore be classified as *general purpose*. *Application dependent selection* schemes on the other hand perform the sample selection based on the objective to enhance a specific type of estimation application.

In this work, we compare nine different global training schemes. Our method also belongs to the class of global methods and is computed in top-down fashion in an application-dependent scheme. The other methods compared to our approach are for general purpose training sample selection.

Very recently, spectral estimation has been used for in-line print inspection for colorimetric quality assessment of printed inks to monitor the colorimetric and spectral quality of printed inks during the production process [3]. In traditional approaches, and even in most current state-of-the-art systems, print quality inspection is performed off-line, meaning that the quality of printing is quantified on sample sheets of printed media [4]. A drawback of such method is the lack of full quality control over all printed material produced, since the reference measurements are only based on individual samples, extracted from the printing process. Apart from that, the measurements are often carried out in time consuming manual fashion and can not be performed in real time.

Dealing with printed ink samples allows us to introduce an alternative approach for gathering the initial pool of samples from which the training set for a spectral estimation system is selected. Rather than using real training data, we model a large dataset of training reflectances using an empirical printer characterization model that has the same CMYK values as input as a four-ink CMYK printer

has. Using, for instance, a step 10 sampling in CMYK space for modeling leads to roughly 14600 reflectance samples that span the gamut of printable reflectances. Further, we propose a novel recursive algorithm to reduce the modeled training data to a set of samples that is optimal for the spectral estimation task.

The algorithm is clustering based and works in top-down fashion by rejecting clusters of samples that are not beneficial for the spectral estimation process. The cluster based rejection makes the algorithm fast on large datasets, compared to most other approaches like [5–9] that select samples one-by-one iteratively. The top-down processing and rejection of samples also allows controlling the decrease in estimation quality for a specific application, an important feature that has not been considered in previous studies.

Our approach overcomes some of the limitations in traditional training of spectral estimation systems. In such systems, training is commonly performed based on datasets consisting of test charts (e.g., Macbeth ColorChecker or IT8.6 test chart) or standard color atlases [e.g., Munsell book of colors or natural color system (NCS)]. One reason for their popularity is the availability of instances of physical samples and that for most of them, spectral data has actually been measured and made publicly available, for instance in [10]. However, using standard datasets does not guarantee a good basis for the selection of training samples, as those samples might not be general enough for a given specific application, or simply not span the spectral or colorimetric data space. By using printer characterization to model reflectance data, samples spanning the whole gamut of printable colors can be generated in a time-saving manner. Furthermore, reflectances of printed inks depend on a combination of several factors such as ink type, printing substrate, and the printing technology. The empirical modeling allows an accurate prediction of reflectances that are specific to the ink, substrate and printer used, whereas standard charts are more general in nature.

The remainder of this article is structured as follows: In Section 2, we introduce the spectral estimation application, the ink reflectance modeling scheme, state-of-the-art training sample selection schemes and in Section 3 our proposed algorithm for training sample selection. Section 4 describes the experimental configuration, experiments, and results of this work are presented in Section 5. Finally, in Section 6 we discuss conclusions of the results and future work.

2. Background

A. Notation

In what follows, a variable x describes a scalar value. Denoted in bold, $\mathbf{x}$ refers to a vector and a matrix $\mathbf{X}$ is indicated in bold capital letters. $\mathbf{X}^+$ refers to the Moore–Penrose pseudoinverse (PI) of matrix $\mathbf{X}$.

B. Spectral Estimation for In-line Print Inspection

As mentioned previously, in-line print inspection aims to provide full-resolution quality assessment in real-time during the printing process. In-line print inspection is mainly applied in offset printing technology. In the experiments described here, a Chromasens GmbH prototype inspection system is used, and is set up by a 12-channel multispectral line-scan camera of type truePIXA and a Chromasens Corona II-D50 daylight LED illumination panel. The illumination system consists of four types of high-power LEDs corresponding to warm-white, cold-white, 400 nm peak-wavelength and 470 nm peak-wavelength. The different LED types allowed us to optimize the color rendering index of the illumination.

The setting used for this work corresponds to a color-rendering index of approximately 90. The camera's 12 channels are created by placing four lenses in front of the cameras RGB sensor and thereby dividing the line into four parts. Each of the four lenses is coupled with a distinct color filter and provides R, G, and B camera responses in the spatial area of the common field-of-view of all lenses. Figures 1 and 2, respectively, show a schematic illustration of the camera system and its responsivities multiplied by the LED illumination spectrum.

Each acquisition of the camera produces only one horizontal line of pixels, whereas the vertical dimension of an image is formed by moving the printed samples under the camera and therefore scanning it. In offset printing, movement of the substrate is part of the process and line-scan cameras are therefore highly applicable for continuous data acquisition. For this work, camera responses have been acquired on a laboratory prototype that is not operated on a printing machine yet; instead a scanning table is used to move the sample while image acquisition takes place.

We previously have investigated the task of spectral estimation of printed ink reflectances for print quality assessment [3]. In that work, several approaches for estimation were compared. The best

estimation results were obtained for the Kernel method [11]. In this work we are only investigating linear least-square regression (PI, estimation method) [12] for the sake of its simplicity. PI is fast compared to other approaches and does not need any parametrization.

The PI method is based on a linear least-square solution of a mapping from 12-dimensional camera response space to the 71-dimensional spectral reflectance space (discrete spectral data in the range from 380 to 730 nm in 5 nm steps), obtained from a set of training samples. Formally, the $71 \times v$ matrix of v estimated reflectances $\mathbf{R_{te}}$ that relates to the $12 \times v$ matrix of test camera responses $\mathbf{P_{te}}$ is computed from a $71 \times u$ matrix of training reflectances $\mathbf{R_{tr}}$ and a $12 \times u$ matrix of training camera response $\mathbf{P_{tr}}$ as

$$\mathbf{R_{te}} = \mathbf{R_{tr}} \times \mathbf{P_{tr}^+} \times \mathbf{P_{te}}. \tag{1}$$

The colorimetric quality of spectral estimation for measured spectral data compared to estimated spectral data is evaluated by CIEDE2000 color difference, which will be referred to as ΔE_{00}, [13]. Calculations were performed assuming D65 illumination and the CIE 1931 2° standard observer. Spectral quality is evaluated by the root mean square error (RMSE), defined for an estimated spectrum $\tilde{\mathbf{r}}$ and its measured counterpart $\mathbf{r}$ as

$$\text{RMSE} = \sqrt{\frac{1}{m}|\mathbf{r} - \tilde{\mathbf{r}}|^2}. \tag{2}$$

Further, d_p is the Pearson distance, which also is known as the complemented GFC (goodness of fit coefficient) [14,15] and defined as

$$d_p = 1 - \frac{\langle \mathbf{r}, \tilde{\mathbf{r}} \rangle}{\|\mathbf{r}\| \cdot \|\tilde{\mathbf{r}}\|}. \tag{3}$$

C. Reflectance Modeling

Building models that predict printed reflectances have been investigated for more than 50 years. From
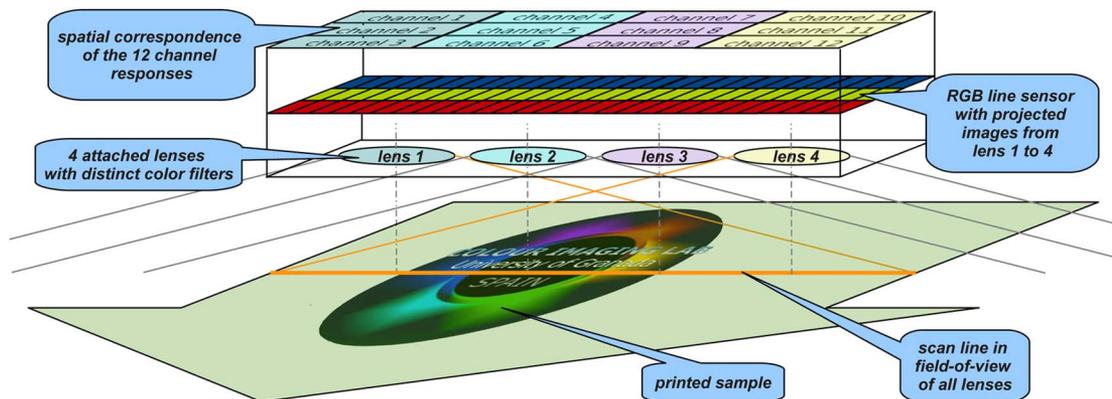
Fig. 1. Schematic illustration of the working principle of the 12 channel multispectral camera system. The orange line indicates the field of view of all lenses of the camera system. To acquire an image scene (illustrated by the multicolor logo of the Color Imaging Laboratory at the University of Granada), the image is translated horizontally under the camera in the direction indicated by the green arrow.
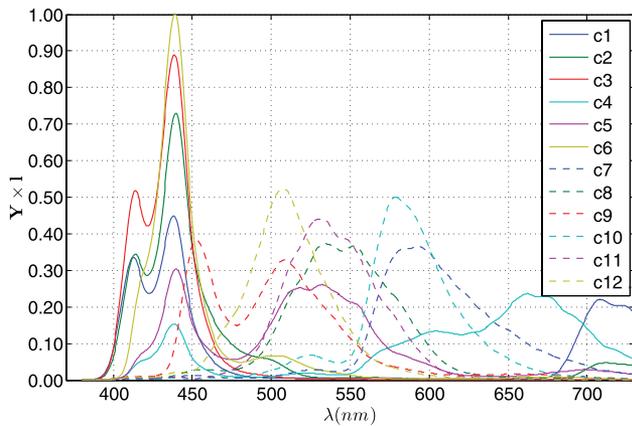
Fig. 2. Normalized product of responsivities **Y** and spectral power distribution of the illumination **l** of the 12-channel multi-spectral camera system.

a historical perspective, modeling evolved over time by expanding simple models with additional parameters taking into account the optical and physical phenomena of the interaction of light, ink and print substrate. A basic overview of reflectance modeling in the printing process is given in [16].

For this work, reflectances were modeled using the enhanced Yule–Nielsen spectral Neugebauer model [17]. For the case of a four-ink printer, as considered here, the inputs of the model are CMYK values, specifying the amount of each ink type per pixel in an image. This model belongs to the class of empirical models, where dot-gain and ink-spreading behavior of the printing process are obtained empirically from measured reflectances of specific printed samples and their CMYK counterparts.

In our experiments, we obtained a mean model accuracy of $\Delta E_{00} = 1.85$ and $\Delta E_{94} = 1.95$ (CIE-Lab CIE94 units) based on data from an Océ ColorWave 600PP inkjet printer on Océ Red Label Paper with a printing resolution of 1200 dpi. The evaluation is based on data from a printed ECI 2000R test chart. Hersch and Crété [17] reported a model accuracy of $\Delta E_{94} = 0.9$ for ink-jet printing technology and a printing resolution of 100 lines per inch for the same printer model. We believe the deviation in model accuracy from our results to be due to the differences in substrate and ink type. The accuracy obtained for our model will suffice for our particular task in the following sections. However, we believe improvements in model accuracy will further positively influence our results.

D. Global Training Sample Selection Schemes

As mentioned in Section 1, we have compared our approach with several state-of-the-art global training set selection methods. In this section we present a complete description of all of them, including details on the implementation. The methods are described using the following notation: the set of m available training samples is $\mathbb{S} = \{x_1, x_2, ..., x_m\}$. The set of n selected training samples is denoted

$\mathbb{S}_{tr} = \{y_1, y_2, ..., y_n\}$ and the set of samples not included in the training is $\overline{\mathbb{S}}_{tr} = \mathbb{S} \backslash \mathbb{S}_{tr}$. Further, $|\mathbb{S}|$ refers to the cardinality of $\mathbb{S}$.

Here are the global training set selection methods to which we compared our approach:

Random selection (RD). The simplest design of a training set is *random selection*, where the objective of the method is a reduction of the number of samples in the training set. In several experiments related to machine learning in different application domains RD was shown to be a rather bad choice [5–7].

Kennard–Stone (KS) design—1996. This method aims at sequentially selecting n samples that are uniformly spaced over the sample space. The ith sample ($i = 1...n$) is selected as $y_i = \arg_v \max\{D(v)|v \in \overline{\mathbb{S}}_{tr}\}$, where $D(v)$ is the minimal distance between v and any point in $\mathbb{S}_{tr}$, calculated as $D(v) = \min\{d(u,v)|u \in \mathbb{S}_{tr}\}$ and $d(u,v)$, is defined as the Euclidean distance between element u and v [5]. The *Kennard–Stone design* was originally proposed in the field of *design of experiments*. We have followed the implementation of Wu *et al.* [5], which was applied to the training of a neural network in a classification task.

Hardeberg (HD) method—1999. Based on the objective that the selected samples should be most distinct from each other (i.e., contain the least amount of mutual information), Hardeberg [8] proposed a bottom-up iterative method based on the criterion of minimum condition number.

The first sample is selected as the one with maximal variance in spectral space among $\mathbb{S}$. The ith sample ($i = 2...n$) is selected as $y_i = \arg_v \max\{\text{cond}(\mathbb{S}_c(v))|v \in \overline{\mathbb{S}}_{tr}\}$, where $\text{cond}(\mathbb{S}_c(v))$ denotes the condition number of the $i \times d$ matrix of i elements with dimensionality d in the set $\mathbb{S}_c = \mathbb{S}_{tr} \cup \{v\}$.

Kang (KG) method—2004. Another method of training sample selection that essentially aims at a simple reduction of training samples was proposed by Kang *et al.* [7]. They used *k-means* to cluster $\mathbb{S}$ into n clusters and for each cluster the sample with the smallest Euclidean distance to the cluster centroid was added to $\mathbb{S}_{tr}$.

We additionally propose a modification of Kang's method that involves substituting the Euclidean distance measurement in *k-means* by the *spectral similarity value* (SSV) distance metric. The SSV metric considers the magnitude as well as scale differences between two spectra, whereas Euclidean distance primarily measures magnitude differences [18]. It has been shown that clustering based on the SSV metric can outperform clustering based on the Euclidean distance metric in some image segmentation tasks [19].

Mohammadi (MH) method—2004. Mohammadi *et al.* [20] cluster $\mathbb{S}$ into n clusters using agglomerative hierarchical clustering with average linkage and correlation distance to measure dissimilarity of reflectances. For each cluster, the reflectance sample having the minimum average vectorial angle to any other reflectance in the set is selected for training.

Cheung–Westland (CW) method—2006. Cheung and Westland [6] proposed several objective functions for an iterative approach that is based on the assumption that representative colors to be selected for training should be most distinct from each other. We adapt the MAXSUMS objective function, which was later tested by Shen *et al.* [9] and found to perform better than the alternative proposals of Cheung and Westland.

Like in HD, the first sample is selected as the one with maximal variance in spectral space among $\mathbb{S}$. The ith sample ($i = 2\ldots n$) is selected as $y_i = \arg_v \max\{D(v)|v \in \overline{\mathbb{S}}_{\mathrm{tr}}\}$, where $D(v) = \sum_{u \in \mathbb{S}_{\mathrm{tr}}} \sqrt{d(u,v)}$ and $d(u,v)$ is defined as the Euclidean distance between element u and v. The square-root term is introduced to penalize small spectral differences.

Shen (SH) method—2008. Shen *et al.* [9] proposed two sample selection schemes for spectral characterization. One is referred to as a *virtual-imaging-based* approach, where the aim is to minimize the error between measured reflectances in the training set and corresponding reflectances estimated from camera responses of a virtual imaging system. The other approach is an iterative method in which the selected samples should produce minimum spectral RMSE for spectral characterization if the selected samples can optimally represent the whole set. The latter approach is compared in this work.

Again, the first sample is selected as the one with maximal variance in spectral space among $\mathbb{S}$. The ith sample ($i = 2\ldots n$) is selected as $y_i = \arg_v \min\{D(\mathbb{S}_c(v), \mathbb{S}'_c(v))|v \in \overline{\mathbb{S}}_{\mathrm{tr}}\}$. The operator $D(\mathbb{S}_c(v), \mathbb{S}'_c(v))$ denotes the Frobenius norm of the residuals between reflectances in $\mathbb{S}_c(v)$ and reconstructed reflectances in $\mathbb{S}_c(v)'$, according to eigenvector decomposition. The reconstruction of set $\mathbb{S}_c(v)'$ is computed from 15 basis vectors of $\mathbb{S}_c(v)$.

In case of KS, KG, MH, and SSV, the design of the algorithms allow input data to be spectral reflectances or camera responses. Using response data instead of reflectance data reduces the amount of processing. We investigated the influence of either selection on the quality of training sample selection.

Most approaches are iterative in nature. The computational complexity of the algorithms vary greatly with the type of computations that are performed in each iteration. However, increasing the initial pool of samples for selection increases the computation time mostly because computations with more data are to be performed within each iteration. Our proposed algorithm does not process the data iteratively. The difference in computational time becomes apparent for large datasets.

3. Adaptive Global Training Sample Selection Scheme: The Recursive Rejection Method

Rather than selecting training samples without prior knowledge of the estimation application and following general premises to select most representative colors like in the methods introduced in Subsection 2.D, the recursive rejection (RR) method proposed in this article, optimizes a training set $\mathbb{S}_{\mathrm{tr}}$ for a certain application by measuring the change in performance in spectral estimation when certain clusters of samples are rejected from the training. Unlike other approaches, the aim of RR is to reduce the number of samples contained in an initially large pool of samples while maintaining or improving estimation quality resulting from estimation with the reduced set.

The method works in a top-down fashion, starting by considering the full set of all available training samples $\mathbb{S}_{\mathrm{tr,init}} = \mathbb{S}$. Then, *k-means* clustering [21] is used to cluster $\mathbb{S}_{\mathrm{tr,init}}$ into k clusters. The importance of each cluster of samples $\mathbb{S}_{ci}$ for estimation is evaluated. This is done by comparing the error e_a obtained from estimation using the whole set of available training samples $\mathbb{S}$ with the estimation error e_r obtained from estimation using a reduced set $\mathbb{S}_{\mathrm{tr}} \backslash \mathbb{S}_{ci}$ for training. For evaluation, an application-dependent dataset $\mathbb{S}_{\mathrm{app}}$ is used. If the estimation error drops more than a predefined threshold th, the cluster of samples $\mathbb{S}_{ci}$ is considered to be important and further processed. Other clusters are rejected from the training set. Clusters of samples that are not rejected are processed in a recursive call of the algorithm with an updated training set to evaluate if they consist of subclusters that could be rejected according to the same criterion as outlined above. This procedure is repeated until either all clusters are processed or the number of samples in the remaining set $\mathbb{S}$ is less or equal to the number of clusters k.

Why are clustering techniques used rather than dividing the training sample space into quantized bins and performing the rejection scheme based on the members of these bins? In fact, performing quantization on high-dimensional data, such as reflectance or camera response data, is often performed by clustering techniques [22,23]. The simple and fast k-means method as such is a member of vector quantization algorithms.

For this work, spectral estimation is performed by the PI method (see Subsection 2.B) and the RMSE is used by the algorithm as a measure of estimation quality. $\mathbb{S}_{\mathrm{tr,init}}$ is intended to be a large set of available training samples. In our work, $\mathbb{S}_{\mathrm{tr,init}}$ is obtained by modeling reflectances from CMYK values using a printer characterization algorithm (see Subsection 2.C). The application dependent set $\mathbb{S}_{\mathrm{app}}$ on the other hand has to represent well real data from the specific application (in our case, $\mathbb{S}_{\mathrm{app}}$ is ideally selected as a representative set of printed samples from the in-line printing process for which print quality is going to be evaluated).

In contradiction to any other previously proposed method, the number of samples in the final training set is not fixed in advance. This characteristic ensures that the final set of selected samples is optimal in terms of estimation quality. However, we can influence the size of the output training set

```
 1: function RR($\mathbb{S}, \mathbb{S}_{app}, e_a, k, th$)
 2:    if $|\mathbb{S}| \leq k$ then
 3:       return $\mathbb{S}$
 4:    else
 5:       $\mathbb{S}_c \leftarrow \mathbb{S}$
 6:       $[\mathbb{S}_{c1}, ..., \mathbb{S}_{ck}] \leftarrow kmeans(\mathbb{S}_c, k)$
 7:       for $i = 1 \rightarrow k$ do
 8:          if $isImportant(\mathbb{S}_c - \mathbb{S}_{ci}, \mathbb{S}_{app}, e_a, th)$ then
 9:             $\mathbb{S}_{ir} \leftarrow RR(\mathbb{S}_{ci}, \mathbb{S}_{app}, e_a, k, th)$
10:             $\mathbb{S}_c \leftarrow \mathbb{S} - \mathbb{S}_{ci} + \mathbb{S}_{ir}$
11:          else
12:             $\mathbb{S}_c \leftarrow \mathbb{S} - \mathbb{S}_{ci}$
13:          end if
14:       end for
15:       return $\mathbb{S}_c$
16:    end if
17: end function

18: function ISIMPORTANT($\mathbb{S}_{tr}, \mathbb{S}_{app}, e_a, th$)
19:    $e_r \leftarrow PI(\mathbb{S}_{tr}, \mathbb{S}_{app})$     ▷ obtain the quality of PI
       estimation with $\mathbb{S}_{tr}$ for training and $\mathbb{S}_{app}$ for testing
20:    if $(e_r - e_a) > th$ then
21:       return $true$
22:    else
23:       return $false$
24:    end if
25: end function
```

Fig. 3. Pseudocode of the RR algorithm proposed in this work.

to some extend by varying the threshold parameter th, and therefore trade off quality of estimation with training set size.

Figure 3 describes the proposed method in pseudocode. Source code of the algorithm for Matlab by MathWorks is available on [24].

4. Experimental Configurations

Here, we compare the global training sample selection schemes introduced in Subsection 2.D with our proposed approach and demonstrate the functionality of the latter in two different experiments. We found that several training sample selection algorithms are not practical for large datasets due to time-intensive computations. Our proposed global training approach that involves a large set of modeled reflectances is therefore not comparable to all methods introduced in Subsection 2.D. However, the proposed RR algorithm also can be operated with a small dataset for training sample selection, which allows comparison to all the other approaches. Experiment 1 does this comparison.

The small dataset consists of 503 sample color patches printed with ink from the Toyo Ink Group on super-calendered paper. Real camera responses of the patches were obtained with the previously introduced prototype system of Chromasens GmbH (see Subsection 2.B). Reflectance data of printed samples were measured with a X-Rite SpectroEye spectrophotometer in the spectral range between 380 and 730 nm. To evaluate Experiment 1 with data that is independent from the training data, we used a set of 140 samples from a X-Rite color checker digital test chart (CC140). Camera responses and spectral

data were acquired in the same conditions as the Toyo dataset. In summary, data for training ($\mathbb{S}_{tr}$) is selected from the Toyo set, and the estimation is tested on the CC140 dataset ($\mathbb{S}_{te}$). For RR, 30% of the test data were reserved as application dataset ($\mathbb{S}_{app}$).

Experiment 2 incorporates the modeled dataset for training sample selection, but only for a constraint set of methods applicable to large datasets (RD, KG, KS, MH, and RR). This dataset consists of 14641 modeled ink reflectances, generated from CMYK values sampled in steps of 10 from 0 to 100 digital counts for each ink type and all combinations of inks (see Subsection 2.C). The corresponding camera responses for the modeled dataset are simulated. Using simulated camera responses of this dataset in the RR method rather than real data is a fundamental characteristic, because it is undesirable to print and measure the huge amount of samples of modeled reflectances to acquire real camera responses. Note that simulated data is used only within the training sample selection and for training. The evaluation of estimation quality of the RR method and its training sample selection are, of course, based on real data. The simulated responses are computed with signal-dependent Gaussian noise according to Eq. (4):

$$\mathbf{P} = \Delta\lambda * \mathbf{Y}' \times \text{diag}(\mathbf{l}) \times \mathbf{R} + \mathbf{b}, \qquad (4)$$

where $\mathbf{P}$ refers to a $c \times n$ matrix of n c-dimensional samples of camera responses. $\mathbf{Y}$ is a $w \times c$ matrix of spectral responsivities of the camera system and $\mathbf{l}$ is the w-dimensional spectral power distribution of the illumination of the image scene, $\mathbf{R}$ a matrix of $w \times n$ spectral reflectances, and $\Delta\lambda$ refers to the sampling interval of the discrete spectral data, which in our case is $\Delta\lambda = 5$ nm.

The noise variance of the signal dependent additive noise term $\mathbf{b}$ was adjusted so that a mean signal-to-noise ratio (SNR) of 54 dB was obtained. This value is close to the SNR computed from real and noiseless simulated camera responses for this system. The term *signal dependent* here means that higher camera response values are assigned with a larger contribution of noise (this behavior models shot noise). The signal independent contribution of noise to real camera responses (e.g., thermal noise) is accounted for in this system by dark signal subtraction and therefore not considered in the simulation.

As application and test dataset in Experiment 2, Toyo was used. A total of 70% were reserved for testing ($\mathbb{S}_{te}$) and 30% for the RR method as application dataset ($\mathbb{S}_{app}$). Figure 4 illustrates CIE-Lab coordinates of Toyo, *CC140* and the modeled set. As several training sample selection methods depend on the initialization of the random number generator of the PC, we repeated each experiment three times and present averaged results to avoid a bias. Computations for this work were performed on a 64-bit system
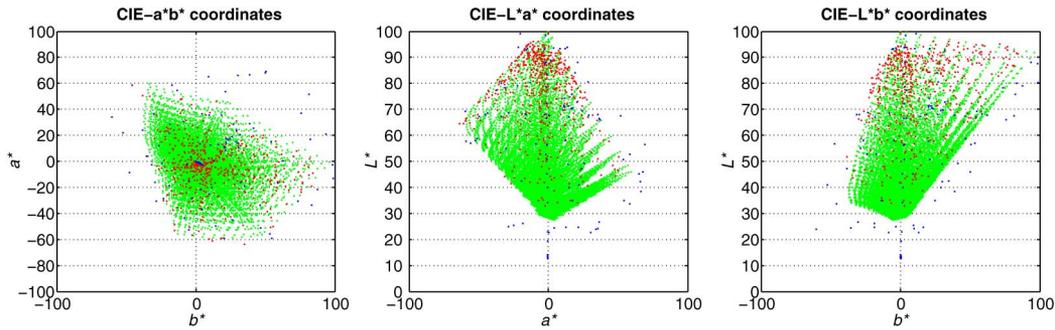
Fig. 4. CIE-L*a*b* color coordinates of modeled (green), Toyo (red), and CC140 (blue) dataset.

with a Intel Core(TM) i5 CPU and 4GB RAM and all algorithms were implemented in Matlab.

5. Experiments and Results

A. Experiment 1: Comparison of Training Sample Selection Methods Using a Small Pool of Selectable Training Data

The RR method proposed in this work does not have the number of desired training samples as input parameter (see Section 3). More precisely, training sample reduction is performed on the premise to enhance spectral estimation quality regardless of the number of training samples. However, other state-of-the-art algorithms compared in this work have the number of training samples as input parameter. To allow a fair comparison, each method should be operated with the number of optimal training samples that minimize the estimation error. We seek this number for each algorithm with an exhaustive search in the range of 10 to 503 selected samples, in steps of 10 samples. In practical applications, an application dataset similar to that of RR ($\mathbb{S}_{app}$) would have to be used to find the optimal n.

Figures 5 and 6 illustrate the results of this first comparison in terms of ΔE_{00} mean estimation error over n. The design of KG, KS, MH, and SSV allows the use of reflectance or camera responses as input
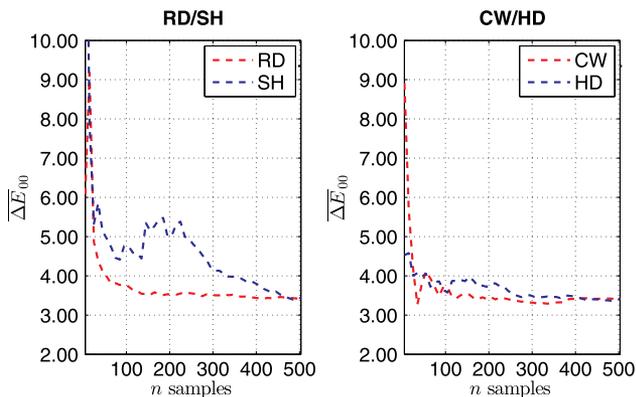
Fig. 5. Mean colorimetric estimation performance ($\overline{\Delta E}_{00}$) over a number of training samples n for RD, SH, CW, and HD method and the small dataset.

data for the training sample selection. For applicable methods, we compared both options.

An important finding in this experiment is that the performance of most algorithms generally improves (despite some variation) with larger number of selected samples. Therefore, a clear number of optimal samples can not necessarily be identified for this dataset. The gradient indicating the increase in estimation quality is large for few samples (up to 100) and gets lower for higher number of samples.

This finding supports the development intention of most algorithms that aim in incrementally selecting most distinct samples: The samples selected at the beginning influence the estimation drastically, while adding more samples in further iterations produces progressively less change in estimation quality. This is in accordance with [25], where the optimal number of samples for training in a spectral reconstruction task are found by comparing the variance in quality of the reconstruction that can be achieved from equal-sized training sets containing different training samples.

Also, one might expect to find a monotonically decreasing error value for larger number of training samples. This is not necessarily the case because the sample selection criteria in the methods compared here is never to optimize the ΔE_{00} error for the least-square spectral estimation illustrated in Fig. 5. Even in the case of SH, where spectral reconstruction is considered in the selection scheme (see Subsection 2.D), the reconstruction algorithm (eigenvector based reconstruction) as well as the metric (Frobenius norm) for measuring reconstruction quality differ. For this method, for instance, the $\overline{\Delta E}_{00}$ error shows large variation when less than 300 samples are used for training.

The comparison of reflectance versus camera response space as input data for KG, KS, MH, and SSV shows that for KS and MH, using reflectances results in a lower mean estimation error for few selected samples, as shown in Fig. 6. The more samples selected, the weaker this effect. In case of KG and SSV, no clear difference can be observed.

Having SSV operated with reflectances as input to the algorithm outperforms KG for a small n. This indicates that using magnitude and shape information in the computation of distances between data points
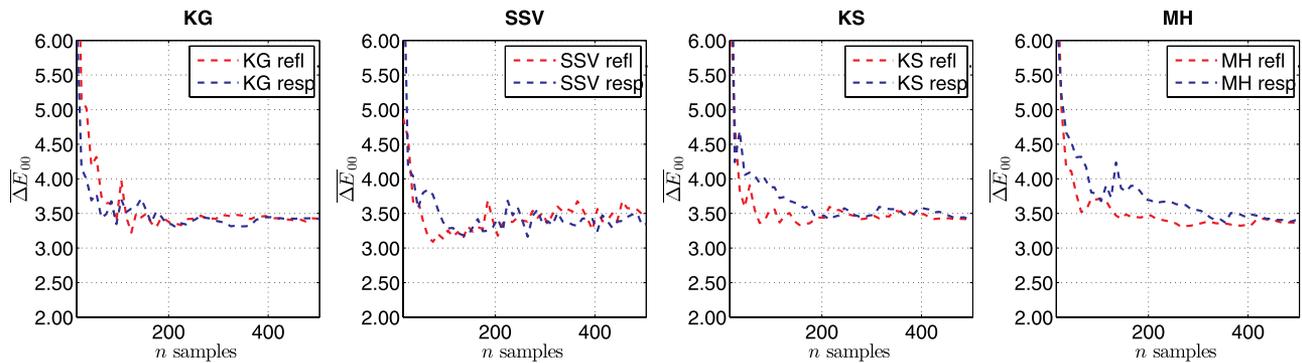
Fig. 6. Mean colorimetric estimation performance ($\overline{\Delta E_{00}}$) over number of training samples n for KG, SSV, KS, and MH and the small dataset. For most methods, the optimal number of samples for training approaches the maximal number of samples.

(in case of SSV) improves the initial KG method, which uses only magnitude information.

In Table 1, the mean estimation error $\overline{\Delta E_{00}}$ for each method at its point of optimal performance is presented and the results for RR are introduced. RR outperforms the other methods significantly in terms of colorimetric error. In terms of spectral error, only CW performs better. We also observed that the number of selected samples in our method is considerably low, compared to most of the other methods.

Regarding the overall spectral and colorimetric estimation performance achieved in this experiment for print quality inspection, a colorimetric error larger than approximately 1 ΔE_{00} is considered high for some applications. The reason for the low performance measured here is that datasets for training and testing are from different populations (the Toyo set consists of printed ink samples, whereas the *CC140* set are painted samples). Figure 4 also shows that several samples of the CC140 test set are out of the gamut of available training samples in the Toyo set. However, this experiment only provides comparative results of global training sample selection schemes and does not offer insight on the performance of the multispectral in-line print inspection system. The results of Experiment 2 will help us draw those conclusions.

Table 1. Experiment 1 Compares Mean Estimation Error for the Method-Depending Optimal n_{opt} Selected Training Samples and the Small Dataset[a]

Method	n_{opt}	$\overline{\Delta E_{00}}$	$\overline{d_p}$	$\overline{RMSE}$
RD	415	3.40	0.0160	0.037
KS refl	155	3.33	0.0163	0.035
HD	485	3.36	0.0159	0.038
KG refl	125	3.22	0.0173	0.036
MH refl	265	3.30	0.0147	0.039
CW	**35**	3.26	**0.0118**	**0.030**
SH	485	3.39	0.0155	0.038
SSV refl	75	3.08	0.0150	0.034
RR	72	**2.64**	0.0124	0.031
ALL	503	3.41	0.0158	0.038

[a]For KS, KG, MH, and SSV numerical results are given for the case where reflectance data was used as input for the selection.

As noted earlier the experiments were repeated three times to avoid a bias caused by the PC's random number generator. This happens mainly in KG and SSV, because their clustering method's cluster centroids must be initialized. Depending on that, the algorithm might or might not converge for a given number of maximal iterations (here, 100). Of course, random initialization also happens in RD because of the intrinsic random nature of the selection process. For KG and SSV, this effect decreases with an increasing n, as convergence in few iterations is more likely for a large n. The influence can be measured with the standard deviation (std) of mean ΔE_{00} estimation error over the number of repetitions. For instance, for $n = 50$ and 10 repetitions $std_{RD} = 0.68$, $std_{KG} = 0.65$ and $std_{SSV} = 0.38$ were found. For the same reason of cluster initialization and possible limitation of the convergence of the clustering, the RR method is also influenced by this ($std_{RR} = 0.23$ for 10 repetitions), however, less dramatically. The reason is that the clustering in this method is only used to partition the data. Whether or not a cluster of samples is rejected from the training set depends on the importance of the samples for training, not on the structure of the cluster itself. Not reaching convergence in clustering is therefore rather likely to lead to an increase in time for sample selection, as a large number of recursive sub-clustering would have to be performed.

B. Experiment 2: Comparative Analysis of the Novel Training Selection Scheme Using a Very Large Pool of Selectable Training Data

We found that RD, KG, KS, and MH were able to better handle large datasets than the other approaches and therefore were compared with the RR method using the modeled set for training sample selection. The th and k parameter in the RR method are found by exhaustive search given the objective to minimize the mean ΔE_{00} error in spectral estimation as $th_{opt} = 5e^{-8}$ and $k_{opt} = 3$. Figure 7 illustrates a surface plot of the $\overline{\Delta E_{00}}$ error over the parameter space for Toyo data. The small optimal number of clusters might not be intuitive, as the obtained clustering appears to be too rough to divide the entire space spanned by the large amount of training reflectances. However, the
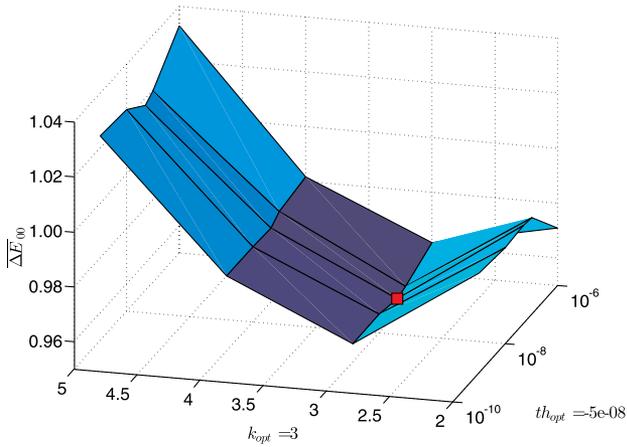
Fig. 7. RR method optimal parameter search: $\overline{\Delta E}_{00}$ error over k and th.

Table 2. Experiment 2 Compares Mean Estimation Error and Time Performance for the Method-Depending Optimal n_{opt} Selected Training Samples and the Large Dataset[a]

Method	RD	KS resp	KG resp	MH refl	RR	ALL
n_{opt}	**410**	610	4010	6410	1015	14641
$\overline{\Delta E}_{00}$	1.04	0.98	1.03	1.00	**0.95**	1.03
$\overline{d_p}$	0.0059	0.0053	0.0058	0.0056	**0.0042**	0.0058
$\overline{\text{RMSE}}$	0.044	0.041	0.044	0.043	**0.036**	0.043
$\overline{t}_{sel}$	<0.1 s	91 s	7.5 min	**14 s**	19 s	—
$\overline{t}_{opt}$	<0.1 s	129 h	8 h	44 min	—	—

[a]For applicable methods, it is indicated whether reflectance data (refl) were used as input or camera responses (resp).

recursive fashion of the algorithm leads to a progressive reduction of the number of samples per cluster in deeper recursions. This results in an adaptive finer clustering for those parts of the training data space that are relevant to maintain the estimation quality.

We repeated Experiment 1 given the current dataset for training to seek for an optimal number n_{opt} of training samples for each method. KG and KS had to be computed with camera responses as input (rather than reflectance data) to make the computations feasible. For MH, we considered reflectances. Figure 8 illustrates the mean estimation error over number of samples. Unlike with the small dataset, for the large dataset and all methods, n_{opt} was found to be less than the total number of samples in the set. That indicates that the modeled dataset for training is to some extend redundant for the task of spectral estimation, which supports the development intention of the RR method. Performing RR selection, 1015 samples are selected. In addition to RD, KS, KG, MH, and RR, we also present results for the case where all available training samples were used for training (further referred to as ALL). Table 2 shows the final results for all methods and n_{opt}.

The lowest ΔE_{00} mean estimation error is found for the RR method, followed by KG, KS, and MH. The difference in colorimetric estimation quality is very little. Spectrally, RR clearly outperforms the other methods. Comparing the time consumption for training set selection (t_{sel}) and finding the optimal n (t_{opt}) makes a big difference. t_{opt} ranges from several minutes in MH to several days in KS. RR, on the other hand, does not require time to select the optimal number of training samples, since this is done as part of the selection routine. The time for finally selecting the training set ranges from a few seconds for MH to a few minutes for KG. RR is with 19 s quite fast. The time for estimation (t_{est}) is upper bounded with $t_{est} < 0.02$ s for ALL, and less time is required for other methods with a smaller number of samples in the training set. However, the increase in estimation quality strongly supports our intent to use the reduced training set rather than all samples.

6. Conclusions and Future Work

In this work, we have proposed what we believe is a novel scheme of global training sample selection for the process of training a spectral estimation system for in-line print inspection. Unlike commonly applied global training, we do not train the system with publicly available test charts, color atlases, or a subset of them, but instead use a printer characterization model to generate a large set of ink reflectances (14600 or more samples). We applied a training sample selection algorithm (the RR method) to reduce the size of the training set while maintaining or enhancing the estimation quality.

The proposed algorithm can handle large datasets, unlike most of the other state-of-the-art global training sample selection algorithms. Our recursive method uses clustering to reject groups of similar samples that do not contribute to enhancing the estimation quality of a realistic application dataset that is required as input for the algorithm. Introducing prior knowledge about a specific estimation application, as proposed in our approach, allows the training sample selection to be tailored in an application-dependent fashion and therefore offers

Fig. 8. Mean colorimetric estimation performance ($\overline{\Delta E}_{00}$) over number of training samples n for RD, KG, KS, and MH method and the large dataset. The optimal number of samples n_{opt} is indicated for all methods, including RR with a squared marker.

an advantage over other methods that aim for a more general design.

It is shown experimentally for real data from a prototype of a 12-channel multispectral camera system for in-line print inspection that the approach of using modeled reflectances as pool for the training together with the RR method for sample selection is outperforming other state-of-the-art methods. This increase in estimation quality is not significant, but a major advantage of our method is that the optimal number of selected training samples is determined by the RR algorithm automatically and does not have to be evaluated in a time-consuming process, as is the case for other methods compared. We conclude that reflectance modeling is a feasible alternative to physically producing a large training set by printing and later measuring spectral reflectances. The RR method can be used to process these large datasets.

In future work, we will further investigate the proposed RR algorithm for training sample selection. The threshold parameter th is currently a fixed value, but could be implemented as a function of the level of recursion of the algorithm, which would allow a different treatment of small clusters compared to big clusters. Also, the number of clusters k itself is fixed in the current implementation. By allowing th and k to change dynamically for consecutive recursive calls of the algorithm we expect to be able to further reduce the final number of samples selected without decreasing the estimation quality obtained for the test sets. As mentioned previously, the current implementation of our selection algorithm uses linear least-square regression (PI estimation) for simplicity.

Another interesting possibility is to extend our experiments to using printed instances of optimized training sets that are obtained by different approaches. By doing so, we want to determine if modeled reflectances and corresponding noisy simulated camera responses (as used in this work) are performing equally as well as real data. Furthermore, reflectances are modeled for an ink-jet printer in this work; however, the multispectral approach for in-line print inspection considered here is primarily intended for offset printing. We regard reflectance modeling for offset printing technology as future work to be considered once the prototype multispectral camera system is integrated in a printing machine.

References

1. H. Haneishi, T. Hasegawa, A. Hosoi, Y. Yokoyama, N. Tsumura, and Y. Miyake, "System design for accurately estimating the spectral reflectance of art paintings," Appl. Opt. **39**, 6621–6632 (2000).
2. M. Vilaseca, R. Mercadal, J. Pujol, M. Arjona, M. de Lasarte, R. Huertas, M. Melgosa, and F. Imai, "Characterization of the human iris spectral reflectance with a multispectral imaging system," Appl. Opt. **47**, 5622–5630 (2008).
3. E. M. Valero, Y. Hu, J. Hernández-Andrés, T. Eckhard, J. L. Nieves, J. Romero, M. Schnitzlein, and D. Nowack, "Comparative performance analysis of spectral estimation algorithms and computational optimization of a multispectral imaging system for print inspection," in *Color Research & Application* (Wiley, 2014), pp. 16–27.
4. M. Tse, D. Forrest, and J. Briggs, "Automated print quality analysis for digital printing technologies," in *Pan-Pacific Imaging Conference/Japan Hardcopy* (Society of Electrophotography of Japan, Tokyo, Japan, 1998).
5. W. Wu, B. Walczak, D. Massart, S. Heuerding, F. Erni, I. Last, and K. Prebble, "Artificial neural networks in classification of NIR spectral data: design of the training set," Chemom. Intell. Lab. Syst. **33**, 35–46 (1996).
6. V. Cheung and S. Westland, "Methods for optimal color selection," J. Imaging Sci. Technol. **50**, 481 (2006).
7. J. Kang, K. Ryu, and H. Kwon, "Using cluster-based sampling to select initial training set for active learning in text classification," *Advances in Knowledge Discovery and Data Mining*, Lecture Notes in Computer Science (Springer, 2004), Vol. **3056**, pp. 384–388.
8. J. Hardeberg, "Acquisition and reproduction of color images: colorimetric and multispectral approaches," Ph.D. thesis (Universal Publishers, 2001).
9. H. Shen, H. Zhang, J. Xin, and S. Shao, "Optimal selection of representative colors for spectral reflectance reconstruction in a multispectral imaging system," Appl. Opt. **47**, 2494–2502 (2008).
10. O. Kohonen, J. Parkkinen, and T. Jääskeläinen, "Databases for spectral color science," Color Res. Appl. **31**, 381–390 (2006).
11. V. Heikkinen, R. Lenz, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, and T. Jääskeläinen, "Evaluation and unification of some methods for estimating reflectance spectra from rgb images," J. Opt. Soc. Am. A **25**, 2444–2458 (2008).
12. J. Nieves, E. Valero, S. Nascimento, J. Hernández-Andrés, and J. Romero, "Multispectral synthesis of daylight using a commercial digital CCD camera," Appl. Opt. **44**, 5696–5703 (2005).
13. CIE, "Improvement to industrial colour-difference evaluation," Tech. Rep., CIE Pub. No. 142–2001 (2001).
14. J. Romero, A. García-Beltrán, and J. Hernández-Andrés, "Linear bases for representation of natural and artificial illuminants," J. Opt. Soc. Am. A **14**, 1007–1014 (1997).
15. J. Viggiano, "Metrics for evaluating spectral matches: a quantitative comparison," in *Proceedings of the 2nd European Conference on Colour Graphics, Imaging and Vision* (Society for Imaging Science and Technology, 2004), pp. 286–291.
16. D. Wyble and R. Berns, "A critical review of spectral models applied to binary color printing," Color Res. Appl. **25**, 4–19 (2000).
17. R. Hersch and F. Crété, "Improving the Yule–Nielsen modified spectral neugebauer model by dot surface coverages depending on the ink superposition conditions," in Proc. SPIE **5667**, 434–445 (2005).
18. J. Granahan and J. Sweet, "An evaluation of atmospheric correction techniques using the spectral similarity scale," in *Geoscience and Remote Sensing Symposium (IGARSS'01)* (IEEE, 2001), Vol. **5**, pp. 2022–2024.
19. A. Rodriguez, J. Nieves, E. Valero, E. Garrote, J. Hernández-Andrés, and J. Romero, "Modified fuzzy c-means applied to a

Bragg grating-based spectral imager for material clustering," in Proc. SPIE **8300**, 830003 (2012).

20. M. Mohammadi, M. Nezamabadi, R. Berns, and L. Taplin, "A prototype calibration target for spectral imaging," in *Tenth Congress of the International Colour Association* Granada, Spain (2005), pp. 387–390.

21. S. Theodoridis and K. Koutroumbas, *Pattern Recognition* (Academic, 2008).

22. P. C. Cosman, K. L. Oehler, E. A. Riskin, and R. M. Gray, "Using vector quantization for image processing," Proc. IEEE **81**, 1326–1341 (1993).

23. D. Lee, S. Baek, and K. Sung, "Modified k-means algorithm for vector quantizer design," IEEE Signal Process. Lett. **4**, 2–4 (1997).

24. T. Eckhard, "Color Imaging Laboratory homepage of the University of Granada," http://www.ugr.es/~colorimg/suppl_docs/global_training.html (2013).

25. M. de Lasarte, M. Arjona, M. Vilaseca, and J. Pujol, "Influence of the number of samples of the training set on accuracy of color measurement and spectral reconstruction," J. Imaging Sci. Technol. **54**, 30501–30510 (2010).