Statistical Linear Destriping of Satellite-Based Pushbroom-Type Images

Hervé Carfantan, *Member, IEEE*, and Jérôme Idier

***Abstract*—This paper introduces a new self-calibration destriping technique for pushbroom-type satellite imaging systems. Self-calibration means that no specific training data are required. It is based on the statistical estimation of each detector gain from the observed image, assuming a linear response. Both theoretical and practical behaviors are studied. Our technique is shown to outperform simpler techniques based on column averages in terms of gain estimation precision while keeping the computational cost within admissible limits.**

***Index Terms*—Calibration, estimation, gain measurement, image restoration, image sensors, radiometry.**

I. Introduction

STRIPING is a well-known phenomenon that arises in a large number of spaceborn multidetector imaging instruments. The main cause of such errors is a difference in the response of the sensors. Despite the prelaunch calibration phase, in-flight calibration often remains a prerequisite step for accurate image analysis.

The striping effect has different characteristics, depending on the instrument scanning of the scene and of the number of detectors. Some instruments, such as Landsat Thematic Mapper (TM) and MultiSpectral Scanner (MSS), operate optomechanically. The charge-coupled device (CCD) array is placed along-track and scans forward and reverse across-track using all the detectors at the same time. Therefore, due to the motion of the satellite, striping appears across-track with a period corresponding to the number of detectors. Other instruments, such as High Resolution Visible (HRV) onboard the Satellite Pour l'Observation de la Terre (SPOT) and Advanced Land Imager (ALI) and Hyperion onboard EO-1 (Earth Observation-1), rather rely on the "pushbroom" technique, where a whole image line is recorded using an array placed across-track. The other dimension is obtained thanks to the motion of the satellite along its orbit, as illustrated in Fig. 1. In the latter case, the striping effect is not periodic. Note that nonperiodic striping also arises in images from instruments with only one detector (such for the Very High Resolution Radiometer (VHRR) of the NOAA weather satellite system), which is caused by random fluctuations in the sensor response.

Manuscript received February 25, 2009; revised August 24, 2009. First published November 24, 2009; current version published March 24, 2010. This work was supported in part by Centre National d'Etudes Spatiales.

H. Carfantan is with the Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, Centre National de la Recherche Scientifique, 31400 Toulouse, France (e-mail: Herve.Carfantan@ast.obs-mip.fr).

J. Idier is with the Institut de Recherche en Communications et Cybernétique de Nantes, École Centrale de Nantes, 44321 Nantes Cedex 3, France (e-mail: Jerome.Idier@irccyn.ec-nantes.fr).

Digital Object Identifier 10.1109/TGRS.2009.2033587

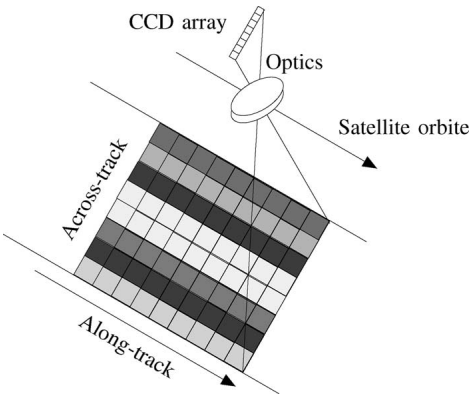

Fig. 1. Image formation of pushbroom-type satellite imaging instrument.

Striping may perturb the qualitative analysis of images. It can also introduce strong biases in the outcome of posterior processing aimed at extracting quantitative parameters, such as those based on geometric structure detection, image segmentation, and image cross correlation. Many destriping methods have been proposed in the literature. Most of them adopt either a filtering or a statistical approach (see, e.g., [1]).

The filtering approach has mainly been proposed to reduce periodic striping in images. Filtering in the spatial frequency domain is usually used [2]–[6]. Recently, wavelet-based filtering methods have been proposed [7], [8]. The main advantage of filtering methods is that they can be used on small images, possibly georectified. However, the spatial frequency components that are filtered out are only partly due to the striping effect. They also contain useful information on the observed image. Such information is lost after the filtering operation, and ringing phenomena may appear in the corrected image. Moreover, optimal filter design is not easy [5]; a Wiener filter approach has been proposed [9], but information on the spatial frequency content of the image is then required. In [9], this information is extracted from a reference image measured on another channel not affected by the striping effect.

Statistical methods account for the assumption that the inputs of all sensors have the same statistical characteristics. Using moment-matching techniques, assuming that the inputs of all sensors have the same mean and standard deviation, affine response functions of the detectors can be corrected [1], [10]–[14]. Histogram matching techniques require that the inputs have the same probability density functions, and they have a capability to deal with nonlinear detector responses [10], [15]. However, they may remain ineffective on small-size images, in particular when the image contains structures oriented along-track [16], [17]. Improved histogram-based methods have been proposed to be more robust to inadequacies of the statistical assumptions [16], [17]. Other methods first select homogeneous

parts of the image and then perform relative equalization of nonlinear detectors using polynomial models [18], [19].

In this paper, we only consider a linear detector response, which is a valid approximation for SPOT satellite instruments, and our goal is to calibrate the gain of the detectors from an observed image without reference or prior knowledge. The proposed method is presented in the Bayesian framework. From a hypothetical statistical model for the image, a gain estimator is deduced. Then, we demonstrate that the proposed gain estimator has statistical convergence properties under broad assumptions. Furthermore, the proposed estimator is shown to provide more accurate results than moment-matching-based estimators in a simulated finite data case, particularly for images with a low number of rows.

The pushbroom image acquisition principle is recalled in Section II, as well as the considered model of detectors, the current calibration procedure, and the main existing destriping techniques. In Section III, a statistical maximum *a posteriori* (MAP) approach based on a Markov random field description of ground scenes is presented, and a new class of gain estimators is deduced. Section IV is devoted to computational issues, which leads us to propose a low-cost iterative algorithm based on solving tridiagonal linear systems. In Section V, it is shown that the resulting gain estimators possess convergence properties in a much wider context than the range of validity of the statistical model initially adopted. Finally, Section VI deals with the practical use of the proposed gain estimator and presents performance comparisons based on realistic simulated data, for both images with low and high number of rows. Practical sensitivity to quantization and image dynamic is also studied. Finally, results on a real SPOT3 image are shown.

II. Problem Statement

A. Image Formation and Detectors Model

Pushbroom-type imaging instruments are designed to acquire complete rows of images using a linear array of detectors. The other dimension is obtained by the column-wise scanning that results from the motion of the satellite along its orbit (see Fig. 1). In this paper, we focus on SPOT imaging instruments, but the proposed method can be used for every pushbroom instrument as long as the sensor's response can be considered linear. The first four generations of SPOT satellites make use of CCD arrays of 3000 pixels in multispectral mode and 6000[1] in panchromatic mode. Those figures are doubled in the case of SPOT5. In practice, the response of the detectors is not strictly uniform along the array. This is the cause of a striping effect in the direction of columns that must be compensated for.

As a first approximation, which is valid for most of the detectors of SPOT satellite instruments, the detector response can be considered as an affine function depending on two parameters, i.e., the gain g_c and the offset o_c. Let $\Gamma = \{(r,c), r = 1,\ldots,R, c = 1,\ldots,C\}$ be a rectangular grid of R rows and C columns used to index the data. The relation between the perfect data (data from perfect detectors) $\boldsymbol{x} = \{x_{r,c}\}_\Gamma$ and the actual data $\boldsymbol{y} = \{y_{r,c}\}_\Gamma$ can thus be written as

$$y_{r,c} = g_c\, x_{r,c} + o_c.$$

[1]Four linear arrays of 1500 detectors each.

As shown in Section II-B, the current calibration procedure for the offsets is very simple and effective, so we focus hereafter on the calibration of the gains and consider a purely linear model

$$y_{r,c} = g_c\, x_{r,c}. \tag{1}$$

With perfect detectors having an identity response ($g_c = 1$), the imperfections then correspond to unknown gains $g_c > 0$, one for each column of pixels.

An equivalent relation can be written using the logarithm of (1): $\log y_{r,c} = \log g_c + \log x_{r,c}$. For simplicity, we will note hereafter $x^\ell_{r,c} = \log x_{r,c}$, $\boldsymbol{x}^\ell = \{x^\ell_{r,c}\}_\Gamma$, $y^\ell_{r,c} = \log y_{r,c}$, $\boldsymbol{y}^\ell = \{y^\ell_{r,c}\}_\Gamma$, $g^\ell_c = \log g_c$, and $\boldsymbol{g}^\ell = \{g^\ell_c\}_{c=1,\ldots,C}$, which leads to the additive relation

$$y^\ell_{r,c} = x^\ell_{r,c} + g^\ell_c \iff \boldsymbol{y}^\ell = \boldsymbol{x}^\ell + \mathbf{1}_R\,(\boldsymbol{g}^\ell)^t \tag{2}$$

where the symbol ^{t} denotes the transpose operator, and $\mathbf{1}_R = [1,\ldots,1]^t$ is a column vector of length R.

B. Calibration of SPOT Instrument

At the present time, identification of both gains and offsets is performed during calibration phases. During such phases, the satellite imaging instruments observe a landscape that is as uniform as possible. To calibrate the offset, a mirror allows observing the sky, which can be considered as a null image. To calibrate the gain, the procedure is more complex since there is no perfectly uniform landscape on Earth. Observations over the poles are generally considered. In such experimental conditions, even a simple moment-matching estimator can give satisfactory estimated gains and offsets. Such a calibration system is periodically used to check and, if necessary, to adjust the detector responses.

Our aim is to estimate the gains $\boldsymbol{g} = \{g_c\}_{c=1,\ldots,C}$ from any image (typically, of $R = 6000$ rows and $C = 6000$ columns), i.e., to avoid supervised calibration procedures using "well-chosen" images. Since the perfect data $x_{r,c}$ are unknown, such a problem is highly indeterminate. On the other hand, we can rely on statistical redundancy, since only C unknown gains have to be estimated from $R \times C$ measured pixels, where the number of rows R may be very large.

C. Simulated Images

To compare the different estimators, simulated data were generated from a perfect image $\{x_{r,c}\}$ and a set of gain values. In what follows, the notation $\widetilde{g_c}$ refers to true gain values.

We simulated the SPOT acquisition process (by convolution and subsampling) from aerial photographs with a resolution of 1 m. We then obtained images with 2360 rows and 376 columns. To come closer to the SPOT configuration in terms of number of detectors, we mirrored the image twice in the column direction, so we obtained an image with $4 \times (376 - 1) = 1500$ columns. In a second stage, to practically check the convergence of the estimators with respect to the number of image rows, we also duplicated the image in the other direction to produce arbitrarily long images. For each duplication, a random circular shift was employed to avoid repeating the same landscape within each long column and to ensure that all image

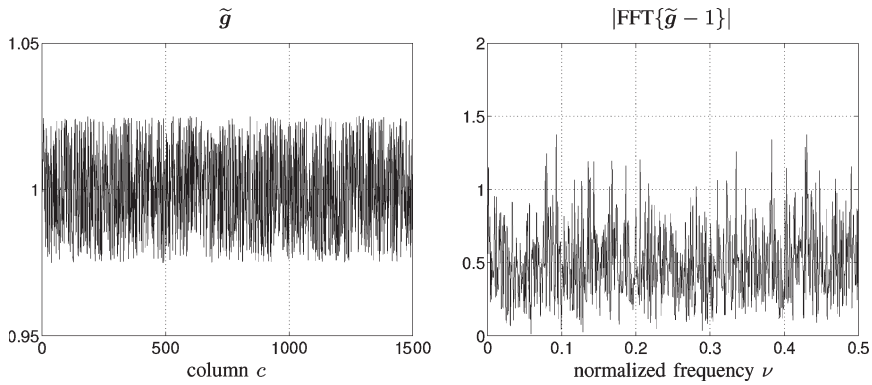

Fig. 2. Spatial and spectral representation of the simulated gains $\widetilde{g}$.

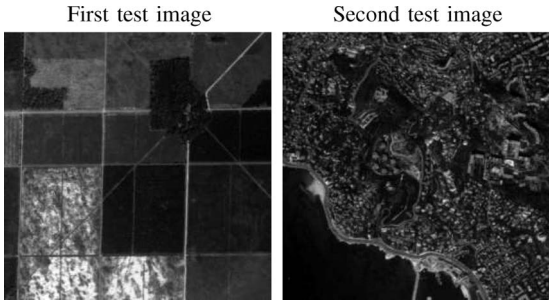

Fig. 3. Subimages of 300 × 300 pixels extracted from the two test images (Copyright CNES/Distribution SPOT IMAGE).

columns contain statistically similar data for an infinitely long image. Of course, such circular shifts have very limited effects for small images.

The applied gains $\widetilde{g_c}$ (see Fig. 2) were generated as independent identically distributed (i.i.d.) random variables distributed uniformly on [0.975, 1.025], which corresponds to the order of magnitude of the gains estimated during the calibration phases of SPOT satellites. Of course, no mirroring has been accounted for in the sampling of the gains.

Two test images have been used, corresponding to two different types of observed landscape, as illustrated in Fig. 3. In particular, let us stress that the first type presents oriented along-track structures, which makes the destriping problem more difficult when the image is of small size [16], [17].

Two indices have been used to quantify the quality of the estimated gains $\widehat{g}_c$, i.e.,

$$\sigma_{\mathrm{E}} = \sqrt{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\widehat{g}_c}{\widetilde{g}_c}-1\right)^2} \qquad \max_{\mathrm{V}} = \max_{c}\left|\frac{\widehat{g}_c}{\widetilde{g}_c}-\frac{\widehat{g}_{c+1}}{\widetilde{g}_{c+1}}\right|$$

namely, *normalized mean square error* σ_{E} and *visual maximum* $\max_{\mathrm{V}}$. The interest of the latter is to provide the maximum calibration error between two neighboring columns, which is the most visible effect in the image (related to high frequencies errors on the estimated gains), whereas the former provides a more global and quantitative calibration error. σ_{E} somehow measures the radiometry distortion in the corrected images as, according to the model, the radiometry of the original image $x_{r,c}$ has been modified by the gains of the detectors according to $y_{r,c} = g_c x_{r,c}$. If the gains are correctly estimated (low σ_{E}), then the corrected image $\widehat{x}_{r,c} = y_{r,c}/\widehat{g}_c$ will have the same radiometry than the original image.

D. Moment-Matching-Based Empirical Estimators

Among the existing destriping methods, we will only study the moment-matching-based empirical estimators, as histogram matching methods are generally used to correct striping from nonlinear detector responses—which is not the considered case—and our goal is to calibrate the gains, without losing spatial information via image filtering.

From an empirical viewpoint, estimators that converge toward the true gains for an infinite number of acquired rows can easily be defined in an ergodic stationary statistical framework. In such a framework, if we assume that each column of the image has the same statistical mean, then the empirical mean of a column $(1/R)\sum_{r=1}^{R} y_{r,c} = g_c(1/R)\sum_{r=1}^{R} x_{r,c}$ converges toward $g_c m_x$, where $m_x = \mathrm{E}[X_{r,c}]$ is the mean of the perfect data, whereas the empirical mean of the whole image $(1/RC)\sum_{r=1}^{R}\sum_{c=1}^{C} y_{r,c}$ converges toward $m_x \mathrm{E}[g_c] = m_x$ (if we assume $\mathrm{E}[g_c] = 1$). Then, one can define the empirical mean-based estimator as

$$\widehat{g}_c^{\mathrm{EM}} = \frac{C\sum_{r=1}^{R} y_{r,c}}{\sum_{c'=1}^{C}\sum_{r=1}^{R} y_{r,c'}}$$

which is expected to converge toward g_c when the number of rows tends toward infinity. Let us note that if $\mathrm{E}[g_c]$ is unknown, then the gains can only be estimated up to a multiplicative constant, so the radiometry of the image will not be preserved. In practice, since the gains to estimate often correspond to deviations from calibrated values, the assumption $\mathrm{E}[g_c] = 1$ can be considered as valid.

However, such an estimator gives unsatisfactory results, even for a large (but finite) number of rows. For a typical number of rows $R = 3000$, it can be seen on the example in Fig. 4 that the gains are very badly estimated. The reason is that such an estimator is very sensitive to the observed landscape. Indeed, it is clear from Fig. 4 that the mirroring effect performed on $\boldsymbol{x}$ for the simulation is retrieved on the estimated gains. On the spectral representation, it can be seen that the high-frequency components $(\nu > 0.2)$ are well estimated but the low frequencies have been overestimated. Of course, such an effect decreases when the number of rows increases, but even for $R = 10^5$ it is still visible.

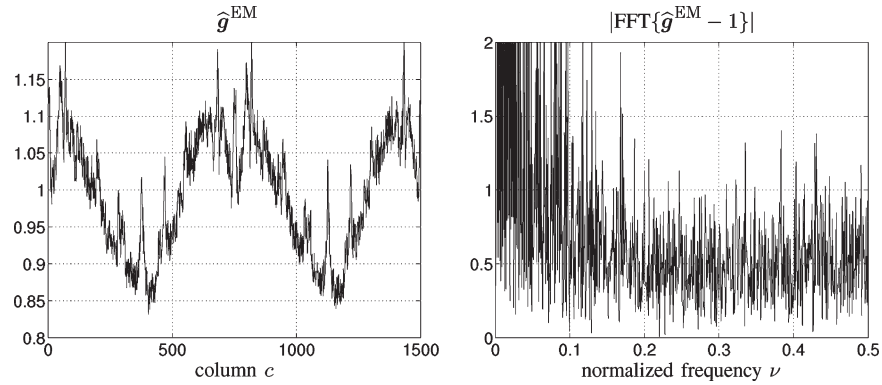

Fig. 4. Spatial and spectral representation of the gains estimated with $\widehat{\boldsymbol{g}}^{\mathrm{EM}}$ on first test image for $R = 3000 : \sigma_{\mathrm{E}} = 7.74\%$ and $\max_{\mathrm{V}} = 3.85\%$.

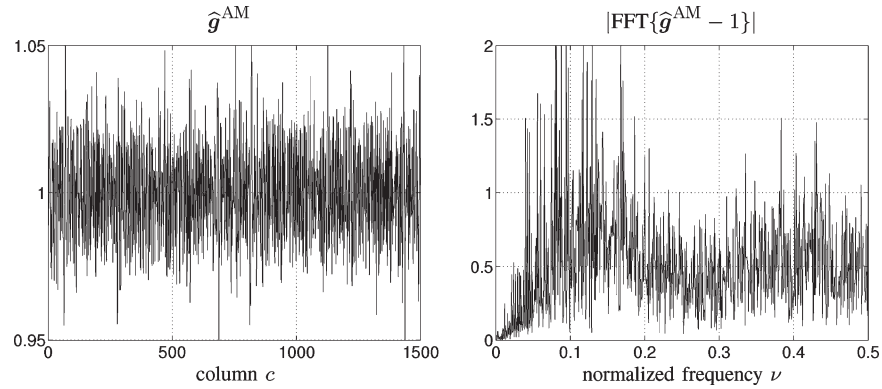

Fig. 5. Spatial and spectral representation of the gains estimated with $\widehat{\boldsymbol{g}}^{\mathrm{AM}}$ (rectangular window $w_k = 1/9$, $k = -4 \ldots, 4$) on first test image for $R = 3000$: $\sigma_{\mathrm{E}} = 0.96\%$ and $\max_{\mathrm{V}} = 2.35\%$.

Such moment-matching empirical estimators have successfully been used on Landsat TM images [12], considering an affine model for the detector response (gain and offset have to be estimated). Note that for the Landsat satellite, the whole images are acquired with only six detectors, so the assumption that each detector acquires data with the same statistical mean is reinforced by the periodicity of the striping: six gains and offsets have to be estimated from $R \times C$ data, each detector d measuring data $y_{r,c} = g_d x_{r,c} + o_d$, where $d = c \bmod 6$.

For pushbroom-type images, as the mean of the image cannot be supposed constant along each column for an image with a low number of rows, a more robust adaptive version of such an empirical estimator can be defined, in which the mean of the image is computed locally with respect to columns using a normalized window $\{w_k\}_{k=-L,\ldots,L}$. For columns $c = L+1, \ldots, C-L$, let the adaptive mean-based estimator be defined by

$$\widehat{g}_c^{\mathrm{AM}} = \frac{\sum_{r=1}^{R} y_{r,c}}{\sum_{c'=c-L}^{c+L} \left(w_{c'-c} \sum_{r=1}^{R} y_{r,c'} \right)}.$$

Such adaptive empirical estimators have successfully been used, e.g., on Landsat and NOAA images, which are associated with detrending techniques [11] or with median filtering [13], to estimate both gains and offsets from Hyperion hyperspectral data.

When the image mean is computed locally, relatively good results can be obtained, as shown in Fig. 5. However, the spectral representation of the gains shows that the low frequencies are not so well estimated. One could still try to improve such a result, for instance, using other window shapes than a rectangular one, or by filtering the estimated gains, but an optimal choice of the window or the filter is not easy in a general case. Finally, more complex procedures based on moment-matching estimators and filtering can be defined for destriping hyperspectral pushbroom data, accounting for correlation of the images in the spectral dimension, such in [14] for Hyperion data.

III. New Estimators

We will focus hereafter on the additive relation (2). The proposed estimators can be introduced within the Bayesian statistical framework. Bayes' rule allows us to combine prior information on the parameters to estimate with the information contained in the data to obtain the posterior probability density

$$p_{G^\ell|Y^\ell}(\boldsymbol{g}^\ell|\boldsymbol{y}^\ell) = \frac{p_{Y^\ell|G^\ell}(\boldsymbol{y}^\ell|\boldsymbol{g}^\ell) p_{G^\ell}(\boldsymbol{g}^\ell)}{p_{Y^\ell}(\boldsymbol{y}^\ell)} \tag{3}$$

where $p_{G^\ell}(\boldsymbol{g}^\ell)$ is the *a priori* density of $\boldsymbol{g}^\ell$, and $p_{Y^\ell|G^\ell}(\boldsymbol{y}^\ell|\boldsymbol{g}^\ell)$ is the data likelihood. The MAP estimator maximizes the posterior density, i.e., the *a posteriori* likelihood.

A. A Priori Model for the Gains

It seems reasonable to consider the gains to be statistically independent and of mean equal to 1 (their logarithm being

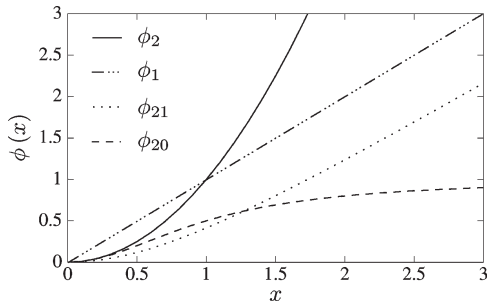

Fig. 6. Four functions defined by (7). The last two are depicted for $s = 1$.

zero mean). Therefore, a centered i.i.d. Gaussian distribution (of variance σ_g^2) can be used for p_{G^ℓ} in a first approximation

$$p_{G^\ell}(\boldsymbol{g}^\ell) \propto \exp\left\{-\frac{1}{2\sigma_g^2}\sum_c \left(g_c^\ell\right)^2\right\}. \tag{4}$$

As for the moment-matching-based estimators, if $\mathrm{E}[g_c]$ is unknown, then the gains can only be estimated up to a multiplicative constant, so the radiometry of the image will not be preserved.

B. A Priori Model for the Image

Given a stochastic model for the perfect data $\boldsymbol{x}^\ell$ described by a probability density p_{X^ℓ}, the data likelihood $p_{Y^\ell}(\boldsymbol{y}^\ell|\boldsymbol{g}^\ell)$ can easily be expressed from the additive relation (2) between $\boldsymbol{X}^\ell$ and $\boldsymbol{Y}^\ell$ as

$$p_{Y^\ell}(\boldsymbol{y}^\ell|\boldsymbol{g}^\ell) = p_{X^\ell}\left(\boldsymbol{x}^\ell + \mathbf{1}_R(\boldsymbol{g}^\ell)^t\right). \tag{5}$$

In the following, we have adopted a basic Markov field model, as widely used in the domain of image analysis and restoration [20], [21], for the logarithm of the image $\boldsymbol{X}^\ell$. More specifically, the chosen model is a first-order Markov field with pairwise interactions restricted to differences between neighboring pixel intensities as

$$p_{X^\ell}(\boldsymbol{x}^\ell) \propto \exp\left\{-\frac{1}{T}\sum_{m\sim n}\phi\left(x_m^\ell - x_n^\ell\right)\right\} \tag{6}$$

where $m \sim n$ means summation over all distinct pairs $\{m, n\}$ of horizontal and vertical neighboring pixels, and T is a constant parameter. Note that such a model is improper since $\int_{\boldsymbol{x}} p_{X^\ell}(\boldsymbol{x}^\ell)d\boldsymbol{x}^\ell = +\infty$, but this happens to be of no practical consequence in the present context.

Classically [20], [22], one chooses ϕ as an even function nondecreasing on $\mathbf{R}^+$. Typical examples are as follows:

$$\begin{aligned}
&\text{quadratic function } (L_2 \text{ norm}):\ \phi_2(u) = u^2\\
&\text{magnitude function } (L_1 \text{ norm}):\ \phi_1(u) = |u|\\
&\text{hyperbolic function [22]:}\ \phi_{21}(u) = \sqrt{s^2+u^2} - s\\
&\text{Geman-McClure function [23]:}\ \phi_{20}(u) = \tfrac{u^2}{s^2+u^2}.
\end{aligned} \tag{7}$$

Whereas the first two examples have a uniform behavior over $\mathbf{R}$ (respectively, quadratic and linear), the last two examples behave quadratically near zero while they are asymptotically linear (ϕ_{21}) or constant (ϕ_{20}), respectively. The change of behavior roughly happens around the threshold parameter s, as seen in Fig. 6.

Differences between neighboring pixel intensities are expected to be small except at the image's edges. Therefore, symmetric functions ϕ that grow more slowly than a parabola are often used: neighboring pixels within homogeneous regions or at the edges between two regions correspond to small or large values of the argument of ϕ, respectively. Such functions are often called *edge preserving*. The function ϕ_1 is not strictly convex and not differentiable at the origin, whereas strictly convex and differentiable functions such as ϕ_{21} make a tradeoff between ϕ_1 and ϕ_2. Nonconvex functions such as ϕ_{20} are even more tolerant to large differences between neighboring pixel intensities at the image edges, but their drawback is to induce a possibly multimodal posterior likelihood, which is then harder to maximize globally.

Let us note that in [24] such kind of model is used for destriping pushbroom-type data, with the detector responses being estimated with a moment-matching-based estimator [24].

C. MAP Estimation

According to (4)–(6) and to Bayes' rule [see (3)], the proposed MAP estimator minimizes the anti-log posterior likelihood, which is proportional to the following criterion:

$$J_{\mathrm{MAP}}(\boldsymbol{g}^\ell) = \sum_{r,c}\phi\left(\delta g_c^\ell - \delta y_{r,c}^\ell\right) + \lambda\sum_c\left(g_c^\ell\right)^2 \tag{8}$$

where $\lambda = T/2\sigma_g^2$, $\delta g_c^\ell = g_c^\ell - g_{c+1}^\ell$, and $\delta y_{r,c}^\ell = y_{r,c}^\ell - y_{r,c+1}^\ell$. Let us note that only the differences between the horizontal neighbor intensities appear in the criterion J_{MAP}.

The aim of the second part of J_{MAP} is to penalize the large values of the gains g_c^ℓ. Note that the first part of the criterion is constant for $\boldsymbol{g}^\ell + \alpha\mathbf{1}_R$, whatever the value of α, so the penalization helps remove this ambiguity according to the assumption that g_c^ℓ is zero mean.

Note that for the extreme case where $\lambda = 0$, the proposed estimator corresponds to a maximum-likelihood (ML) estimator $\widehat{\boldsymbol{g}}_{\mathrm{ML}}^\ell$ for the chosen model. However, it corresponds either to $\sigma_g^2 = \infty$, i.e., an improper uniform prior law on $\mathbf{R}$ for $\boldsymbol{g}_c^\ell$, or to $T \to 0$, which corresponds to a uniform landscape, both assumptions being quite unrealistic in practice. The reader may refer to [25] for a more complete study of such an ML-based gain estimator.

Of course, the use of the logarithm of the data instead of the data themselves introduces differences instead of divisions. As a consequence, the resulting criterion J_{MAP} has a simpler structure that facilitates both the practical computations and the statistical study.

IV. Computation of the Estimate

The proposed estimator is defined as the minimizer of the MAP criterion [see (8)]. The goal of this section is to propose an efficient minimization technique.

A. MAP Estimate in the Quadratic Case

In the case $\phi = \phi_2$, the MAP criterion [see (8)] is quadratic and also reads

$$J_{\mathrm{MAP}}(\boldsymbol{g}^\ell) = \sum_r \left\|\mathbf{D}\boldsymbol{g}^\ell - \mathbf{D}\boldsymbol{y}_r^\ell\right\|^2 + \lambda\|\boldsymbol{g}^\ell\|^2$$

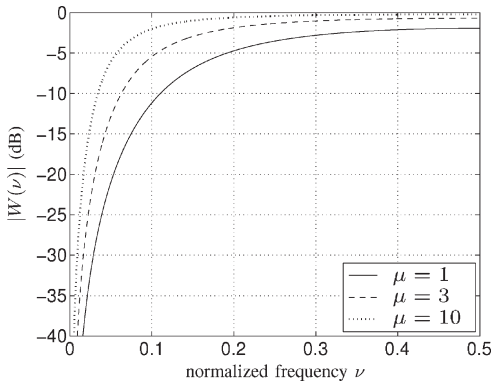

Fig. 7. Frequency response $W(\nu)$ defined by (11) for several values of $\mu = \lambda/R$.

where $\mathbf{D}$ is the Toeplitz matrix of first-order differences, and $\boldsymbol{y}_r^\ell = [y_{r,1}^\ell, y_{r,2}^\ell, \dots y_{r,C}^\ell]^t$. The solution thus has an explicit expression

$$\widehat{\boldsymbol{g}}_{\mathrm{MAP}}^\ell = (\mathbf{D}^t\mathbf{D} + \mu\mathbf{I})^{-1}\mathbf{D}^t\mathbf{D}\widehat{\boldsymbol{g}}_{\mathrm{ML}}^\ell \tag{9}$$

where $\mu = \lambda/R$ and

$$\widehat{\boldsymbol{g}}_{\mathrm{ML}}^\ell = \frac{1}{R}\sum_r \boldsymbol{y}_r^\ell \tag{10}$$

and $\mathbf{I}$ is the identity matrix. The matrix $\mathbf{D}^t\mathbf{D} + \mu\mathbf{I}$ being tridiagonal, the computation cost of $\widehat{\boldsymbol{g}}_{\mathrm{MAP}}^\ell$ is only linear with C.

Let us note that according to (10), $\widehat{\boldsymbol{g}}_{\mathrm{ML}}^\ell$ can be viewed as an empirical estimator for $\boldsymbol{g}^\ell$ that converges toward $\widetilde{\boldsymbol{g}}^\ell + m_x^\ell\mathbf{1}$ when $R \to \infty$, with $m_x^\ell = \mathrm{E}[X_{r,c}^\ell]$. Moreover, the matrix $\mathbf{D}^t\mathbf{D}$ is nearly circulant (only its outer rows and columns do not respect the circulant structure). Therefore, expression (9) can approximately be written in the Fourier domain as

$$\mathrm{FT}\left\{\widehat{\boldsymbol{g}}_{\mathrm{MAP}}^\ell\right\}(\nu) = W(\nu)\mathrm{FT}\left\{\widehat{\boldsymbol{g}}_{\mathrm{ML}}^\ell\right\}(\nu)$$

with

$$W(\nu) = \frac{|\mathrm{FT}\{\boldsymbol{d}\}(\nu)|^2}{|\mathrm{FT}\{\boldsymbol{d}\}(\nu)|^2 + \mu} \tag{11}$$

where $\boldsymbol{d} = [1, -1, 0, \dots, 0]^t$, so the MAP estimate $\widehat{\boldsymbol{g}}_{\mathrm{MAP}}^\ell$ corresponds to a filtered version of $\widehat{\boldsymbol{g}}_{\mathrm{ML}}^\ell$ with appropriate boundary conditions. The frequency response $W(\nu)$ is shown in Fig. 7 for several values of μ. It corresponds to a high-pass filter with a cutoff frequency that decreases for an increasing number of rows R in the image (and also for decreasing values of λ). Its effect is to remove the low-frequency components that are contaminated with the low-frequency components of the observed landscape. This is quite comparable to the result obtained by windowing the mean-based estimator according to the adaptive formulation in Section II-D.

B. MAP Estimate in the General Case

For functions ϕ different from ϕ_2, the MAP estimate has no explicit expression, and its computation amounts to solve an optimization problem that involves C gain variables using an iterative scheme.

```
% Matrices I and D (sparse structure)
D = spdiags([ones(C-1,1),-ones(C-1,1)],0:1,C-1,C);
I = speye(C,C);

Dyl = D*log(y)';
gl = zeros(1,C); % Initialization
for it=1:n_it    % Iterations
    b = zeros(C-1);
    w = zeros(C-1,1);
    % Compute b and W from eq. (12)
    for r=1:R    % Sweeping rows
        w_r = phiprimeover2x(D*gl-Dyl(:,r));
        b = b + w_r.*Dyl(:,r);
        w = w + w_r;
    end
    W = spdiags(w,0,C-1,C-1); % Sparse matrix W
    % Solve tridiagonal system of eq. (13)
    gl = (D'*W*D + lambda*I)\(D'*b);
end
g = exp(gl);
```

Fig. 8. Matlab version of the IRLS algorithm to minimize (8). The function `phiprimeover2x` is assumed to compute the vector $\boldsymbol{w}_r = [w_{r,1} \dots, w_{r,C}]$ according to (12).

One can still rely on solving tridiagonal linear systems, one at each iteration. In particular, let us adopt an iterative reweighted least square (IRLS) algorithm [26] (see also [27], [28] and Appendix I-A for a brief introduction). At each iteration, it consists of iteratively solving the following linear system:

$$\left(\mathbf{D}^t\left(\sum_r \mathbf{W}_r\right)\mathbf{D} + \lambda\mathbf{I}\right)\boldsymbol{g}^\ell = \mathbf{D}^t\left(\sum_r \mathbf{W}_r\mathbf{D}\boldsymbol{y}_r^\ell\right)$$

where $\mathbf{W}_r = \mathrm{diag}\{w_{r,1}, \dots, w_{r,C}\}$, and

$$w_{r,c} = \frac{\phi'\left(\delta g_c^\ell - \delta y_{r,c}^\ell\right)}{2\left(\delta g_c^\ell - \delta y_{r,c}^\ell\right)}. \tag{12}$$

In practice, this amounts to compute the diagonal matrices $\{\mathbf{W}_r\}_{r=1,\dots,R}$ and $\mathbf{W} = \sum_r \mathbf{W}_r$, the vector $\boldsymbol{b} = \sum_r \mathbf{W}_r\mathbf{D}\boldsymbol{y}_r^\ell$, and then to solve the following tridiagonal linear system:

$$\left(\mathbf{D}^t\mathbf{W}\mathbf{D} + \lambda\mathbf{I}\right)\boldsymbol{g}^\ell = \mathbf{D}^t\boldsymbol{b}. \tag{13}$$

Compared with other descent algorithms, the IRLS algorithm is easy to implement since only a tridiagonal linear system is solved at each iteration, and no step-size tuning is required. Fig. 8 depicts a Matlab code accounting for the sparse structure of the involved matrices.

Convergence toward $\widehat{\boldsymbol{g}}_{\mathrm{MAP}}^\ell$ is guaranteed for the function ϕ_{21} and many other differentiable convex functions ϕ [29], and also in the case of ϕ_1 provided that no component of $\mathbf{D}\boldsymbol{y}_r^\ell - \mathbf{D}\boldsymbol{g}^\ell$ cancels, so that the $w_{r,c}$'s defined in (12) are always properly defined [30]. In the case of a nonconvex function such as ϕ_{20}, convergence is still granted, but only in a local sense (meaning that the algorithm can be trapped in a local minimizer that is not the global solution) [28], [31]. In our context, convergence is reached in fewer than 100 iterations for $\phi = \phi_1$ and for $\phi = \phi_{21}$ for small values of s, and in fewer than 20 iterations otherwise.

V. Convergence Properties

To prove the consistency of the proposed estimator, one possibility is to rely on the asymptotic properties of the ML and Bayes estimates [32]. However, the practical impact of the result would depend on the fact that the Markov model

[see (6)] is statistically valid to describe the logarithm of the image. In the sequel, the proposed MAP estimator is shown to be consistent under much broader hypotheses.

A. Convergence Results

Definition 1: A real-valued function f is said to be n-*type unimodal* if for some value m (i.e., the mode), $f(x)$ is increasing for $x \leq m$ and decreasing for $x \geq m$. It is said to be u-*type unimodal* if $-f$ is n-type unimodal.

Proposition 1: Let us assume the following:

a) ϕ is a continuous even u-type unimodal function. Without loss of generality, we also assume that $\phi(0) = 0$.
b) $(\delta X^{\ell}_{r,c})_{r\in\mathbb{Z}}$ is a stationary ergodic random process.[2]
c) $\delta X^{\ell}_{1,c}$ admits a probability density p, which is even and n-type unimodal.
d) $\phi * p(0) < \infty$.

Let $\widehat{\boldsymbol{g}}^{\ell}$ denote the minimizer of (8), and $\delta\widehat{\boldsymbol{g}}^{\ell} = (\delta\widehat{g}^{\ell}_c)$, with $\delta\widehat{g}^{\ell}_c = \widehat{g}^{\ell}_c - \widehat{g}^{\ell}_{c+1}$. Then, $\delta\widehat{\boldsymbol{g}}^{\ell}$ converges in probability toward the true value $\delta\widetilde{\boldsymbol{g}}^{\ell}$ when the number of rows tends to infinity.

Proof: The proof is based on the fact that $\delta\widehat{\boldsymbol{g}}^{\ell}$ is a *minimum contrast estimator* [33, Sec. 3.2]. A minimum contrast estimator of parameter $\boldsymbol{\theta}$ minimizes a cost function $\mathcal{L}_N(\boldsymbol{\theta})$ such that we have the following:

A1. when $N \to \infty$, $\mathcal{L}_N(\boldsymbol{\theta})$ tends toward a function $\mathcal{L}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ of the true parameters $\widetilde{\boldsymbol{\theta}}$;
A2. $\mathcal{L}(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})$ is a contrast function for $\widetilde{\boldsymbol{\theta}}$, i.e., it reaches its minimum, as a function of $\boldsymbol{\theta}$, only at $\boldsymbol{\theta} = \widetilde{\boldsymbol{\theta}}$.

Under quite general regularity conditions, the minimum contrast estimators are weakly consistent, i.e., the minimizer $\widehat{\boldsymbol{\theta}}_N$ of $\mathcal{L}_N(\boldsymbol{\theta})$ converges in probability toward $\widetilde{\boldsymbol{\theta}}$ as

$$\forall\, \epsilon > 0, \quad \lim_{N\to\infty} \Pr\left\{ \|\widehat{\boldsymbol{\theta}}_{\boldsymbol{N}} - \widetilde{\boldsymbol{\theta}}\|^2 > \epsilon \right\} = 0.$$

Let us prove that the minimizer of J_{MAP} yields a minimum contrast estimator associated with the following limiting function:

$$\mathcal{L}_c(\delta\boldsymbol{g}^{\ell}, \delta\widetilde{\boldsymbol{g}}^{\ell}) = \sum_c \mathcal{L}_c\left(\delta g^{\ell}_c, \delta\widetilde{g}^{\ell}_c\right)$$

with

$$\begin{aligned}\mathcal{L}_c\left(\delta g^{\ell}_c, \delta\widetilde{g}^{\ell}_c\right) &= \left[\phi\left(\delta g^{\ell}_c - \delta\widetilde{g}^{\ell}_c - \delta X^{\ell}_{1,c}\right)\right] \\ &= \phi * p\left(\delta g^{\ell}_c - \delta\widetilde{g}^{\ell}_c\right). \end{aligned} \tag{14}$$

First, it is equivalent to minimize $J^R_{\mathrm{MAP}} = J_{\mathrm{MAP}}/R$ rather than J_{MAP}. Given (8), J^R_{MAP} takes the following form:

$$J^R_{\mathrm{MAP}}(\boldsymbol{g}^{\ell}) = \sum_c \mathcal{L}^R_c\left(\delta g^{\ell}_c\right) + \frac{\lambda}{R}\sum_c \left(g^{\ell}_c\right)^2$$

where

$$\mathcal{L}^R_c\left(\delta g^{\ell}_c\right) = \frac{1}{R}\sum_r \phi\left(g^{\ell}_c - y^{\ell}_{r,c}\right).$$

[2]A stationary random process $(X_r)_{r\in\mathbb{Z}}$ is said to be ergodic if $\mathrm{plim}_{R\to\infty}\sum_{r=1}^{R}\phi(X_r) = \mathrm{E}[\phi(X_1)]$ for every function ϕ for which $\mathrm{E}[|\phi(X_1)|]$ is finite.

Let us consider the linear change of the variable $(\boldsymbol{g}^{\ell}) \to (g^{\ell}_1, \delta\boldsymbol{g}^{\ell})$, which is invertible given that

$$g^{\ell}_c = g^{\ell}_1 - \sum_{k=1}^{c-1}\delta g^{\ell}_k$$

for all $c > 1$. Thus, we have

$$\begin{aligned}&\min_{\boldsymbol{g}^{\ell}} J^R_{\mathrm{MAP}}(\boldsymbol{g}^{\ell}) \\ &\quad= \min_{g^{\ell}_1, \delta\boldsymbol{g}^{\ell}}\left(\sum_c \mathcal{L}^R_c\left(\delta g^{\ell}_c\right) + \frac{\lambda}{R}\sum_c\left(g^{\ell}_1 - \sum_{k=1}^{c-1}\delta g^{\ell}_k\right)^2\right) \\ &\quad= \min_{\delta\boldsymbol{g}^{\ell}}\left(\sum_c \mathcal{L}^R_c\left(\delta g^{\ell}_c\right) + \frac{\lambda}{R}\min_{g^{\ell}_1}\sum_c\left(g^{\ell}_1 - \sum_{k=1}^{c-1}\delta g^{\ell}_k\right)^2\right) \\ &\quad= \min_{\delta\boldsymbol{g}^{\ell}}\left(\sum_c \mathcal{L}^R_c\left(\delta g^{\ell}_c\right) + \frac{\lambda}{R}Q(\delta\boldsymbol{g}^{\ell})\right)\end{aligned}$$

where Q is a quadratic function of $\delta\boldsymbol{g}^{\ell}$ whose exact expression does not matter here, save for the fact that it does not depend on R. Thus, in the limit $R \to \infty$, the second term vanishes, whereas the first term tends toward $\mathcal{L}(\delta\boldsymbol{g}^{\ell}, \delta\widetilde{\boldsymbol{g}}^{\ell})$.

It remains to justify that $\mathcal{L}$ is a contrast function for $\delta\widetilde{\boldsymbol{g}}^{\ell}$, or equivalently that $\mathcal{L}_c$ is a contrast function for $\delta\widetilde{g}^{\ell}_c$ for all c. Given (14), this amounts to showing that the minimum of $\phi * p$ is uniquely attained at zero. The latter property holds since $\phi * p$ is even and u-type unimodal as a well-defined convolution of even unimodal functions. This result stems from a variation on Wintner's result about even unimodal probability distributions [34]. Since Wintner's result is only applicable if ϕ can be normalized under the form of a probability density, we formulate and prove a standalone result in Appendix I-B. ■

Let us note that the consistency of $\delta\widetilde{\boldsymbol{g}}^{\ell}$ does not necessarily imply that of $\widetilde{\boldsymbol{g}}^{\ell}$. However, such an implication can be derived if an additional condition holds for the true gains, for instance, that their logarithm is statistically centered.

B. Practical Considerations

All functions ϕ classically used in the Markov field image modeling and robust estimation satisfy condition *(a)* of Proposition 1. Moreover, the stationarity and ergodicity required by condition *(b)* are quite general assumptions. Therefore, one has to give particular attention to the conditions of Proposition 1 on $\delta X^{\ell}_{r,c}$, and particularly to conditions *(c)* and *(d)*.

As $\delta x^{\ell}_{r,c} = x^{\ell}_{r,c} - x^{\ell}_{r,c+1}$ is a difference between pixels, it is quite natural to suppose that $\delta X^{\ell}_{r,c}$ has an even density. The unimodality assumption on p is stronger. Such a property is impossible to be assessed in general, but it can be checked on estimated densities from a case by case basis. Fig. 9 shows the estimated densities using simple histograms on the two test images. On such examples, it seems reasonable to assume that p is even and unimodal.

Finally, condition *(d)* also reads $\int \phi(u)p(u)du < \infty$. Since p and ϕ are unimodal continuous functions, an equivalent property is that $\phi(u)p(u)$ vanishes more rapidly than $|u|^{-1}$ for large values of $|u|$. Since the considered functions ϕ do not increase more rapidly than $|u|^2$, or even than $|u|$ in the case of edge-preserving functions, $p(u)$ must vanish more rapidly than

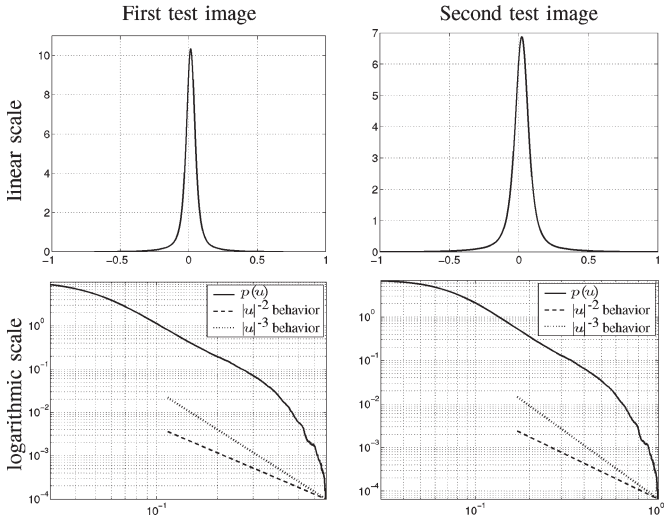

Fig. 9. Histogram-based estimated density $p(u)$ for the two test images, i.e., in linear and logarithmic scales. In the latter case, only the right tail of $p(u)$ is represented, which makes visible that $p(u)$ vanishes much more rapidly than $|u|^{-2}$ and $|u|^{-3}$ in both cases.

$|u|^{\gamma}$ with $\gamma = -3$ or $\gamma = -2$ in the edge-preserving case. Once again, such a property cannot be established in a general way. However, Fig. 9 indicates that it holds for the two test images.

As a conclusion, Proposition 1 is based on realistic conditions, i.e., that can be justified by either mathematical or practical considerations.

VI. Simulation Results

A. Proposed Estimator

Table I compares the results obtained with the empirical mean and adaptive-mean-based estimators $\widehat{\boldsymbol{g}}^{\mathrm{EM}}$, $\widehat{\boldsymbol{g}}^{\mathrm{AM}}$, with the proposed estimators for different functions ϕ, for 3000 or 10 000 rows of the test images.

All of the proposed estimators depend on the parameter λ, and $\widehat{\boldsymbol{g}}^{2,1}_{\mathrm{MAP}}$ and $\widehat{\boldsymbol{g}}^{2,0}_{\mathrm{MAP}}$ also depend on the parameter s.

The parameter $\lambda = T/2\sigma_g^2$ depends on both the variance of the true gain values σ_g^2 and the image (via T). In the quadratic case, λ has been shown to tune the high-pass filtering of the empirically estimated logarithm of gains (see Section IV-A). While this interpretation is no more valid in the nonquadratic case, λ still plays the same role, the level of "filtering" being set by the parameter λ. The example in Fig. 10 clearly indicates that for a small value of λ the low frequencies are overestimated, whereas for a large value they are underestimated. Moreover, Table I shows that gain estimation is not very sensitive to the value of λ, so a constant value can be adopted independently of the test image and of the value of s. In our case, an appropriate choice is $\lambda = 10^3$ for ϕ_2, ϕ_1, and ϕ_{21}, and $\lambda = 10^4$ for ϕ_{20}. However, one must pay attention to the fact that such a value also depends on the variance of the gains, i.e., of the severity of the striping effect. In all cases, the display of the estimated gains in the spectral domain should be useful to adjust the value of λ in an empirical way.

In the case of $\widehat{\boldsymbol{g}}^{2,1}_{\mathrm{MAP}}$ and $\widehat{\boldsymbol{g}}^{2,0}_{\mathrm{MAP}}$, the value of the parameter s must also be chosen, which happens to be rather easy. At first glance, s is image dependent since it enters the *a priori* law of the image. However, this parameter does not depend on the image scale, since our model (6) deals with differences between the logarithm of pixel values. For the function ϕ_{21}, the best results are obtained for a low value for s, so that ϕ_{21} may be seen as a differentiable approximation of ϕ_1. However, the computational burden of the algorithm proposed in Section IV-B increases for very small value of s, whereas the estimated gains are very similar. Therefore, a good tradeoff is reached at 10^{-2}. The situation is quite comparable for function ϕ_{20}, and we have found that $s = 0.1$ was an appropriate value for all the tested images.

Finally, Table I shows that the proposed approach gives better results than the moment matching one in all situations both in terms of σ_{E} and $\mathrm{max}_{\mathrm{V}}$. In particular, the function ϕ_{20} generally provides better results than ϕ_1 and ϕ_{21}, and the latter perform better than ϕ_2.

The typical numbers of iterations and computation time for the simulations in Table I for $R = 3000$ (respectively, $R = 10\,000$) rows image are as follows:

- 50 iterations and 20 s (respectively, 70 s) for ϕ_1;
- 12 iterations and 8 s (respectively, 30 s) for ϕ_{21};
- 12 iterations and 5 s (respectively, 20 s) for ϕ_{20}.

B. Practical Convergence

We empirically checked the convergence of the proposed estimators toward theoretical values for images with a growing number of rows (from $R = 10^3$ to 10^5). Fig. 11 depicts the quality indices of the different estimates as functions of R. Note that σ_{E} is not represented for the empirical mean estimator $\widehat{\boldsymbol{g}}^{\mathrm{EM}}$ because the values are too large. The proposed approach gives much better results than the adaptive mean moment matching solution $\widehat{g}^{\mathrm{AM}}$, particularly for small numbers of rows (typically, $R = 6000$ for SPOT images). The results obtained using ϕ_2 are of intermediate quality between that of the adaptive mean moment-matching estimator $\widehat{\boldsymbol{g}}^{\mathrm{AM}}$ and MAP estimation using edge-preserving functions ϕ.

As expected, ϕ_1 and ϕ_{21} for small values of s always give very similar results. The function ϕ_{20} gives even slightly better results in terms of mean square error σ_{E} and comparable results in terms of visual maximum $\mathrm{max}_{\mathrm{V}}$. Let us also remark that the performances are more contrasted for the first test image than for the second test image. However, the ordering remains the same in all cases and whatever the number of rows.

C. Sensitivity to Quantization and to Image Dynamics

In practice, quantization is performed on the acquired images, so the aim of this section is to quantify the sensitivity of the proposed estimators to quantization. Fig. 12 depicts the quality indices of the different estimators as functions of the number of bits used for quantization. Equivalently, such quality indices can be considered as functions of the image dynamics, as using one bit less for quantization is strictly equivalent to halving the dynamics. The proposed approach is clearly more sensitive to the quantization effect than the adaptive mean moment-matching solution. For very low number of bits, the latter would become preferable. However, the proposed estimators remain superior as long as the image intensities are quantified using more than 6 bits, which covers all practical situations.

TABLE I
ERROR INDICES FOR THE GAINS ESTIMATED WITH THE MOMENT-MATCHING AND MAP ESTIMATORS FOR DIFFERENT FUNCTIONS ϕ AND VARIOUS VALUES FOR PARAMETER s AND λ; ESTIMATION PERFORMED ON THE FIRST AND SECOND TEST IMAGES FOR $R = 3000$ AND $R = 10\,000$

			Image 1				Image 2			
			$R = 3000$		$R = 10000$		$R = 3000$		$R = 10000$	
	s	λ	σ_{E}	$\max_{\mathrm{V}}$	σ_{E}	$\max_{\mathrm{V}}$	σ_{E}	$\max_{\mathrm{V}}$	σ_{E}	$\max_{\mathrm{V}}$
$\widehat{\boldsymbol{g}}^{\mathrm{EM}}$			7.74%	3.85%	5.00%	1.52%	6.89%	1.98%	4.50%	1.34%
$\widehat{\boldsymbol{g}}^{\mathrm{AM}}$			0.96%	2.35%	0.66%	1.61%	0.87%	2.52%	0.63%	1.52%
$\widehat{\boldsymbol{g}}^{2}_{\mathrm{MAP}}$		10^2	1.22%	2.59%	0.95%	1.08%	0.89%	1.89%	0.77%	1.22%
		$\mathbf{10^3}$	**0.69%**	**1.63%**	**0.61%**	**0.92%**	**0.71%**	**1.64%**	**0.56%**	**1.15%**
		10^4	1.01%	2.87%	0.76%	1.64%	1.02%	2.98%	0.76%	1.81%
		10^5	1.37%	4.50%	1.24%	3.86%	1.37%	4.50%	1.24%	3.86%
$\widehat{\boldsymbol{g}}^{1}_{\mathrm{MAP}}$		10^2	0.88%	0.84%	0.55%	0.44%	0.95%	1.13%	0.78%	0.62%
		$\mathbf{10^3}$	**0.54%**	**0.78%**	**0.44%**	**0.45%**	**0.57%**	**0.97%**	**0.47%**	**0.62%**
		10^4	0.56%	1.08%	0.44%	0.55%	0.72%	1.70%	0.54%	0.91%
		10^5	1.01%	3.05%	0.74%	1.74%	1.20%	3.89%	0.95%	2.62%
$\widehat{\boldsymbol{g}}^{2,1}_{\mathrm{MAP}}$	0.01	10^2	0.88%	0.94%	0.56%	0.46%	0.95%	1.13%	0.79%	0.63%
		$\mathbf{10^3}$	**0.55%**	**0.82%**	**0.45%**	**0.43%**	**0.58%**	**0.96%**	**0.47%**	**0.60%**
		10^4	0.57%	1.06%	0.45%	0.56%	0.73%	1.70%	0.55%	0.99%
		10^5	1.03%	3.08%	0.76%	1.77%	1.21%	3.89%	0.96%	2.69%
	0.1	10^2	1.08%	1.74%	0.74%	0.69%	0.92%	1.32%	0.80%	0.84%
		$\mathbf{10^3}$	**0.63%**	**1.19%**	**0.54%**	**0.62%**	**0.62%**	**1.21%**	**0.51%**	**0.82%**
		10^4	0.72%	1.63%	0.54%	0.85%	0.83%	2.12%	0.62%	1.30%
		10^5	1.21%	3.73%	0.96%	2.63%	1.28%	4.11%	1.07%	3.17%
	1	10^2	0.98%	2.20%	0.82%	0.98%	0.78%	1.65%	0.65%	1.17%
		$\mathbf{10^3}$	**0.71%**	**1.65%**	**0.58%**	**1.00%**	**0.75%**	**1.78%**	**0.57%**	**1.08%**
		10^4	1.16%	3.52%	0.90%	2.28%	1.17%	3.64%	0.91%	2.43%
		10^5	1.41%	4.68%	1.33%	4.31%	1.41%	4.68%	1.33%	4.29%
$\widehat{\boldsymbol{g}}^{2,0}_{\mathrm{MAP}}$	0.1	10^2	0.99%	0.68%	0.66%	0.32%	1.60%	1.14%	1.37%	0.60%
		10^3	0.61%	0.62%	0.42%	0.33%	0.75%	1.02%	0.59%	0.61%
		$\mathbf{10^4}$	**0.46%**	**0.60%**	**0.38%**	**0.36%**	**0.57%**	**1.16%**	**0.46%**	**0.62%**
		10^5	0.68%	1.50%	0.50%	0.79%	0.91%	2.68%	0.67%	1.43%
	0.05	10^2	1.03%	0.53%	1.01%	0.28%	1.76%	1.27%	1.56%	0.66%
		10^3	0.62%	0.51%	0.41%	0.28%	0.79%	1.24%	0.61%	0.62%
		$\mathbf{10^4}$	**0.44%**	**0.50%**	**0.35%**	**0.32%**	**0.55%**	**1.14%**	**0.45%**	**0.61%**
		10^5	0.57%	1.07%	0.43%	0.58%	0.82%	2.41%	0.61%	1.10%
	0.01	10^2	1.34%	1.02%	1.63%	0.45%	2.38%	3.45%	2.04%	1.55%
		10^3	0.73%	0.97%	0.52%	0.42%	1.09%	3.32%	0.89%	1.45%
		$\mathbf{10^4}$	**0.49%**	**0.94%**	**0.38%**	**0.44%**	**0.71%**	**3.29%**	**0.54%**	**1.42%**
		10^5	0.57%	1.68%	0.43%	0.65%	0.90%	3.68%	0.61%	2.23%

D. Real SPOT Images

To illustrate the proposed gain calibration method, a 6000×1500 panchromatic SPOT3 image has been processed, given that SPOT3 makes use of four linear arrays of 1500 detectors each to acquire images. Fig. 13 displays a small part of the image before and after correction using estimated gains $\widehat{\boldsymbol{g}}^{2,0}_{\mathrm{MAP}}$ with $s = 0.1$ and $\lambda = 10^4$. The striping effect is clearly visible on the original image (after enhancement), whereas it has totally disappeared after correction. The computation time is less than 4 s on a standard personal computer. Note that on such a subimage, one cannot see any difference between the images corrected using the estimated gains $\widehat{\boldsymbol{g}}^{\mathrm{AM}}$ and $\widehat{\boldsymbol{g}}^{2,0}_{\mathrm{MAP}}$. Indeed, compared with $\widehat{\boldsymbol{g}}^{\mathrm{AM}}$, the proposed estimator $\widehat{\boldsymbol{g}}^{2,0}_{\mathrm{MAP}}$ has been shown in Section VI-A to improve the estimation of the low frequencies of the gain, which correspond to large-scale errors in the image. As only a small part of the image is shown here, such estimation differences are not visible.

VII. CONCLUSION AND PERSPECTIVES

Imperfect detectors in a CCD linear array generate a striping effect when the array is used to scan images. This is a non-negligible source of degradation in pushbroom-type satellite imaging systems. However, compensation is easy provided that accurate estimation of detector responses has been achieved.

This paper has introduced a statistical self-calibration method to perform linear response correction. Self-calibration means that no specific training data are required, in contrast with commonly found approaches that involve specific calibration procedures. Our approach is based on a Markov random field model for the logarithm of pixel intensities. The best results are obtained with a nonconvex energy, e.g., when the *a posteriori* distribution is maximized with respect to the logarithm of detector gains.

The corresponding estimators have been shown well behaved in both theoretical and practical sense. They yield significantly higher precision than empirical estimation based on column averages at a computational cost that remains moderate.

The proposed method assumes a linear response of the detectors, and its efficiency depends on the validity of this model. For example, if the offset of an affine response has not been correctly calibrated, estimating the gains (and offsets) with a moment-matching-based method may give better results than with the proposed method. However, partial simulation results on the test images in Fig. 3 show that our method is relatively insensitive to offset residuals up to a value of 2 or 3. However, if the striping is due to the offsets but the gains are known, the proposed method can directly be applied to estimate the offsets (up to an additive constant) using the image itself instead of its logarithm. We are currently studying an extension of the method to jointly estimate the gains and offsets of the detectors. However, even if the practical results are encouraging, most of the theoretical properties presented in this paper collapse in this case. In particular, the criterion to minimize may have local minima, and convergence properties may not be guaranteed.

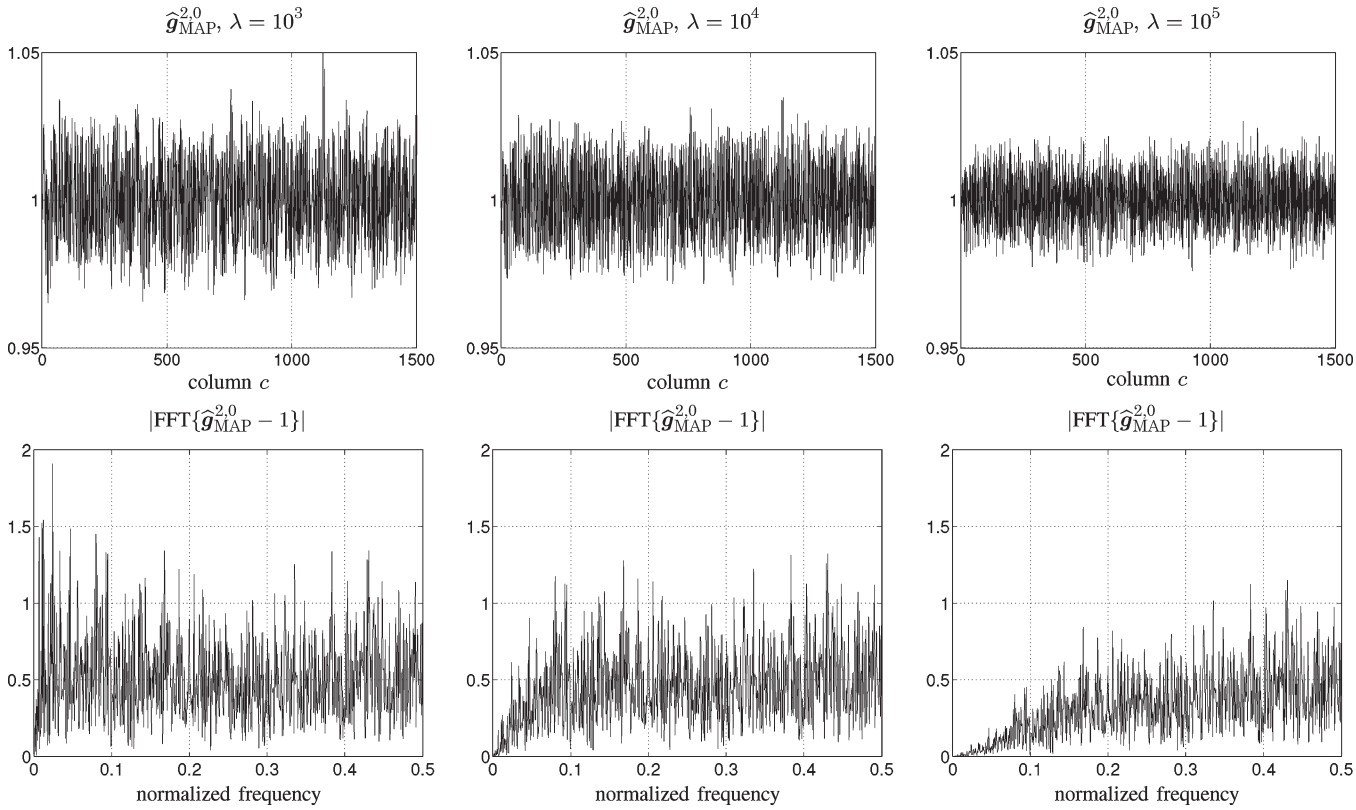

Fig. 10. Spatial and spectral representation of the gains estimated with the MAP estimator $\widehat{g}^{2,0}_{\mathrm{MAP}}$ with ϕ_{20} for $s = 0.1$ and various values of parameter λ; estimation performed on the first test image for $R = 3000$.

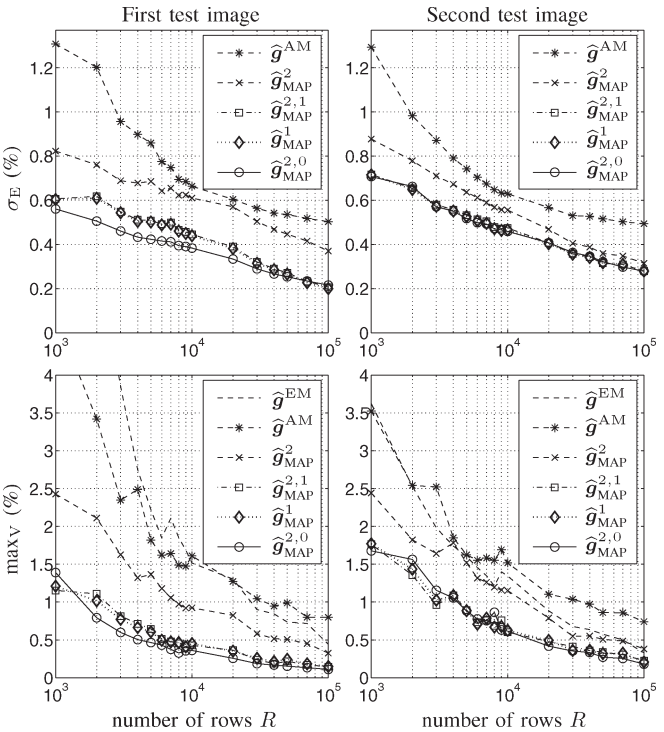

Fig. 11. Quality of the moment-matching and proposed estimators as a function of the number of rows on the first and second test images. $\lambda = 10^3$ for ϕ_2 and ϕ_1, $s = 10^{-2}$ and $\lambda = 10^3$ for ϕ_{21}, and $s = 0.1$ and $\lambda = 10^4$ for ϕ_{20}.

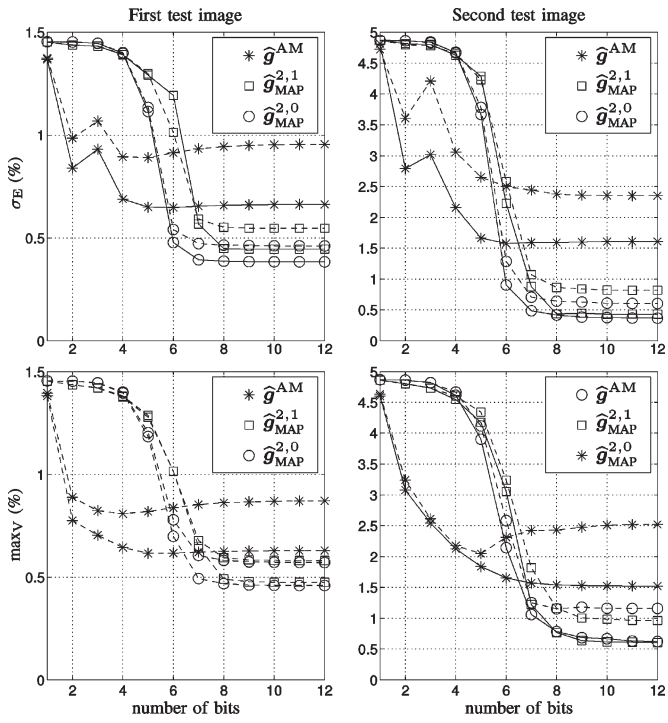

Fig. 12. Effect of quantization on the quality indices of the different estimators. Plain lines and dashed lines correspond to $R = 10000$ and $R = 3000$ rows, respectively. $s = 10^{-2}$ and $\lambda = 10^3$ for ϕ_{21}, and $s = 0.1$ and $\lambda = 10^4$ for ϕ_{20}.

Fig. 13. Zoom on a small part (200 × 200) of an original SPOT3 image before and after the image correction. Both images have been enhanced using xv (cf., http://www.trilon.com/xv) and the color map set from 50 to 150 to emphasize the striping effect.

Another extension of the proposed method to hyperspectral data may also be interesting, accounting for correlation in images acquired at different wavelength. Such an attempt is proposed in [14] for Hyperion data based on a moment-matching estimator.

APPENDIX I

A. IRLS

The IRLS algorithm can be justified within several theoretical frameworks. According to a unifying viewpoint, it can be seen as a particular instance of a majorize–minimize (MM) algorithm [35]. Let J be a real-valued criterion to be minimized such that

$$J(\widetilde{\boldsymbol{u}}) \leq K(\widetilde{\boldsymbol{u}}, \boldsymbol{u})$$
$$J(\boldsymbol{u}) = K(\boldsymbol{u}, \boldsymbol{u})$$

for some bivariate real-valued function K for all couples $\boldsymbol{u}$, $\widetilde{\boldsymbol{u}}$. The function $K(\cdot, \boldsymbol{u})$ is said to be a *majorizing approximation* of function J at $\boldsymbol{u}$. It is then easy to prove that $J(\widehat{\boldsymbol{u}}) \leq K(\widehat{\boldsymbol{u}}, \boldsymbol{u}) \leq J(\boldsymbol{u})$, where $\widehat{\boldsymbol{u}} = \arg\min_{\widetilde{\boldsymbol{u}}} K(\widetilde{\boldsymbol{u}}, \boldsymbol{u})$, so the following MM iteration produces a nonincreasing series $\{J(\boldsymbol{u}_k)\}$:

$$\boldsymbol{u}_{k+1} = \arg\min_{\boldsymbol{u}} K(\boldsymbol{u}_k, \boldsymbol{u}).$$

Under broad assumptions, MM algorithms are provably convergent toward a minimizer of J [35].

The interpretation of the IRLS algorithm as an MM scheme is a consequence of the following results [22], [28], [29]. Let ϕ be a $\mathcal{C}^1$ even function such that $\phi(\sqrt{\cdot})$ is concave over $\mathbf{R}^+$. Then, there exists ψ such that

$$\phi(u) = \inf_{w \in \mathbf{R}} \left(wu^2 + \psi(w) \right). \tag{15}$$

Moreover

$$\arg\min_{w} \left(wu^2 + \psi(w) \right) = \frac{\phi'(u)}{2u} \tag{16}$$

which is understood to be extended by continuity for $u = 0$.

As a consequence

$$K(\widetilde{\boldsymbol{g}}^\ell, \boldsymbol{g}^\ell) = \sum_{r,c} \left(w_{r,c} \left(\delta \widetilde{g}_c^\ell - \delta y_{r,c}^\ell \right)^2 + \psi(w_{r,c}) \right) + \lambda \sum_c \left(\widetilde{g}_c^\ell \right)^2$$

is a majorizing function for J_{MAP}, with each $w_{r,c}$ being a function of g_c^ℓ according to (12). Moreover, K is a quadratic function of its first argument, and minimizing it with respect to this argument amounts to solve the tridiagonal linear system (13).

B. A Unimodality Result for the Convolution Product of Unimodal Functions

Proposition 2: Let ϕ be a continuous even u-type unimodal function such that $\phi(0) = 0$, and p an even n-type unimodal probability density such that $\phi * p(0) < \infty$. Then, the convolution product $\phi * p$ is a continuous, even, and u-type unimodal function.

The main part of the proof follows similar lines to Purkayastha's proof of Wintner's result [36].

Proof: First, $\phi * p$ is continuous since ϕ is continuous and takes finite values since $\phi * p(0) < \infty$. Moreover, $\phi * p$ is an even function since both ϕ and p are even.

Let us show that $\phi * p$ is an increasing function on $[0, \infty)$. Suppose that X is a random variable with density p, i.e.,

$$\phi * p(x) = \mathrm{E}\left[\phi(X - x)\right] = \int_0^\infty \Pr(\phi(X - x) > u) du.$$

Now, given that ϕ is even and u-unimodal, $\{\phi(X - x) > u\}$ is the same event than $\{|X - x| > \phi^{-1}(u)\}$, where $\phi^{-1}(u)$ conventionally denotes the unique nonnegative solution of $\phi(v) = u$. Hence

$$\phi * p(x) = \int_0^\infty f(x, u) du \tag{17}$$

where

$$f(x,u) = \Pr\left(X < x - \phi(u)\right) + \Pr\left(X > x + \phi(u)\right)$$
$$= \int_{-\infty}^{x - \phi(u)} p(y) dy + \int_{x + \phi(u)}^{\infty} p(y) dy$$

so that

$$\frac{\partial f(x,u)}{\partial x} = p\left(x - \phi(u)\right) - p\left(x + \phi(u)\right)$$

which is positive for $x > 0$ for all values of u. Thus, f is an increasing function for all $x > 0$, and the same holds for $\phi * p$, given (17). ■

ACKNOWLEDGMENT

The authors would like to thank B. Beghin and P. Kubik for helpful discussions and the two anonymous reviewers for helpful comments on the initial version of this paper.

REFERENCES

[1] P. Mather, *Computer Processing of Remotely-Sensed Images: An Introduction*, 3rd ed. Hoboken, NJ: Wiley, 2004.
[2] R. Srinivasan and M. C. J. White, "Landsat data destriping using power spectral filtering," *Opt. Eng.*, vol. 27, no. 11, pp. 939–943, Nov. 1988.
[3] R. E. Crippen, "A simple spatial filtering routine for the cosmetic removal of scan-line noise from LANDSAT TM P-tape imagery," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 327–331, Mar. 1989.
[4] D. L. Helder, B. K. Quirk, and J. J. Hood, "A technique for the reduction of banding in LANDSAT Thematic Mapper images," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 10, pp. 1425–1431, Oct. 1992.
[5] J. J. Simpson, J. R. Stitt, and D. M. Leath, "Improved finite impulse response filters for enhanced destriping of geostationary satellite data," *Opt. Eng.*, vol. 66, no. 3, pp. 235–249, Dec. 1998.
[6] J. Chen, Y. Shao, H. Guo, W. Wang, and B. Zhu, "Destriping CMODIS data by power filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2119–2124, Sep. 2003.
[7] J. Torres and S. O. Infante, "Wavelet analysis for the elimination of striping noise in satellite images," *Opt. Eng.*, vol. 40, no. 7, pp. 1309–1314, Jul. 2001.
[8] J. Chen, H. Lin, Y. Shao, and L. Yang, "Oblique striping removal in remote sensing imagery based on wavelet transform," *Int. J. Remote Sens.*, vol. 27, no. 8, pp. 1717–1723, Apr. 2006.
[9] J. J. Simpson and S. R. Yhann, "Reduction of noise in AVHRR channel 3 data with minimum distortion," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 2, pp. 315–328, Mar. 1994.
[10] B. K. P. Horn and R. J. Woodham, "Destriping LANDSAT MSS images by histogram modification," *Comput. Graph. Image Process.*, vol. 10, pp. 69–83, 1979.
[11] V. R. Algazi and G. E. Ford, "Radiometric equalization of nonperiodic striping in satellite data," *Comput. Graph. Image Process.*, vol. 16, pp. 287–295, 1981.
[12] F. L. Gadallah, G. Csillag, and E. J. M. Smith, "Destriping multisensor imagery with moment matching," *Int. J. Remote Sens.*, vol. 21, no. 12, pp. 2505–2511, Aug. 2000.
[13] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.
[14] L. Sun, R. Neville, K. Staenz, and P. White, "Automatic destriping of Hyperion imagery based on spectral moment matching," *Can. J. Remote Sens.*, vol. 34, pp. S68–S81, 2008, Suppl. 1.
[15] D. J. Poros and C. J. Peterson, "Methods for destriping Landsat Thematic Mapper images—A feasibility study for an online destriping process in the Thematic Mapper Image Processing System (TIPS)," *Photogramm. Eng. Remote Sens.*, vol. 51, no. 9, pp. 1371–1378, Nov. 1985.
[16] M. Wegener, "Destriping multiple sensor imagery by improved histogram matching," *Int. J. Remote Sens.*, vol. 11, no. 5, pp. 859–875, May 1990.
[17] P. Rakwatin, W. Takeuchi, and Y. Yasuoka, "Stripe noise reduction in MODIS data by combining histogram matching with facet filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1844–1856, Jun. 2007.
[18] P. Antonelli, M. di Bisceglie, R. Episcopo, and C. Galdi, "Destriping MODIS data using IFOV overlapping," in *Proc. Int. Geosci. Remote Sens. Symp.*, Madison, WI, 2004, vol. 7, pp. 4568–4571.
[19] G. Corsini, M. Diani, and T. Walzel, "Striping removal in MOS-B data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1439–1446, May 2000.
[20] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 6, pp. 576–586, Jun. 1995.
[21] G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.
[22] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.
[23] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," in *Proc. 46th Session ICI, Bull. ICI*, 1987, vol. 52, pp. 5–21.
[24] H. Shen and L. Zhang, "A MAP-based algorithm for destriping and in painting of remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1492–1502, May 2009.
[25] H. Carfantan, J. Idier, B. Beghin, A. Meygret, and B. Rougé, "Statistical self-calibration of SPOT satellite imaging instrument," in *AIP Conf. Proc., Bayesian Inference Maximum Entropy Methods*, A. Mohammad-Djafari, Ed., Gif-sur-Yvette, France, Jul. 2000, vol. 568, pp. 501–512.
[26] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, pp. 147–185, 1974.
[27] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. Int. Conf. Image Process.*, Austin, TX, Nov. 1994, vol. 2, pp. 168–172.
[28] M. Allain, J. Idier, and Y. Goussard, "On global and local convergence of half-quadratic algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1130–1142, May 2006.
[29] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.
[30] J.-J. Fuchs, "Convergence of a sparse representations algorithm applicable to real or complex data," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 598–605, Dec. 2007.
[31] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 204–221, Feb. 1998.
[32] H. Strasser, "Consistency of maximum likelihood and Bayes estimates," *Ann. Stat.*, vol. 9, no. 5, pp. 1107–1113, 1981.
[33] D. Dacunha-Castelle and M. Duflo, *Probability and Statistics*, vol. 2. New York: Springer-Verlag, 1986.
[34] A. Wintner, *Asymptotic Distributions and Infinite Convolutions*. Ann Arbor, MI: Edwards, 1938.
[35] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Stat.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.
[36] S. Purkayastha, "Simple proofs of two results on convolutions of unimodal distributions," *Stat. Prob. Lett.*, vol. 39, no. 2, pp. 97–100, Aug. 1998.

Hervé Carfantan (M'06) was born in France in 1968. He received the Engineer degree in control and signal processing from the École Supérieure d'Infomatique Électronique Automatique, Paris, France, in 1992 and the M.S. degree in control and signal processing and the Ph.D. degree in physics from the University of Paris-Sud, Orsay, France, in 1992 and 1996, respectively.

He is currently an Assistant Professor with the Université Paul Sabatier, Toulouse, France, and a Researcher with the Signal, Image and Instrumentation Team, Laboratoire d'Astrophysique de Toulouse-Tarbes, which is part of the Université de Toulouse and the Centre National de la Recherche Scientifique (CNRS). His main interests are in inverse problems, estimation, spectral analysis, and optimization. His application fields essentially concern astronomical and geophysical data analysis.

Jérôme Idier was born in France, in 1966. He received the Engineer degree in electrical engineering from the École Supérieure d'Électricité, Gif-sur-Yvette, France, in 1988 and the Ph.D. degree in physics from the University of Paris-Sud, Orsay, France, in 1991.

Since 1991, he has been with the Centre National de la Recherche Scientifique. He is currently a Senior Researcher with the Institut de Recherche en Communications et Cybernétique de Nantes, École Centrale de Nantes, Nantes, France. His major scientific interests are in probabilistic approaches to inverse problems for signal and image processing.