Evaluating logarithmic kernel for spectral reflectance estimation—effects on model parametrization, training set size, and number of sensor spectral channels

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

[1]*Optics Department, University of Granada, ES-18071 Granada, Spain*
[2]*School of Computing, University of Eastern Finland, FN-80101 Joensuu, Finland*
**Corresponding author: timo.eckhard@gmx.com*

Received July 2, 2013; revised December 12, 2013; accepted December 25, 2013;
posted January 6, 2014 (Doc. ID 193051); published February 13, 2014

In this work, we evaluate the conditionally positive definite logarithmic kernel in kernel-based estimation of reflectance spectra. Reflectance spectra are estimated from responses of a 12-channel multispectral imaging system. We demonstrate the performance of the logarithmic kernel in comparison with the linear and Gaussian kernel using simulated and measured camera responses for the Pantone and HKS color charts. Especially, we focus on the estimation model evaluations in case the selection of model parameters is optimized using a cross-validation technique. In experiments, it was found that the Gaussian and logarithmic kernel outperformed the linear kernel in almost all evaluation cases (training set size, response channel number) for both sets. Furthermore, the spectral and color estimation accuracies of the Gaussian and logarithmic kernel were found to be similar in several evaluation cases for real and simulated responses. However, results suggest that for a relatively small training set size, the accuracy of the logarithmic kernel can be markedly lower when compared to the Gaussian kernel. Further it was found from our data that the parameter of the logarithmic kernel could be fixed, which simplified the use of this kernel when compared with the Gaussian kernel.

OCIS codes: (150.0150) Machine vision; (100.3010) Image reconstruction techniques; (100.3190) Inverse problems; (110.4234) Multispectral and hyperspectral imaging; (330.1710) Color, measurement.
http://dx.doi.org/10.1364/JOSAA.31.000541

1. INTRODUCTION

Kernel methods for pattern recognition or machine learning tasks have proven to be powerful [1–3]. In this work, we concentrate on the problem of spectral reflectance estimation of image data, using kernel-based regression. The data variables are camera responses of an image scene, captured by a multispectral imaging system and corresponding spectral reflectance data. The aim of the regression is to determine their relation, using a set of training data. Once the regression model is constructed, spectral reflectances can be estimated from camera responses.

It has been shown that regularized regression is useful for reflectance estimation. Especially, a kernel-based (ridge) regression has been demonstrated to provide improvement in accuracy over traditional regularized regression models [4–7]. Valid kernel functions are positive definite (*further: PD*) or, with additional constraints imposed on the regression model, kernel functions that are conditionally positive definite (*further: CPD*). Some instances of PD functions are the linear or the Gaussian kernel, and from the CPD class the thin plate spline kernel [5,8].

In this paper, the so-called *logarithmic kernel* [9] (a member of the CPD class) is to our knowledge for the first time introduced for spectral reflectance estimation, and compared with the linear and the Gaussian kernel functions. The logarithmic kernel function has been used previously in SVM based image recognition [9].

In our experiments, we use the Pantone and HKS reflectance datasets and measured as well as simulated camera response data. The performance of the three kernel functions is demonstrated with a varying number of training samples and sensor spectral channels. In addition, the performance of the logarithmic kernel is also evaluated in case the kernel is used via fixed parameter.

This article is structured as follows: after introducing the notation in Section 2, Section 3 revises the spectral reflectance estimation problem and the kernel-based regression. Further, all kernel functions compared in this work are summarized, and the CPD logarithmic kernel function is introduced. The experiments are explained in Section 4. The results and discussion follow in Section 5, and finally the most relevant conclusions are summarized in Section 6.

2. NOTATION

In what follows, a column vector is denoted by a boldface letter such as $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$. Matrices are denoted by capital letters, such as A.

3. THEORY

A. Image Acquisition Model

For image formation of the multispectral imaging system, we assume the general discrete image transfer function:

$$\mathbf{x} = W\mathbf{r} + \mathbf{b}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes a camera response sample, W is the $n \times m$ matrix that combines the effect of spectral responsivities

of the camera system and scene illumination, and $\mathbf{r} \in \mathbb{R}^m$ is a spectral reflectance vector. The vector $\mathbf{b} \in \mathbb{R}^n$ accounts for noise or nonlinearities present in the acquisition process. In our work, we assume fixed-measurement geometry and that the scene illumination is uniform over the field of view of the camera.

B. Kernel Ridge Regression Applied To Spectral Estimation

Spectral estimation can be considered as the problem of finding a function that best approximates a set of training data with respect to some loss function. In spectral reflectance estimation, the training dataset S consists of l camera responses $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_l\} \subset \mathbb{R}^n$, acquired by an n channel multispectral imaging system and corresponding known spectral reflectance vectors $R = \{\mathbf{r}_1, \ldots, \mathbf{r}_l\} \subset \mathbb{R}^m$. Once this function is found, spectral reflectance vectors of unknown objects can be estimated from camera responses acquired under the same viewing conditions.

Following [4,5,8], our optimization problem is to find the $N \times m$ matrix F such that we minimize

$$\underset{F}{\operatorname{argmin}} \left(\sum_{i=1}^{l} \|\mathbf{r}_i - F^T \Phi(\mathbf{x}_i)\|^2 + \lambda \|F\|_F^2 \right), \tag{2}$$

where $\Phi : \mathbb{R}^n \to \mathcal{F}$ is a feature mapping from camera response space to a feature space $\mathcal{F}$ of dimension N, $\|F\|_F = (\sum_{i=1}^{m} \sum_{j=1}^{N} |F_{ij}|^2)^{1/2}$ denotes the Frobenius norm, and λ is a regularization parameter. The regularization term is added to avoid numerical instability and overfitting to training data.

The feature mapping $\Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \ldots, \phi_N(\mathbf{x}))^T$ is often nonlinear and can be infinite dimensional. Assuming that the feature mapping $\Phi(\mathbf{x})$ is induced by a PD kernel [1] and using the corresponding kernel representation (kernel trick), the minimization problem above is equivalent to finding a matrix A such that

$$\underset{A}{\operatorname{argmin}} \left(\sum_{i=1}^{l} \|\mathbf{r}_i - A^T \mathbf{k}(\mathbf{x}_i)\|^2 + \lambda \operatorname{Tr}(A^T K A) \right), \tag{3}$$

where $A = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m]_{l \times m}$ is a matrix of weight vectors $\alpha \in \mathbb{R}^l$, $\operatorname{Tr}(\cdot)$ denotes matrix trace, $K = [k(\mathbf{x}_i, \mathbf{x}_j)]_{l \times l}$ is the kernel matrix of training data, and $\mathbf{k}(\mathbf{x}_i) = (k(\mathbf{x}_1, \mathbf{x}_i), \ldots, k(\mathbf{x}_l, \mathbf{x}_i))^T \in \mathbb{R}^l$ is a vector containing the kernel evaluations between the camera response training set and sample $\mathbf{x}_i$.

By differentiating Eq. (3) with respect to A and setting the resulting function equal to zero we obtain

$$(K + \lambda I_l) A = R^T, \tag{4}$$

where I_l is the identity matrix of size $l \times l$, and $R = [\mathbf{r}_1, \ldots, \mathbf{r}_l]_{m \times l}$ contains training reflectances. The solution to the minimization is $A = (K + \lambda I_l)^{-1} R^T$. Reflectance $\tilde{\mathbf{r}}$ can now be estimated from camera response $\mathbf{x}$ as

$$\tilde{\mathbf{r}} = A^T \mathbf{k}(\mathbf{x}) = R(K + \lambda I_l)^{-1} \mathbf{k}(\mathbf{x}). \tag{5}$$

C. PD Kernels in Ridge Regression

Positive definiteness of a function is defined as [10] (p. 67): let X be a closed and bounded subset of $\mathbb{R}^n$. A function k is called positive definite if $\sum_{i,j=1}^{l} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$, for all $l \in \mathbb{N}$, $\mathbf{x}_i$, $\mathbf{x}_j \in X$, and α_i, $\alpha_j \in \mathbb{R}$.

Positive definite kernel functions considered in this work are the linear and Gaussian kernels. These kernels are probably the most widely used kernel functions in spectral estimation [4,5,7]. The linear kernel function [1] is in fact a homogeneous polynomial kernel of degree $d = 1$, defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j)^d, \quad \text{with } d = 1, \tag{6}$$

where $\mathbf{x}_i$, $\mathbf{x}_j \in \mathbb{R}^n$. The isotropic Gaussian kernel [1,3–5] is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right), \tag{7}$$

where $\mathbf{x}_i$, $\mathbf{x}_j \in \mathbb{R}^n$ and $\sigma > 0$.

Heikkinen *et al.* discussed that the selection of too small values for σ (corresponding to a small effective area of the kernel) can lead to over-fitting of the training data and therefore a risk of poor generalization. A too large selected value σ can lead to oversimplification [5].

D. CPD Kernels in Ridge Regression

There exists a larger class of kernel functions that can be used within the framework of kernel ridge regression, given that certain constraints can be imposed on the regression model. These functions are CPD kernel functions, and they are defined as follows [1] (p. 49): let X be a closed and bounded subset of $\mathbb{R}^n$. A function k is called CPD if $\sum_{i,j=1}^{l} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$, for all $l \in \mathbb{N}$, $\mathbf{x}_i$, $\mathbf{x}_j \in X$, and α_i, $\alpha_j \in \mathbb{R}$ with $\sum_{i=1}^{l} \alpha_i = 0$.

For CPD kernels and d-conditionally positive kernels, a semi-parametric model of ridge regression is formulated introducing a polynomial expansion of feature vectors in Eq. (3) [1,8]. In the case of the logarithmic kernel, the semi-parametric model is formed by adding only a constant term to the model. The related minimization problem is

$$\underset{\{A, \mathbf{a}\}}{\arg\min} \left(\sum_{i=1}^{l} \|\mathbf{r}_i - A^T \mathbf{k}(\mathbf{x}_i) - \boldsymbol{a}\|^2 + \lambda \operatorname{Tr}(A^T K A) \right) \quad \text{s.t. } \mathbf{1}^T A = \mathbf{0}^T \tag{8}$$

where $A = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m]_{l \times m}$ and $\mathbf{k}(\mathbf{x}_i) \in \mathbb{R}^l$ are the same as in the PD case (Eq. 3). The vector corresponding to the additional constant term is $\boldsymbol{a} = (a_1, \ldots, a_m)^T$ and $\mathbf{1}$ and $\mathbf{0}$ are the all-one and all-zero column vectors of length l and m, respectively.

The block matrix notation corresponding to the minimization problem in Eq. (8) is

$$\begin{bmatrix} K + \lambda I_l & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} A \\ \mathbf{a}^T \end{bmatrix} = \begin{bmatrix} R^T \\ \mathbf{0}^T \end{bmatrix}, \tag{9}$$

where matrix R and K are the same as in the PD case, $R = [\mathbf{r}_1, \ldots, \mathbf{r}_l]_{m \times l}$ and $K = [k(\mathbf{x}_i, \mathbf{x}_j)]_{l \times l}$. A reflectance, $\tilde{\mathbf{r}}$, can now be estimated from camera response $\mathbf{x}$ as

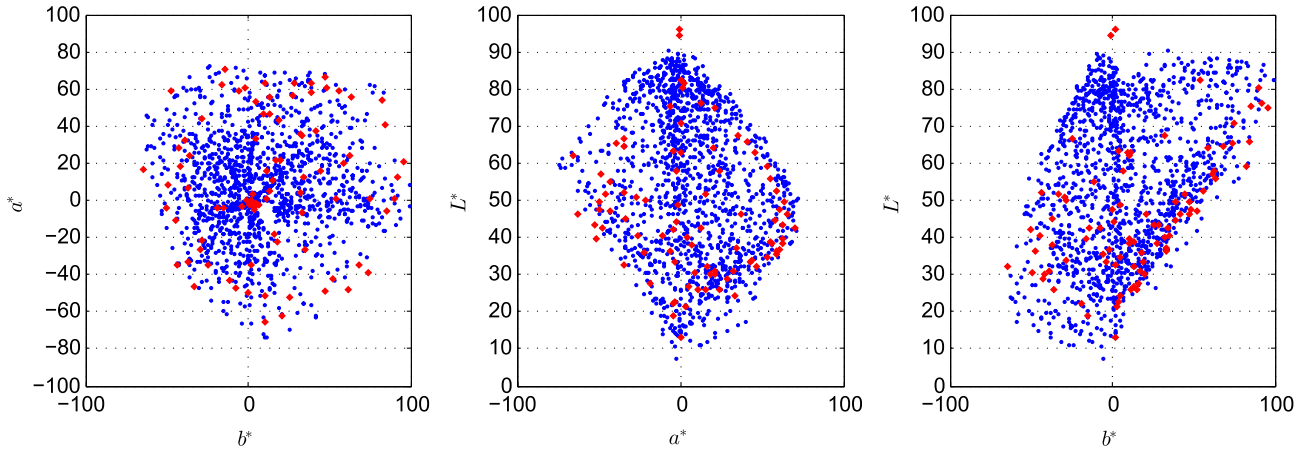

Fig. 1. CIE-L*a*b* coordinates for the Pantone (blue circle markers) and HKS dataset (red diamond markers).

$$\tilde{\mathbf{r}} = A^T \mathbf{k}(\mathbf{x}) + \mathbf{a}. \tag{10}$$

In our experiments, we focused on evaluating the logarithmic kernel defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = -\log(1 + \|\mathbf{x}_i - \mathbf{x}_j\|^\beta), \quad \text{with } 0 < \beta \leq 2, \tag{11}$$

where $\mathbf{x}_i$, $\mathbf{x}_j \in \mathbb{R}^n$.

The CPD-ness of the logarithmic kernel can be demonstrated as follows: all kernels of form $-\|\mathbf{x_i} - \mathbf{x_j}\|^\beta$ are CPD, if $0 \leq \beta \leq 2$ [11] (p. 49). Further, $-\log(1-k)$ is a CPD kernel, if $k: X \times X \mapsto [-\infty, 0]$ is a CPD kernel [10], [11] (p. 50). It follows that $-\log(1 + \|\mathbf{x}_i - \mathbf{x}_j\|^\beta)$ is a CPD kernel when $0 < \beta \leq 2$.

4. EXPERIMENTAL CONFIGURATION AND EVALUATION

We designed a set of experiments to demonstrate the performance of the logarithmic kernel in kernel ridge regression spectral estimation. In this section, the acquisition system and datasets used for our experiments are introduced. Further, we illustrate the evaluation scheme that is the basis for the experimental work.

A. Datasets and Acquisition System

The performance of kernel regression depends on the selection of the kernel function, training data, and the acquisition system. We consider two spectral datasets, namely 1314 samples of a coated Pantone color chart and 91 full-tone samples of a coated HKS color chart. The corresponding inks for these spot colors are widely used in the offset printing industry. The color gamut spanned by these datasets is illustrated in Fig. 1, with CIE-L*a*b* coordinates calculated for the CIE-1964 10° standard observer and CIE-D65 illumination. It can be observed that both datasets span a similar area in the color space, but some lighter samples of the HKS set are beyond the gamut boundaries of the Pantone set. The spectral data was measured with an X-Rite Spectro-Eye spectrophotometer and is sampled from 380 to 730 nm in 10 nm steps.

The multispectral acquisition system considered here consists of the prototype 12-channel line-scan camera truePIXA and a prototype LED illumination system, both developed by Chromasens GmbH [12]. The camera sensor used in the truePIXA system is in fact a high-resolution RGB line-scan sensor with four lenses that are mounted in series in front of the camera. The four lenses have a common field-of-view from which the final multichannel image is constructed. Each lens has a distinct color filter attached, which, in combination with the RGB filter in front of the sensor and the transmittance of the lens, modulate the effective responsivity of the system. The RGB channel of the sensor in combination with the four filters allow acquisition of 12 camera responses at once. The acquisition principle is illustrated schematically in Fig. 2, and spectral curves of the camera system's effective responsivity are illustrated in Fig. 3. The measurement geometry of the system is close to 45/0, with some deviation of the observation angle due to the image projection of corresponding spatial locations of the measurement surface over the four lenses on different sensor locations.

The short exposure times required for acquisition with the line-scan system allow an over-sampling of the target and corresponding temporal averaging. This temporal pixel-wise averaging of replicated measurements increases the signal-to-noise ratio of the measurement with the square root of the number of averaged measurements, given the noise is a random signal with zero mean [13]. Further, the temporal

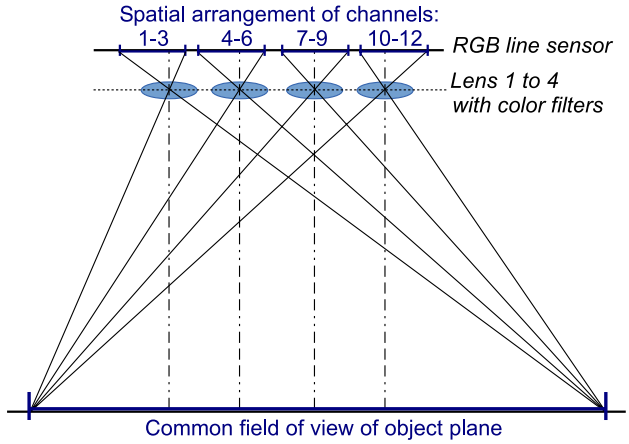

Fig. 2. Schematic illustration of the working principle of the 12-channel line-scan camera truePIXA. The RGB channel of the line sensor in combination with the four filters allow acquisition of 12 camera responses per image pixel at once.

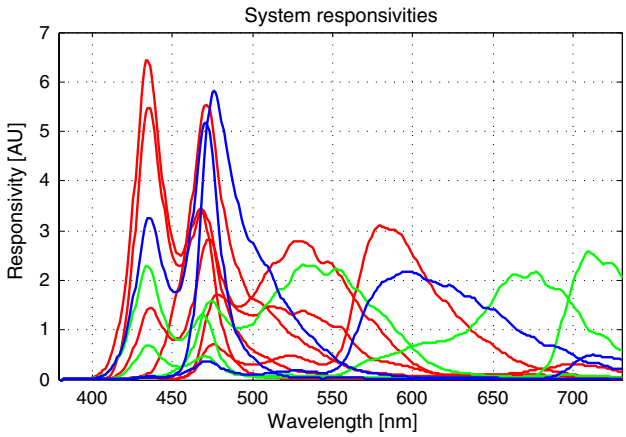

Fig. 3. Illustration of the system responsivities [sensor spectral responsivity plus illumination in arbitrary units (AU)]. Blue corresponds to the three-channel system, blue+green to the six-channel, blue+green+red to the 12 channel system used for experiments in Section 5.E.

averaged camera responses are spatially averaged over an area that corresponds roughly to that of the aperture of the spectrophotometer used for reflectance measurements.

In real capture systems, both noise and nonlinearities inherent to the acquisition process influence the estimation quality. In this work, we perform our analysis using measured data and simulated data. The simulated case can be considered to correspond to ideal noise-free conditions. Accordingly, simulated camera responses were calculated using the linear model in Eq. (1) with $\mathbf{b} = 0$.

B. Evaluation Scheme and Implementation Details

We used tenfold cross validation [11] for optimizing and evaluating the models. This means that in each of 10 rounds of evaluation, a subset of 10% of the initially shuffled data is used for testing, and the rest for training the system. This process allows using the entire dataset for testing the spectral estimation, while keeping training and test data separated. Further, we implemented our evaluation scheme such that free parameters are selected based on average RMSE minimization and tenfold cross-validation: in each evaluation fold, the available training data is further partitioned and processed in a tenfold manner to find the optimal parameter as the one minimizing the average RMSE over all folds with the chosen parameter grid. A proper selection of the model parameters in kernel ridge regression is a key factor for obtaining a satisfactory estimation performance.

In the case of kernel ridge regression estimation, parameters are the regularization term λ and others depending on the type of kernel function. For the Gaussian kernel [Eq. (7)], we have σ controlling the width of the Gaussian and, for the logarithmic kernel [Eq. (11)], the power term β in the logarithmic function.

So, in each fold, an optimal parameter has to be selected from either a 1D or 2D parameter search space. We adjusted the parameter grid such that throughout all experiments, optimal parameters were never selected to be on the edge of the parameter grid (unless the edge of the grid corresponds to the parameter range limit, as is the case with $\beta = 2$ for the logarithmic kernel). We used the Gaussian kernel parameter σ in the range $[10^{-4}, 50]$, the logarithmic kernel parameter β in the range [0.1,2], and the regularization parameter λ in the range $[10^{-13}, 0.5]$. We evaluated several sampling positions for these parameters to get information about the differences between the performance of the chosen optimization grids. It was found that these differences were small. The experimental results are calculated using 34 sampling points for σ, 14 sampling points for β, and 13 sampling points for λ.

Camera responses are normalized to the range [0,1] by dividing each camera response channel-wise by the maximal camera response value over the entire dataset. Training reflectances are centered by subtracting the mean, prior to estimation. Consequently, the mean of the training reflectances has to be added to the recovered spectra after estimation. Our implementation of kernel ridge regression with the logarithmic kernel function can be accessed through the web page of the first author's institution [14].

We evaluated estimation performance in terms of spectral as well as color difference metrics. As discussed previously by several authors, there is not any metric that is conclusively superior to others for all purposes [15,16].

The RMSE is defined for an estimated spectrum, $\tilde{\mathbf{r}}$, and its measured counterpart, $\mathbf{r}$, as

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\|\mathbf{r} - \tilde{\mathbf{r}}\|^2}. \tag{12}$$

Further, d_p is the Pearson distance, also known as the complemented GFC (goodness of fit coefficient) [16,17], and defined as

$$d_p = 1 - \frac{\mathbf{r}^T\tilde{\mathbf{r}}}{\|\mathbf{r}\|\|\tilde{\mathbf{r}}\|}. \tag{13}$$

The formulation of color differences aims at predicting the magnitude of the perceived color difference between two color stimuli [18]. In this study, we followed the latest recommendations of the International Commission on Illumination (CIE) and used ΔE_{00} color difference formula in CIE L*a*b* color space [19]. We computed CIE L*a*b* coordinates of reflectance spectra, assuming the CIE-1964 10° standard observer and CIE-D65 standard illuminant. The white point was set to the perfect reflecting diffuser {vector with values equal to 1, [20] (p. 48)}.

5. RESULTS

The following subsections summarize results and findings related to training set size, number of spectral sensor channels, and kernel parametrization.

A. Selection of Model Parameters

We have analyzed the RMSE surfaces over the parameter search grid for logarithmic and Gaussian kernel function and our experimental data.

The selection of the regularization parameter λ depends on the type of data as well as the number of sensor spectral channels in the acquisition system and has to be optimized for all kernel functions compared in this work. For the Gaussian kernel, we identified from our data that the selection of σ influences the choice of λ and vice versa. Figure 4 illustrates one typical instance of the RMSE surface over the parameter grid, as can be found from one fold of cross-validation with Pantone data and the 12 channel system for the Gaussian

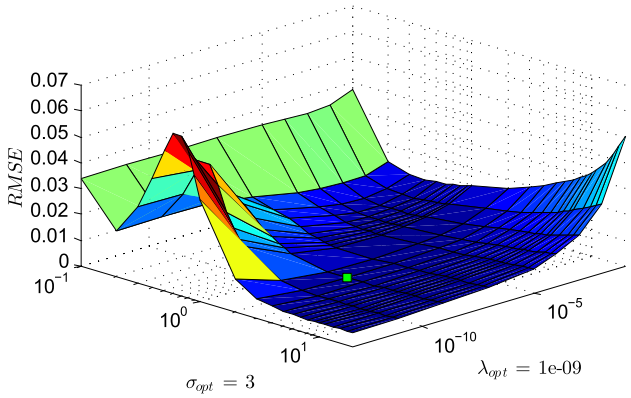

Fig. 4. Illustration of the mean RMSE (over 10 folds of parameter optimization) over the parameter search space for Pantone 12C measured camera response data and the Gaussian kernel. The optimal parameter selection is illustrated by a green square. The parameter search space has been cropped for illustration purposes.

kernel function. As explained previously, one fold of evaluation includes 10 folds of parameter optimization, so the resulting RMSE is a mean value over these folds. From the figure, the interdependence of the two parameters can be observed.

Figure 5 illustrates the RMSE surface over parameter space for the logarithmic kernel, corresponding to the same setting as used for Fig. 4. The β parameter of the logarithmic kernel behaves very differently from σ of the Gaussian kernel, as the selection of β does not seem to influence the selection of λ for our data. This is an interesting observation regarding the logarithmic kernel, as it allows us to perform parameter selection sequentially for each parameter. For the Gaussian kernel function, this is not the case.

Further, we identified that $\beta = 2$ led to the lowest estimation errors for both evaluated datasets, simulated or measured data, a different number of sensor spectral channels as well as a varying number of training samples. For example, we illustrate the RMSE surfaces over parameter space for a for varying number of training samples (200, 400, 600, 800) in Fig. 6. The optimal parameters for this example were found to be ($\beta_{1\ldots4} = 2$, $\lambda_{1\ldots4} = 10^{-4}$). From these results, it seems that the shape of the RMSE surface is in a wide area invariant to the number of training samples. A range where this does

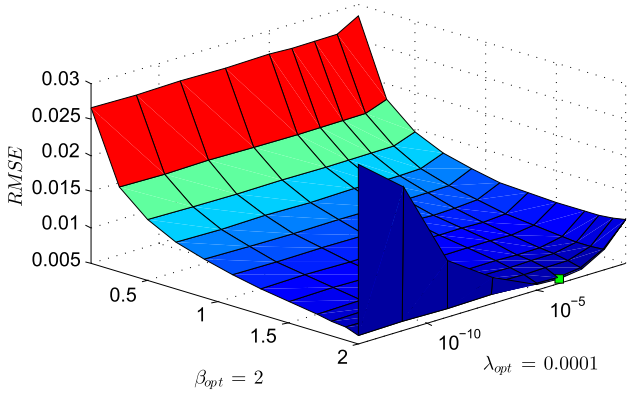

Fig. 5. Illustration of the mean RMSE (over 10 folds of parameter optimization) over the parameter search space for Pantone 12C measured camera response data and the logarithmic kernel. The optimal parameter selection is illustrated by a green square. The parameter search space has been cropped for illustration purposes.

not apply is when $\beta = 2$, for which the selection of λ influences the corresponding RMSE differently depending on the number of training samples.

Using $\beta = 2$ is the special case of Eq. (11), which directly incorporates the negative squared distance kernel. Fixing the β parameter reduces the parametrization of kernel ridge regression to a optimization problem in a 1D search space (only λ has to be optimized).

B. Influence of the Number of Training Samples

Throughout our experimental work related to kernel ridge regression, we identified a strong dependence of the number of samples used for training and the estimation performance. Obviously, a too small number of training samples leads to poor estimation quality because the trained model does not generalize well for any other than the training data. Further, it is not only the amount of samples that influences the system performance, but also the type of data used for training. In this subsection, we limit our analysis to a discussion about the amount of samples used for training in the case of the logarithmic, Gaussian, and linear kernel. To analyze this factor, we compared a tenfold cross-validation spectral estimation performance for different numbers of training samples of the shuffled Pantone dataset. Varying dataset sizes considered here were from 50 to 100 samples in steps of 10 and from 200 to 1314 samples in steps of 100. The optimal parameter was again selected by average RMSE minimization over 10 folds using the parameter ranges described in Section 4.B.

The results of this experiment for the case of noiseless simulated camera responses show that the estimation performance (here, RMSE) as a function of training samples, l, behaves differently depending on the choice of the kernel function. The logarithmic kernel shows the strongest dependence on the number of training samples, especially for small l, showing a clear decrease in performance as compared with the Gaussian and linear kernels. At approximately 530 samples, the linear and logarithmic kernel share similar estimation performance. With a larger l, the logarithmic kernel approaches the estimation performance of the Gaussian kernel. Figure 7 illustrates these results graphically.

Also for measured camera responses of the Pantone dataset, fewer training samples result in lower RMSE. Unlike in the simulated data case, the relation of RMSE and the number of training samples for the logarithmic, linear, and the Gaussian kernel is found to be quite similar in shape. Only for the linear kernel, the RMSE is higher than that of the logarithmic and Gaussian kernel for every number of training samples evaluated. This result is illustrated in Fig. 8. The dissimilarity found in the results for these two conditions indicates strong differences between measured and simulated data.

C. Estimation Performance with Measured Sensor Responses from the 12-Channel System

We have compared the spectral and colorimetric estimation results obtained with the logarithmic, Gaussian, and linear kernel for the Pantone and the HKS data of the 12-channel acquisition system. Numerical results are illustrated in Table 1.

For Pantone data, it can be observed that the mean colorimetric estimation quality is similar to that of the Gaussian kernel function ($\Delta E_{00} = 0.33$), the mean spectral quality in

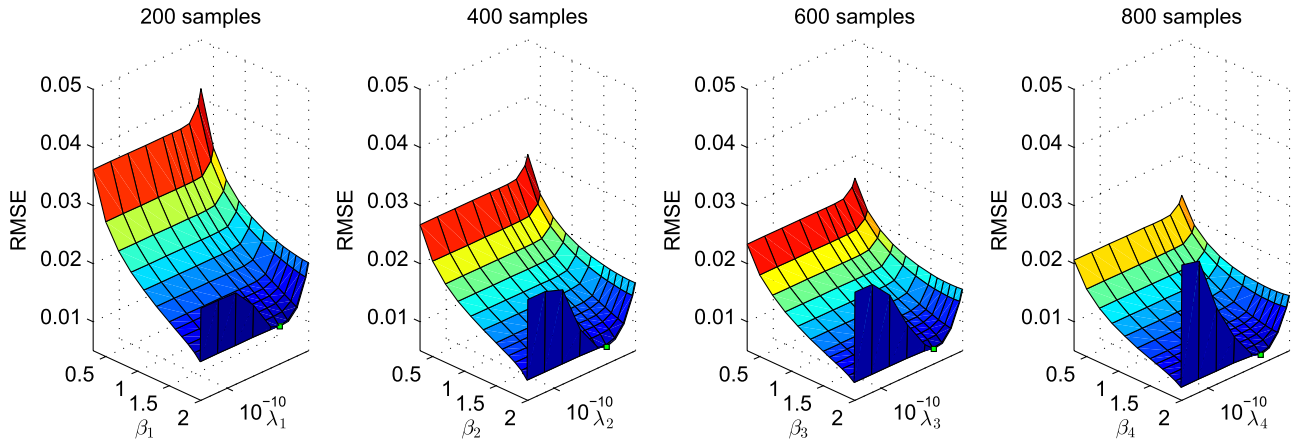

Fig. 6. Illustration of the mean RMSE (over 10 folds of parameter optimization) over the parameter search space for Pantone 12C measured camera response data and varying number of training samples for the logarithmic kernel. The optimal parameters (indicated as a green square) are $\beta_{1\ldots4} = 2$ and $\lambda_{1\ldots4} = 10^{-4}$.

terms of d_p leads to similar results ($d_p = 0.0004$), and the RMSE is slightly worse in case of the logarithmic kernel (RMSE $= 0.0054$ for logarithmic and RMSE $= 0.0050$ for Gaussian). The linear kernel function results are the worst, both spectrally and colorimetrically. The logarithmic kernel results in a slightly lower standard deviation for ΔE_{00} and d_p as compared with the Gaussian kernel. For RMSE, the Gaussian kernel has a slightly lower standard deviation. The linear kernel has the highest standard deviations for all metrics. The maximum ΔE_{00} and d_p error is lowest for the logarithmic kernel. The maximum RMSE is lowest for the linear kernel.

When it comes to HKS data, the Gaussian ($\Delta E_{00} = 0.67$), $\log(\Delta E_{00} = 0.73)$, and the linear ($\Delta E_{00} = 0.75$) kernel have a relatively close average color difference value. Spectrally, when considering mean d_p, the Gaussian kernel performs best ($d_p = 0.0006$). In terms of the mean RMSE, the linear kernel results in the smallest error (RMSE $= 0.0092$).

Overall, the mean estimation quality is approximately by a factor of 2 worse for the HKS data as compared with Pantone in case of the logarithmic and Gaussian kernel (in ΔE_{00} and RMSE). One reason for the overall worse results of HKS compared to Pantone data is the low number of training samples. We have shown similar trends in the case of measured Pantone data for the Gaussian and the logarithmic kernel in Subsection 5.B.

In Fig. 9, we illustrate an instance of estimated and measured spectral reflectances from the Pantone dataset obtained with the logarithmic kernel. As reference, results obtained from the linear and Gaussian kernel are also illustrated. Figure 9 (left) corresponds to the sample for which the lowest RMSE was obtained (RMSE $= 0.0006$). It can be seen that the estimated reflectance by the Gaussian kernel is only slightly worse (RMSE $= 0.0009$), but in case of the linear kernel much worse (RMSE $= 0.0036$). In Fig. 9 (middle), we illustrate the sample corresponding to highest RMSE for the logarithmic kernel. Here it can be observed that the estimated reflectances for the logarithmic, Gaussian, and linear kernel are quite similar but consistently deviate from the measured reflectance. It was identified that for this particular sample, measured camera responses and reflectance data do not correspond. The sample with the second highest RMSE for the logarithmic kernel (RMSE $= 0.0457$) is illustrated in Fig. 9 (right). For this sample, the linear kernel performs best (RMSE $= 0.0138$), followed by the Gaussian kernel (RMSE $= 0.0332$).

D. Estimation Performance with Simulated Sensor Responses from the 12-Channel System

A similar setting as in the previous subsection was used for the evaluation of simulated data. Namely, we compare results

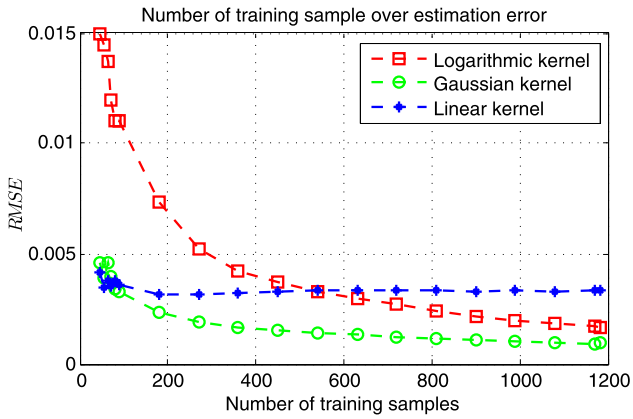

Fig. 7. Analysis of the influence of the number of training samples on the estimation performance (RMSE) for the logarithmic, Gaussian, and linear kernel function. Camera responses were simulated noiseless for the Pantone dataset.

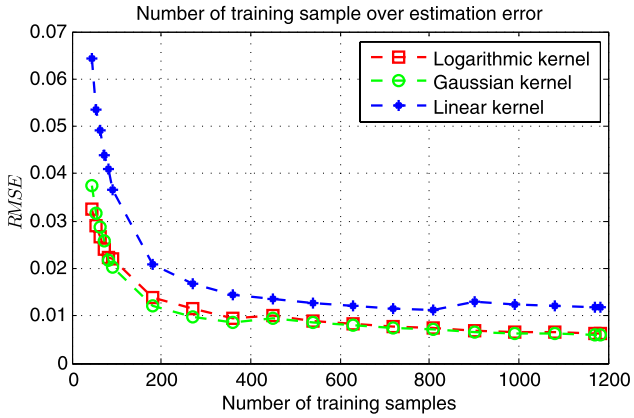

Fig. 8. Analysis of the influence of the number of training samples on the estimation performance (RMSE) for the logarithmic, Gaussian, and linear kernel function and measured camera responses of the Pantone dataset.

Table 1. Spectral Estimation Results for Measured Data of the Pantone and HKS Dataset and the 12-Channel Acquisition System, Logarithmic, Gaussian, and Linear Kernel

	Logarithmic Kernel				Gaussian Kernel				Linear Kernel			
Pantone Measure	Mean	Std	Max	p95	Mean	Std	Max	p95	Mean	Std	Max	p95
ΔE_{00}	0.33	0.26	4.59	0.73	0.33	0.28	4.62	0.73	0.68	0.50	5.02	1.46
d_p	0.0004	0.0011	0.0299	0.0016	0.0004	0.0015	0.0426	0.0016	0.0014	0.0048	0.0930	0.0044
RMSE	0.0054	0.0043	0.0984	0.0114	0.0050	0.0040	0.1001	0.0102	0.0086	0.0047	0.0974	0.0158
HKS Measured												
ΔE_{00}	0.73	0.74	4.56	2.22	0.67	0.58	3.25	2.10	0.75	0.55	3.34	1.73
d_p	0.0014	0.0028	0.0152	0.0064	0.0006	0.0014	0.0096	0.0029	0.0012	0.0025	0.0207	0.0052
RMSE	0.0129	0.0189	0.1451	0.0393	0.0106	0.0073	0.0380	0.0258	0.0092	0.0106	0.0763	0.0262

obtained with the logarithmic, Gaussian, and linear kernel for Pantone and HKS data of the 12-channel acquisition system using noiseless simulated camera response data. The numerical results are illustrated in Table 2.

For Pantone data and comparison of the logarithmic kernel performance, the following results are found: the logarithmic and Gaussian kernel have the same mean ΔE_{00} and d_p error ($\Delta E_{00} = 0.01$, $d_p = 0.0001$), the RMSE is lower in case of the Gaussian kernel. The linear kernel performs worst on average.

The results for the HKS dataset are different from the Pantone data. Here, the logarithmic kernel performs considerably worse than the Gaussian kernel, which gives the best, and the linear kernel with intermediate results. This finding applies to the mean, standard deviation, maximum, and the 95th percentile of the spectral and colorimetric error measures. It seems that the poor performance of the logarithmic kernel is due to the relatively small training set used.

For the comparison of measured and simulated data results (Tables 1 and 2), we concentrate on the Pantone dataset, which does not inherit the limitation of the number of training samples. Overall, a large discrepancy can be identified for all kernel functions.

The increase in estimation performance from one kernel function to another can be up to 48% (comparing mean ΔE_{00} between the linear and the logarithmic or the Gaussian kernel for Pantone data in Table 1).

E. Influence of the Number of Sensor Responses

We divided the 12 channels into three different sets with three, six, and 12 channels, respectively, in order to investigate the influence of the number of sensor spectral channels in the estimation. The spectral responsivities of the resulting systems are illustrated in Fig. 3. It has to be mentioned that the corresponding three- and six-channel systems are far from ideal for the task of spectral recovery. However, using measured data from the same system allows us to analyze the influence of the number of channels on the estimation performance of the logarithmic kernel without the potential effect of differences that are introduced by using measured data from different estimation systems, such as different noise and nonlinearity behaviors.

The numerical results of the comparison are illustrated in Table 3. Pantone 12-channel measured data results have already been presented in Table 1, but are unified in Table 3

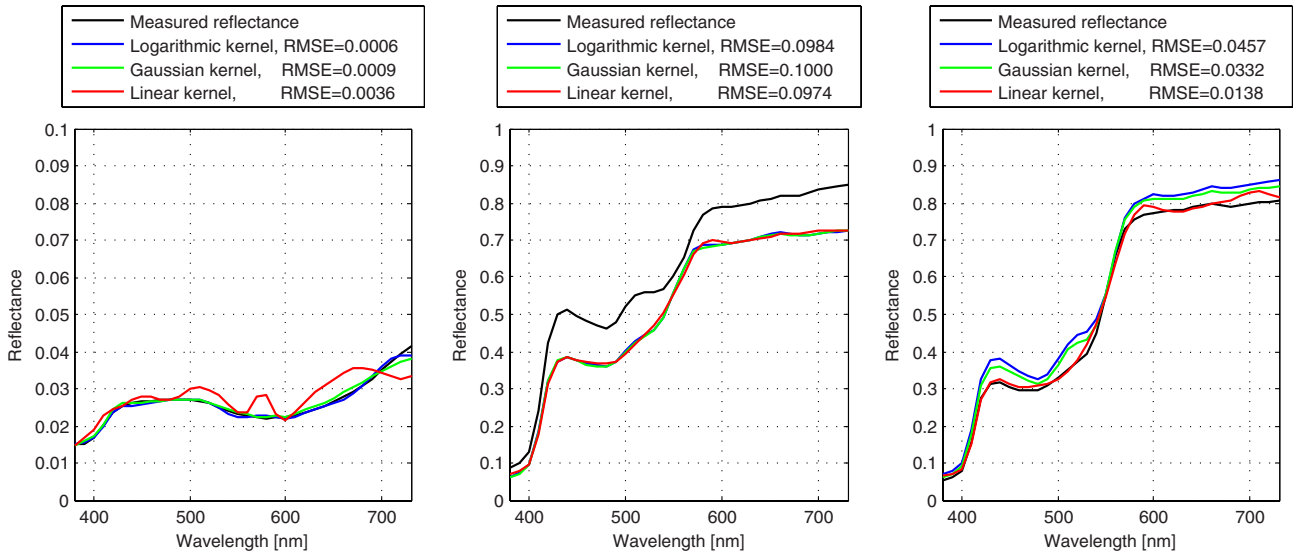

Fig. 9. Estimation results for the Pantone dataset and measured camera responses: sample reflectance with lowest (left) and highest (middle) RMSE error for the logarithmic kernel and corresponding estimated spectra of the linear and Gaussian kernel. The plot on the right illustrates the sample corresponding to the 2nd highest RMSE for the logarithmic kernel.

Table 2. Spectral Estimation Results for Simulated Data of the Pantone and HKS Dataset and 12-Channel Acquisition System, Logarithmic, Gaussian, and Linear Kernel

	Logarithmic Kernel				Gaussian Kernel				Linear Kernel			
Pantone Simulated	Mean	Std	Max	p95	Mean	Std	Max	p95	Mean	Std	Max	p95
ΔE_{00}	0.01	0.01	0.14	0.04	0.01	0.01	0.19	0.04	0.03	0.03	0.26	0.08
d_p	0.0001	0.0001	0.0016	0.0001	0.0001	0.0001	0.0018	0.0001	0.0002	0.0008	0.0158	0.0008
RMSE	0.0017	0.0019	0.0192	0.0052	0.0009	0.0011	0.0233	0.0022	0.0033	0.0021	0.0169	0.0073
HKS Simulated												
ΔE_{00}	0.28	0.43	3.62	0.84	0.03	0.03	0.17	0.08	0.06	0.05	0.36	0.15
d_p	0.0005	0.0014	0.0113	0.0023	0.0001	0.0003	0.0025	0.0004	0.0004	0.0011	0.0097	0.0016
RMSE	0.0074	0.0091	0.0579	0.0205	0.0030	0.0033	0.0172	0.0093	0.0049	0.0046	0.0328	0.0148

Table 3. Comparison of the Number of Channels for Measured Data of the Pantone Dataset, the Logarithmic, Gaussian, and Linear Kernel

	Logarithmic Kernel				Gaussian Kernel				Linear Kernel			
12-Channel Measured	Mean	Std	Max	p95	Mean	Std	Max	p95	Mean	Std	Max	p95
ΔE_{00}	0.33	0.26	4.59	0.73	0.33	0.28	4.62	0.73	0.68	0.50	5.02	1.46
d_p	0.0004	0.0011	0.0299	0.0016	0.0004	0.0015	0.0426	0.0016	0.0014	0.0048	0.0930	0.0044
RMSE	0.0054	0.0043	0.0984	0.0114	0.0050	0.0040	0.1001	0.0102	0.0086	0.0047	0.0974	0.0158
6-Channel Measured												
ΔE_{00}	0.43	0.33	4.96	0.96	0.44	0.35	5.81	0.99	0.72	0.53	5.11	1.74
d_p	0.0006	0.0015	0.0222	0.0031	0.0006	0.0014	0.0218	0.0030	0.0029	0.0060	0.0988	0.0124
RMSE	0.0068	0.0049	0.0975	0.0148	0.0065	0.0048	0.0988	0.0136	0.0146	0.0081	0.1003	0.0304
3-Channel Measured												
ΔE_{00}	1.50	1.34	12.10	4.29	1.49	1.33	12.22	4.26	2.66	2.41	23.69	7.12
d_p	0.0055	0.0115	0.1552	0.0196	0.0054	0.0114	0.1581	0.0200	0.0173	0.0354	0.3098	0.0845
RMSE	0.0223	0.0183	0.1119	0.0626	0.0223	0.0182	0.1141	0.0612	0.0365	0.0234	0.1618	0.0836

with the results of the three- and six-channel system to allow a more clear illustration.

The general trend found in other experiments with 12 channels (i.e., the lower performance of the linear kernel in comparison with the Gaussian and logarithmic kernels) can also be identified for six or three channels.

The logarithmic and Gaussian kernel, on the other hand, perform quite similar. When considering color measurement accuracy, the 12-channel as well as the six-channel case give average ΔE_{00} errors for all three kernel functions smaller than one unit, which is an error that is close to the just noticeable difference. For the three-channel case and all three kernel functions, the average ΔE_{00} errors are above one unit ($\Delta E_{00} = 1.50$ for the logarithmic kernel, $\Delta E_{00} = 1.49$ for the Gaussian kernel, and $\Delta E_{00} = 2.66$ for the linear kernel). From all metrics, it can be seen that the error increases less when comparing 12 and six channels and more when comparing six and three channels.

6. DISCUSSION AND CONCLUSION

In this work, we introduce the logarithmic kernel function to kernel ridge regression. Specifically, we concentrate on the problem of spectral estimation, in which spectral reflectance values are to be estimated from few camera responses of a multispectral image-acquisition system. The logarithmic kernel belongs to the class of CPD kernel functions, which require an additional parametric part to be included in the regression model.

We considered two datasets (Pantone, 1314 samples; HKS, 91 samples) in our evaluations and a 12-channel multispectral imaging system. The results based on measured camera responses for the Pantone dataset indicate that the logarithmic kernel outperforms the linear kernel and leads to comparable colorimetric and spectral estimation performance with the Gaussian kernel. When considering the smaller HKS dataset, the logarithmic kernel had the worst performance.

In further analysis on the larger Pantone dataset, it was identified that the logarithmic kernel requires more training data to perform similar or better than the linear or Gaussian kernel, when simulated noiseless camera responses are considered. For measured camera response data, logarithmic and Gaussian kernel performance is similar, whereas the linear kernel performs worse for any number of training samples evaluated.

We believe that the influence of the number of training samples can explain the relatively poor performance of the logarithmic kernel (compared to other kernels), which was found for the HKS set when noiseless simulated camera responses were considered. However, the results for measured camera response data of the HKS set were close to the results obtained with the Gaussian kernel.

For the data considered in this work, it was identified that the parametrization of the logarithmic kernel can be simplified, as the power term β of the logarithmic kernel did not

influence the selection of the regularization parameter λ of the regression. This allows us to search for the optimal parameter sequentially rather than in grid search for all combinations of parameters in the parameter search space. For our data, we found that $\beta = 2$ had the best performance.

We also investigated the influence of the number of sensor spectral channels of the acquisition system on the estimation performance. In comparison were three, six, and 12 channels and measured camera response data of the Pantone set. We observed that the logarithmic and Gaussian kernel performance is very close and compared to the linear kernel, which is significantly higher for any number of channels. Further, the colorimetric performance of the 12-or 6-channel system seems to be suitable for color measurements, whereas the three-channel system could be considered for applications where less color precision is required.

ACKNOWLEDGMENTS

The authors thank the reviewers for valuable comments and suggestions that helped us to improve the quality of this work. This study was supported by Chromasens GmbH through UGR grant number 2936 and by the Spanish Ministry of Research and Innovation through grant number DPI2011-23202. We further thank Chromasens GmbH for providing the data for the empirical part of this studies.

REFERENCES

1. B. Schölkopf and A. Smola, *Learning with Kernels* (MIT, 2002).
2. J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge University, 2004).
3. M. G. Genton, "Classes of kernels for machine learning: a statistics perspective," J. Mach. Learn. Res. **2**, 299–312 (2002).
4. V. Heikkinen, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, T. Jääskeläinen, and S. D. Lee, "Regularized learning framework in the estimation of reflectance spectra from camera responses," J. Opt. Soc. Am. A **24**, 2673–2683 (2007).
5. 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).
6. T. Eckhard, E. M. Valero, and J. Hernández-Andrés, "A comparative analysis of spectral estimation approaches applied to print inspection," in *Proceedings of the Annual German Colour Group Meeting* (IDD TU Darmstadt, GRIS TU Darmstadt, IGD Fraunhofer, 2012), pp. 13–24.
7. W.-F. Zhang, P. Yang, D.-Q. Dai, and A. Nehorai, "Reflectance estimation using local regression methods," Lect. Notes Comput. Sci. **7367**, 116–122 (2012).
8. V. Heikkinen, "Kernel methods for estimation and classification of data from spectral imaging," Dissertations in Forestry and Natural Sciences (University of Eastern Finland, 2011).
9. S. Boughorbel, J.-P. Tarel, and N. Boujemaa, "Conditionally positive definite kernels for svm based image recognition," in *Proceedings of IEEE Conference on Multimedia and Expo* (IEEE, 2005), pp. 113–116.
10. C. Berg, J. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups Theory of Positive Definite and Related Functions* (Springer, 1984).
11. B. Schöllkopf, "The kernel trick for distances," in *Proceedings of Advances in Neural Information Processing Systems*, T. K. Leen, ed. (MIT, 2001), pp. 301–307.
12. M. Schnitzlein and M. Hund, "Chromasens GmbH," http://www.chromasens.de/en.
13. R. G. Lyons, *Understanding Digital Signal Processing* (Prentice Hall, 2004).
14. T. Eckhard, "Color imaging web page of the University of Granada," http://www.ugr.es/local/colorimg/suppl_docs/log_kernel.html.
15. F. H. Imai, M. R. Rosen, and R. S. Berns, "Comparative study of metrics for spectral match quality," in *CGIV, First European Conference on Colour in Graphics, Imaging and Vision* (Society for Imaging Science and Technology, 2002), pp. 492–496.
16. J. A. S. Viggiano, "Metrics for evaluating spectral matches: a quantitative comparison," in *Second European Conference on Colour Graphics, Imaging and Vision* (Society for Imaging Science and Technology, 2004), pp. 286–291.
17. J. Romero, A. García-Beltrań, and J. Hernández-Andrés, "Linear bases for representation of natural and artificial illuminants," J. Opt. Soc. Am. A **14**, 1007–1014 (1997).
18. G. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed. (Wiley, 1982).
19. "Improvement to industrial color-difference evaluation," CIE Technical Report, Pub. No. 142 (CIE, 2001).
20. J. Schanda, *Colorimetry: Understanding the CIE System* (Wiley, 2007).