EXTENSION AND EVALUATION OF SELECTION CRITERIA FOR THE ESTIMATION OF SPECTRAL DATA

P. Carvalho, L. Mendes, P. Martins, A. Santos

Centre of Informatics and Systems, University of Coimbra
Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal
Email: {carvalho, lgmendes, pjmm, amancio}@dei.uc.pt

ABSTRACT

Spectral data estimation from image data is an ill-posed problem since (i) due to the integral nature of imaging sensors, the same output can be obtained from an infinity of input signals and (ii) color signals are spectrally smooth in nature and, therefore, limit the number of linear independent data than can be collected. To enable the solution of these problems the solution's search space has to be constrained. The question that arises is how to select/parameterize these constraints? In this paper several model selection criteria are extended for spectral data estimation and evaluated in the context of spectral sensitivity function estimation of CCD sensors.

***Index Terms*—** Imaging, Color, Machine Learning

1. INTRODUCTION

Ill-posed modeling problems are frequently found in several image processing and computer vision domains. Multi-spectral data estimation from low dimensional imaging device responses are typical ill-posed modeling problems whose solution is important or even fundamental in several computer vision and image processing operations, such as demosaicking [1], color constancy [2] and color space mapping [3]. Usually, it involves the estimation, for each wavelength λ, of some data distribution $X(\lambda)$. Namely, let $I(\lambda)$ be the spectral power distribution (SPD) of the sensor's excitation signal, $S(\lambda)$ the sensor's spectral sensitivity and $b \propto \int I(\lambda) S(\lambda)\, d\lambda$ be the imaging device response (it is assumed the image has been corrected for radiometric distortions and static non-linearities). For spectral sensitivity estimation, given a set of device responses and known SPDs of excitation signals, one wants to estimate $S(\lambda)$, while for high-dimensional spectral signals estimation the goal is to estimate $I(\lambda)$ given the device's responses and sensitivities $S(\lambda)$. From the image formation equation it is observed that this is an ill-posed problem, since the imaging sensor performs space reduction through integration over wavelengths. Furthermore, natural colors can be accurately approximated with just a few (typically between 3 and 9) basis functions [4]. Therefore, (i) the same device output can be produced by an infinite number of stimuli and (ii) only a limited set of linear independent data may be collected for the estimation task. Fortunately, there are some assumptions that can be made to constrain the problem. For image sensor sensitivity estimation, the most commonly applied constraints are the positivity of the sensor's spectral sensitivities and the smoothness of the sensitivity function [5][2][4]. There are several alternative strategies to account for smoothness: (i) Sharma *et al.* [6] impose an upper bound on the second derivative of the solution, while (ii) other authors [2][5] apply a Tikhonov formulation where a regularization term is added to the object function, i.e. let $x \in \mathbb{R}^n$ be a discrete version of $X(\lambda)$ such that $x_i \equiv X(\lambda_i)$, $\lambda_i = \lambda_0 + (i-1)\Delta\lambda$, $i = 1...n$, and $\Delta\lambda$ is the sampling interval, then x can be computed from

$$\min \left\{ \frac{1}{m} \|Ax - b\|^2 + \frac{\alpha}{\Delta\lambda^4} \|Dx\|^2 \right\} \tag{1}$$

$$\text{subject to } Cx \leq h \tag{2}$$

where $Ax = b$ ($A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $D \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{q\times n}$, $H \in \mathbb{R}^q$) are the m equations (usually $m \ll n$) that can be obtained from the sensor's outputs, and $\alpha \in \mathbb{R}^+$ is a regularization gain that controls the trade-off between the roughness of the solution as measured by $\|Dx\|^2$ ($\Delta\lambda^{-2} Dx$ approximates the second derivative of x) and the infidelity to the data as measured by $\|Ax - b\|^2$.

The solution to (1) is equivalent to a Butterworth low-pass filter where α controls the cutoff frequency. This means that the attainable solution is band-limited and therefore can be modeled using the d dimensional set of functions $f(z, \alpha) = \sum_{k=1}^{d} \alpha_k \varphi_k(z)$, where $\{\varphi_k\}$ form a orthonormal basis. For instance, taking a Fourier basis, it is seen that $X(\lambda) = x_0 + \sum_{w=1}^{k} x_w \cos(w\varpi) + y_w \sin(w\varpi)$ for a domain $\lambda_0 \leq \lambda < \lambda_v$ and $\varpi \equiv 2\pi(\lambda - \lambda_0)/(\lambda_v - \lambda_0)$. In this formulation x can be computed from (3) subject to (2), such that $A \in \mathbb{R}^{m\times d}$, $d = 2k + 1$.

$$\min \left\{ \frac{1}{m} \|Ax - b\|^2 \right\} \tag{3}$$

This project was partially financed by FCT/PRAXIS XXI (project praxis/36564/99) and by the POSI Program, supported by the Portuguese Government and the European Union.

This idea has been explored by several authors. Hardeberg [7] utilizes a truncated pseudo-inverse to constrain smoothness, whereas Finlayson *et al.* [4] suggested using (3) directly. Solving (3) provides a more compact representation of $X(\lambda)$ and lowers significantly the complexity of the algorithm. This is significant for SPD estimation problems using low dimensional imaging device outputs where, for each cluster of similar pixels, one has to solve (3) or (1).

In the aforementioned spectral estimation algorithms the constraints are defined based on exact *a priori* knowledge on x, being the solution very dependant on these values. To adaptively identify the optimal set of constraints, the expected empirical risk or equivalently the bias-variance trade-off should be minimized. Unfortunately, the empirical risk is only defined assuming a parametric formulation for the problem, which in practical applications is usually not possible, since the exact noise variance is unknown. In order to solve this, several indirect selection criteria may be applied, such as the Generalized Cross-Validation (GCV), the Bayes Information Criterion (BIC) or the Akaike Information Criterion (AIC). These are known to asymptotically approach the expected empirical risk. In [5] we have developed a data driven algorithm to identify the best set of constraints using the formulation in (1) and a modified GCV criterion. More recently we introduced a less complex solution to the problem using an adapted GCV criterion and the formulation in (3) [8]. It is well known that the expectation efficiency (ratio between the global optimum and the achieved solution) of these criteria approach 1 when the number of observation is very large (theoretically ∞). For limited number of observations, these criteria tend to exhibit different degrees of bias. In this paper, some of the most significant selection criteria presented in literature are adapted to the problem defined in (3) subject to (2). Their efficiency is tested using sensitivity estimation of CCD sensors in RGB cameras.

The paper is organized as follows: in section 2 our data-driven sensitivity estimation algorithm reported in [8] is briefly outlined. In section 3 the model selection criteria are adapted. Experimental results using these criteria are discussed in section 4. Finally, in section 5 some main conclusions are presented.

2. THE DATA-DRIVEN ALGORITHM

In our method a similar formulation as in [4] is applied. Namely, using a Fourier basis, (3) is minimized subject to positivity constraints, i.e., $X(\lambda) \geq 0$. Modality constraints have to be applied in order to avoid rapid oscillations between peaks of the sensitivity function. For example, an uni-modal sensitivity function with a peak at wavelength $\lambda = \lambda_p$ can be expressed as a set of linear constraints as in (4).

$$\begin{aligned} X(\lambda_{i+1}) &\geq X(\lambda_i), i = 0, ..., p-1 \\ X(\lambda_{i+1}) &\leq X(\lambda_i), i = p, ..., v-2 \end{aligned} \tag{4}$$

In order to constrain the search space for the number of peaks in $X(\lambda)$ and their location, we observed that formulating the estimation problem as in (1) subject to positivity constraints, can be equivalently computed for p active constraints from

$$\min_{\widetilde{s}} \left\{ \left\| \widetilde{A}\widetilde{s} - b \right\|^2 + \gamma \left\| \widetilde{D}\widetilde{s} \right\|^2 \right\} \tag{5}$$

where $\widetilde{A} \in \mathbb{R}^{m \times (n-p)}$, $\widetilde{D} \in \mathbb{R}^{(n-p-1)\times(n-p)}$, if $m \leq n$ and $p < n$. Further, $\widetilde{s}_i(\gamma)$ is obtained by (6), where $c = U^T b$, $\widetilde{A} = U\Sigma Z$ and $\widetilde{D} = V\Omega Z$ are the GSVD decomposition of matrixes $\widetilde{A}$ and $\widetilde{D}$. Finally $\Sigma = \begin{pmatrix} 0 & D_M \end{pmatrix} \in \mathbb{R}^{m\times(n-p)}$, $\Omega = \begin{pmatrix} D_B & 0 \end{pmatrix}^T \in \mathbb{R}^{(n-p+1)\times(n-p)}$, $D_M = diag(\alpha_1, ..., \alpha_m)$ and $D_B = diag(\beta_1, ..., \beta_{n-p})$, $1 \geqslant \beta_1 \geqslant ... \geqslant \beta_{n-p} \geqslant 0$. It can be shown that the solution to (5) is

$$\widetilde{s}_i(\gamma) = \sum_{j=n-p-m+1}^{n-p} \frac{\alpha_{j-n-p+m} Z_{i,j}^{-1} c_j}{\alpha_{j-n-p+m} + \gamma\beta_j^2}, i = 1, ..., n-p \tag{6}$$

From (6) it is seen that for large values of the regularization gain γ, the solution $\widetilde{s}_i(\gamma)$ will be dominated by the linear combination of a small set of terms, those where $\beta_j \approx 1$. The influence of these terms will be persistent and, therefore, $\widetilde{s}_i(\gamma)$ will tend to decrease/increase monotonically as γ increases. Therefore, it is observed that if $\widetilde{s}_i(\gamma)$ is a local maximum, then $\widetilde{s}(\gamma + \Delta\gamma)$, $\Delta\gamma > 0$, will also tend to be a local maximum. From this observation, the strategy for constraining the search space is straightforward to define: compute $s(\gamma)$ using a set of regularization gains $\gamma_1 < \gamma < \gamma_t$, such that $\widetilde{s}(\gamma_1)$ is a smooth solution (typically a set of 4 to 10 distinct γ). From these solutions persistent and pronounced peaks can be easily identified to formulate the modality constraints. For a fully automatic algorithm a clustering technique based on an oscillation measure $O(\gamma)$ (e.g. the second derivative) of the solution may be applied to compute the range of necessary regularization gains (in this case the smallest applied γ should lead to a highly oscillating solution, to enable the identification of at least 2 clusters). In our implementation of the algorithm, for each identified peak at wavelength λ_p, the search space is limited to $\lambda_p \pm g\Delta\lambda$,$g = 0, 1, ...$ As for the search space of the model order, we apply $k = 2, ..., \lfloor \frac{m-1}{2} \rfloor$. Finally, for each pair (k, λ_p) the solution to (3) subject to the linear constraints is computed. The optimal pair (k^*, λ_p^*) is identified by the solution that minimizes the selection criterion.

3. SELECTION CRITERIA

In estimation it is observed that data as well as model constraints are information sources for the modelling problem. Data represent the underlying process without bias. However, it is contaminated with noise, hence exhibiting uncertainty. On the other hand, the model enables to reduce the uncertainty of the phenomena, but it tends to induce bias if wrongly chosen. Therefore, in these type of problems one has to minimize the bias-variance trade-off, which, unfortunately, can not be directly computed from the empirical risk defined as the optimization criterion, but rather from the expected risk. In practical applications, this optimization has to be performed using indirect selection criteria, since the exact degree of uncertainty is unknown. These selection criteria are usually formulated for estimation problems of the form defined in (1) or (3) without additional constraints. Using the active set theory, it is observed that these criteria may be adapted using the following result:

Theorem 1 *Let $\widehat{x}$ be the vector that minimizes $\|Ax-b\|^2$ subject to $Cx \leq h$, where $A \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{s\times n}$. Let $C^*\widehat{x} = h^*$ be the active set of constraints, such that $C^* \in \mathbb{R}^{p\times n}$ and $rank\left(C^*\right) = p < n$. The same solution may be obtained using*

$$\min_{w_2} \left\|Hw_2 - \widetilde{b}\right\|^2$$

where $\widehat{x} = Z^{-1}\left(\widehat{w}_1^T\ \widehat{w}_2^T\right)^T$, $\widehat{w}_1 = D_C^{-1}V^T h^* \in \mathbb{R}^p$, $H = U\begin{pmatrix} 0 & D_{M2}^T & 0\end{pmatrix}^T$ and $A = U\begin{pmatrix} D_M^T & 0\end{pmatrix}^T Z$, $C^* = V\begin{pmatrix} D_C^T & 0\end{pmatrix}^T Z$ are the Generalized Singular Value Decompositions of matrixes A and C^*. Finally,

$$D_M \equiv \begin{pmatrix} D_{M1} & 0 \\ 0 & D_{M2}\end{pmatrix}, D_{M1} \in \mathbb{R}^{p\times p}$$

and $\widetilde{y} = y - D_{M1}D_C^{-1}V^T h^*$.

Proof 1 *Proof is a consequence of the GSVD of matrixes A and C^* and by noting that the empirical risk is not altered by changing to an alternative orthonormal basis.*

It should be noted that if C^* is not full line rank, then the singular value decomposition of C^* may be applied to transform C^* and h^* appropriately. Using theorem 1, the optimal constraints (model order, peak number and their wavelength location, for the problem defined in section 2) may be identified using the most commonly applied selection criteria. In this section, the following selection criteria will be extended: GCV, BIC and Akaike's FPE. An extension to Akaike's FPE for small sample regression is also considered [9]. Finally, AIC will not be considered, since it asymptotically approaches GCV.

GCV utilizes the leave-one-out principle. The idea is to minimize $\frac{1}{m}\sum_{k=1}^{m}\left(A_k x^{[k]} - b_k\right)^2$, where $x^{[k]}$ is the achieved solution using all but the kth data point (hence $A_k x^{[k]} - b_k$ represents the prediction error of the kth data point using solution $x^{[k]}$). It can be shown using theorem 1, that this average prediction error may be computed in close form using (7) (see [8]).

$$GCV = \frac{\frac{1}{m}\|b-Ax\|^2}{\left(\frac{1}{m}trace\left(I - AZ_2U_2\right)\right)^2} \quad (7)$$

$$U^T = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} \begin{matrix} \{ & p \\ \{ & n-p \\ \{ & m-n \end{matrix}, Z^{-1} = \begin{pmatrix} \overbrace{Z_1}^{p} & \overbrace{Z_2}^{n-p} \end{pmatrix}$$

BIC may be interpreted using the a posteriori likelihood maximization, i.e. $max\left\{P\left(d, \lambda_{p1}, ..., \lambda_{pr}|Y\right)\right\}$. Assuming $b = Ax + \epsilon$, $\epsilon \sim N\left(0, \sigma^2 I\right)$, it can be shown using theorem 1 and Schwartz's work that (let $\widehat{\sigma}^2$the estimate of the variance in the likelihood sense, i.e. $\widehat{\sigma}^2 \equiv \frac{1}{m}\left\|H\widehat{w}_2 - \widetilde{b}\right\|^2$, and $P\left(d, \lambda_{p1}, ..., \lambda_{pr}\right)$ be the prior distribution on the model's order and the sensitivity peak locations)

$$BIC = m\ln\widehat{\sigma}^2 + (n-p)\ln m - 2\ln P\left(d, \lambda_{p1}, ..., \lambda_{pr}\right) \quad (8)$$

Regarding the Akaike's Future Prediction Error, it is observed that under the conditions assumed in this section it may be formulated as in (9).

$$FPE = \left\|H\widehat{w}_2 - \widetilde{b}\right\|^2\left(1 + \frac{n-p}{m}\right)\left(1 - \frac{n-p}{m}\right)^{-1} \quad (9)$$

In [9], Chapelle and co-works have introduced a redefinition of the penalization therm of the FPE in order to adapt it for situations when a small set of data samples is available. In this situation the problem's dimension is directly estimated from the eigenvalues of the covariance matrix, i.e. $q = E\sum_{i=1}^{n-p}\frac{1}{\xi_i}$, where ξ are the eigenvalues of the covariance matrix. Using this modification and (9), (10) follows.

$$FPEVap = \left\|H\widehat{w}_2 - \widetilde{b}\right\|^2\left(1 - \frac{n-p}{m}\right)^{-1} \times \left(1 + \frac{E\sum_{i=1}^{n-p}\frac{1}{\xi_i}}{m}\right) \quad (10)$$

4. RESULTS

In this section some results for spectral sensitivity estimation in common CCD sensors applied in RGB cameras are presented. In order to measure the performance of the described model selection measures in spectral estimation problems a simulation program was developed. The shown test results are (i) for an asymmetrical Gaussian model for the spectral sensitivities (these are typical sensitivity curves for some cameras such as the Sony DXC-930 color video camera [2] - see fig. 1 right) and (ii) for the spectral sensitivity curves from a Kodak DCS200 camera as described in [1] - see fig. 1 left. These two types of sensitivity functions were chosen to evaluate the method's performance for curves with distinct smoothness and modality. In these tests 24 ($m = 24$) patches of the MacBeth-Color Checker map were applied and g was limited to $g = 0, 1, ..., 10$. The sampling step was fixed to $\Delta\lambda = 2nm$, $\lambda_0 = 400nm$, $\lambda_v = 700nm$. Table 1 summarizes the achieved results. In this table η represents the estimation efficiency, i.e. $\eta \equiv \frac{\left\|x^{\text{real}} - \widehat{x}\right\|^2}{\left\|x^{\text{real}} - \widehat{x}^*\right\|^2}$ where x^{real} represents

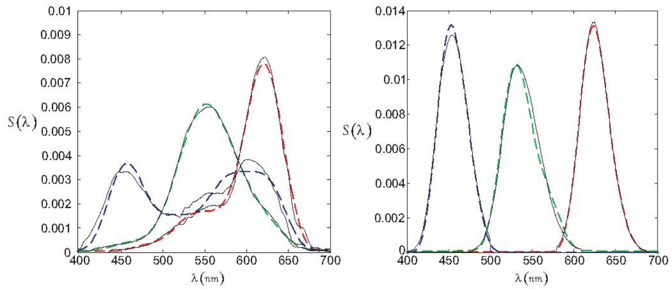

Fig. 1. Real (solid line) vs. estimated (dashed line) spectral sensitivities for RGB sensors using BIC. (left) Sony DXC-930. (right) Kodak DCS 200.

$Average$	$Criterion$	η	$k-k^*$	$\lambda_p-\lambda_p^*$
	GCV	2.73	-1.33	1.00
	BIC	1.60	-0.83	0.67
	FPE	2.01	-1.00	1.00
	FPEVap	2.41	-0.67	1.33
Max	GCV	6.37	2	9
	BIC	2.50	2	11
	FPE	3.79	2	10
	FPEVap	3.79	2	10

Table 1. Results: maximum and average values.

the real function and $\widehat{x}^*$ represents the best achievable solution using the specified estimation process. Further results shown are the deviation from the ideal model order $(k-k^*)$ and the error in peak estimation $\left(\lambda_p-\lambda_p^*\right)$. Finally, since no prior knowledge was assumed, the prior $P\left(d,\lambda_p\right)$ for BIC was modeled as an uniform distribution.

From table 1 it is observed that the best and the most consistent estimation results are achieved using BIC criterion. In fact, in average terms its efficiency is 20% less compared to FPE and FPEVap and almost 42% lower compared to GCV. Furthermore, the accuracy concerning peak detection as well as model order selection using BIC is also higher compared to the other selection criteria. The obtained results show that the BIC criterion is the one that induces less estimation bias, although it is closely followed by Akaike's FPE. FPEVap is the one that is most consistent in selecting the appropriate model order, but it exhibits low performance in selecting other constraints, which leads to a degradation in estimation efficiency. Finally, the worst performance is achieved by GCV.

5. CONCLUSIONS

This paper discusses model selection criteria in the context of spectral estimation problems. Several selection criteria are extended for this purpose and applied to CCD spectral sensitivity estimation with unknown smoothness and maxima locations. The obtained results suggest that the BIC criterion is the one that induces less estimation bias, although it is closely followed by Akaike's FPE. This is an interesting result, since BIC exhibits a natural way for integrating prior knowledge by modelling the prior distribution of constraints. This could have an important role for SPD estimation of neighboring pixels.

6. REFERENCES

[1] P. Vora, J. Farrell, J. Tietz, and D. Brainard, "Digital color cameras - 2 - spectral response," Tech. Rep. HPL-97-53, Hewlett-Packard Lab., 1997.

[2] K. Barnard and B. Funt, "Camera calibration for color research," in *Proc. SPIE Electronic Imaging IV*, 1999.

[3] F. Martinez-Verdu, J. Pujoe, and P Capilla, "Characterization of digital camera as an absolute tristimulus colorimeter," *Imaging Science and Technology*, vol. 47, no. 4, pp. 279–295, 2003.

[4] G. Finlayson, S. Hordley, and P. Huble, "Recovering device sensitivity with quadratic programming," in *IST/SID's Color Imaging Conf.: Color Science, Systems and Applications*, 1998, pp. 90–95.

[5] P. Carvalho, A. Santos, B. Ribeiro, and A. Dourado, "Learning spectral calibration parameters for color inspection," in *Int. Conf. On Computer Vision*, 2001, pp. 660–667.

[6] G. Sharma and H. Trussel, "Characterization of scanner sensitivity," in *Color Imaging Conf.: Transformations and Transportability of Color*, 1993, pp. 103–107.

[7] J. Hardeberg, *Acquisition and reproduction of colour images:colorimetric and multispectral approaches*, Ph.D. thesis, Ecole Nationale Superieure des Telecommunications, 1999.

[8] P. Carvalho, A. Santos, and P. Martins, "Recovering imaging device sensitivities: a data-driven approach," in *Int. Conf. On Image Processing*, 2004, pp. 2411–2414.

[9] O. Chapelle, V. Vapnik, and Y. Bengio, "Model selection for small sample regression," *Machine Learning*, vol. 48, no. 1, pp. 9–23, 2001.