

PAPER • OPEN ACCESS

Blind image fusion for hyperspectral imaging with the directional total variation

To cite this article: Leon Bungert *et al* 2018 *Inverse Problems* **34** 044003

View the [article online](#) for updates and enhancements.

Related content

- [Sparsity regularization for parameter identification problems](#)
Bangti Jin and Peter Maass
- [A hybrid stochastic-deterministic gradient descent algorithm for image reconstruction in cone-beam computed tomography](#)
Davood Karimi and Rabab K Ward
- [On the convergence of a linesearch based proximal-gradient method for nonconvex optimization](#)
S Bonettini, I Loris, F Porta et al.

Blind image fusion for hyperspectral imaging with the directional total variation

Leon Bungert¹, David A Coomes², Matthias J Ehrhardt³,
Jennifer Rasch⁴, Rafael Reisenhofer⁵
and Carola-Bibiane Schönlieb³

¹ Department of Mathematics, Friedrich-Alexander University Erlangen-Nürnberg, Cauerstr. 11, 91058 Erlangen, Germany

² Forest Ecology and Conservation Group, Department of Plant Sciences, University of Cambridge, Cambridge, CB2 3EA, United Kingdom

³ Department for Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, CB3 0WA, United Kingdom

⁴ Fraunhofer Heinrich Hertz Institute, Einsteinufer 37, 10587 Berlin, Germany

⁵ University of Bremen, Fachbereich 3, Postfach 330440, 28334 Bremen, Germany

E-mail: m.j.ehrhardt@damtp.cam.ac.uk

Received 3 October 2017, revised 21 December 2017

Accepted for publication 14 February 2018

Published 8 March 2018



CrossMark

Abstract

Hyperspectral imaging is a cutting-edge type of remote sensing used for mapping vegetation properties, rock minerals and other materials. A major drawback of hyperspectral imaging devices is their intrinsic low spatial resolution. In this paper, we propose a method for increasing the spatial resolution of a hyperspectral image by fusing it with an image of higher spatial resolution that was obtained with a different imaging modality. This is accomplished by solving a variational problem in which the regularization functional is the directional total variation. To accommodate for possible mis-registrations between the two images, we consider a non-convex blind super-resolution problem where both a fused image and the corresponding convolution kernel are estimated. Using this approach, our model can realign the given images if needed. Our experimental results indicate that the non-convexity is negligible in practice and that reliable solutions can be computed using a variety of different optimization algorithms. Numerical results on real remote sensing data from plant sciences and urban monitoring show the potential of the proposed method and suggests that it is robust with respect to the regularization parameters, mis-registration and the shape of the kernel.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: remote sensing, super-resolution, pansharpening, blind deconvolution, hyperspectral imaging

(Some figures may appear in colour only in the online journal)

1. Introduction

Hyperspectral imaging is an earth observation technique that measures the energy of light reflected off the earth's surface within many narrow spectral bands, from which reflectance curves can be calculated. The shape of these reflectance curves provides information about the chemical and physical properties of materials such as minerals in rocks, vegetation, synthetic materials and water, which allows these materials to be classified and mapped remotely (see [1, 2], for instance).

Usually, hyperspectral images of the earth are being recorded by imaging devices that are mounted on planes or satellites. In either case, the speed of the carrier and narrowness of the spectral bands result in limited energy reaching the sensor for each spectral band. Therefore, most hyperspectral cameras have an intrinsic trade-off between spatial and spectral resolution [3, 4]. A standard approach to overcoming this trade-off is to record a second image with low spectral but high spatial resolution (e.g. a photograph or a laser scan) [5] and to fuse these data together in order to create an image that has the high spectral resolution of the hyperspectral image and the high spatial resolution of the photograph [3–13]. These techniques have become known as *pansharpening* or *image fusion*. The problem is illustrated by three example data sets in figure 1.

Most image fusion techniques are based on certain assumptions regarding the data and the corresponding imaging devices. First, it is often assumed that the image of high spatial resolution is a linear combination of the spectral channels with known weights [4]. Second, the loss of resolution is usually modeled as a linear operator which consists of a subsampled convolution with known kernel (point spread function). While both assumptions may be justified in some applications, it may be difficult to measure or estimate the weights and the convolution kernel in a practical situation.

Instead of having accurate descriptions or measurements about the system response of our cameras [4, 6, 12], we make use of the fact that both images are acquired from the same scenery. While the actual image intensities will depend on the wavelengths that are being recorded and are therefore dissimilar, the geometric structure of those images is likely to be very similar as they are both taken from the same terrain [6, 7, 9, 12]. This approach has been studied extensively in the context of medical imaging where traditionally an anatomical imaging modality with high spatial resolution is used to aid the reconstruction of a functional imaging modality, see e.g. [14–20].

Note that the hyperspectral data image and the high-resolution photograph usually undergo a georectification procedure which aligns them using a digital elevation map of the study area. However, this adjustment is often imprecise, particularly in topographically complex regions, so further steps are required to co-register the images precisely [21]. Moreover, having a good description of the point spread function of the system is very unnatural. Thus, we take a blind approach and estimate the point spread function within our reconstruction and thereby learn this information from the recorded data. This approach is intrinsically related to blind deconvolution [22–25]. Figure 2 shows an example of blind versus non-blind image fusion applied to the real remote sensing data with unknown blurring kernel. It can be seen that the shape of the kernel that is estimated during the reconstruction clearly differs from the fixed kernel. Equally important is the translation of the estimated kernel, which can compensate for shifts between the given image pair. To the best of our knowledge this is the first contribution where structural

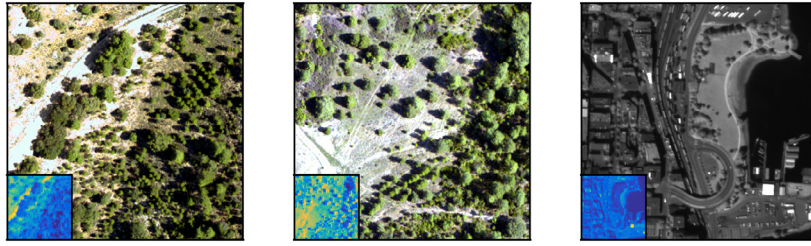


Figure 1. Three example data for image fusion in remote sensing. They each consist of a hyperspectral image (small image, only one channel shown) and an image of higher spatial resolution (large image). The goal is to create an image that has both high spatial and high spectral resolution.

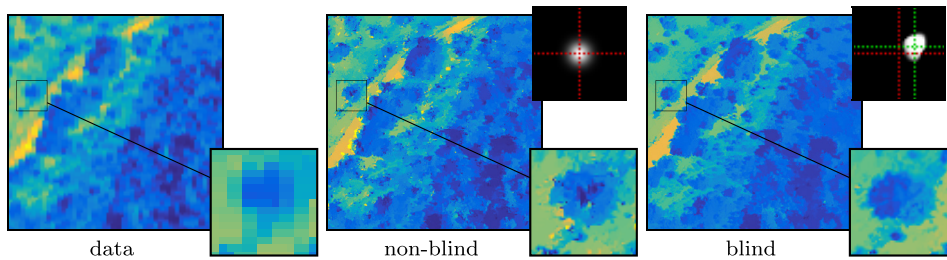


Figure 2. Comparison of *non-blind* versus *blind* image fusion. From left to right: hyperspectral image (one channel, four times enlarged), result of non-blind and blind image fusion together with the kernels that are used or estimated, respectively. The plotted kernels are of size 41×41 and have been enlarged for better visibility. The dotted green lines highlight the center of mass of the obtained kernel and indicate the amount of translational displacement between the fused images.

side information / anatomical priors are combined with registration which removes the unnatural assumption of perfectly registered images that is usually made, see [14–20], for instance.

2. Mathematical model and notation

In this work we consider the problem of simultaneous image fusion and blind image deconvolution which can be cast as the optimization problem

$$(u^*, k^*) \in \arg \min_{(u,k) \in U \times K} \left\{ \frac{1}{2} \|\mathbf{A}_k u - f\|^2 + \mathcal{R}_u(u) + \mathcal{R}_k(k) \right\} \quad (1)$$

where the *forward operator* \mathbf{A}_k is a sampled blur with kernel k . Thus, it relates an image $u \in U := \mathbb{R}^m$ of size $m = (m_1, m_2)$ to blurred and subsampled data $f \in \mathbb{R}^n, n = (n_1, n_2)$. Note that we used *multi-index notation* which means $\mathbb{R}^m = \mathbb{R}^{m_1 \times m_2}$, for instance. Here, we consider the case of *super-resolution* $n < m$ (true if and only if $n_1 < m_1$ and $n_2 < m_2$). The data discrepancy is measured in terms of the Euclidean / Frobenius norm $\|x\|^2 := \sum_i x_i^2$ and the optimal solution is *regularized* through the penalty functionals \mathcal{R}_u and \mathcal{R}_k .

Since—up to now—the ingredients of our model are only vaguely defined, the remainder of this section is devoted to a detailed explanation of the operators and regularization functionals involved.

2.1. The forward operator

The forward operator $\mathbf{A}_k := \mathbf{S} \circ \mathbf{B} \circ \mathbf{C}_k$ is the composition of three operators: \mathbf{C}_k represents a *convolution* with a kernel k , \mathbf{B} performs *boundary treatment* and \mathbf{S} is a *sampling operator*.

The convolution operator $\mathbf{C}_k : U \rightarrow U$ with kernel $k \in K := \mathbb{R}^r$, $r := (r_1, r_2)$ is defined by

$$\mathbf{C}_k u := \mathbf{J}k * u \quad (2)$$

where $*$ is the cyclic convolution of images in $U = \mathbb{R}^m$. Since the kernel k is generally assumed to have smaller support than the image u , i.e. $r < m$, we introduced the embedding operator $\mathbf{J} : K \rightarrow U$ which embeds the ‘small’ convolution kernel $k \in K$ into the image space U by padding it with zeros. By assuming periodic boundary conditions on u and $\mathbf{J}k$, the convolution can be efficiently computed via the discrete Fourier transform \mathbf{F} , i.e. $\mathbf{C}_k u = \mathbf{F}^{-1}(\mathbf{F}\mathbf{J}k \odot \mathbf{F}u)$ where \odot denotes the Hadamard product (pointwise multiplication). With a slight abuse of notation, we can exploit the symmetry of the convolution to infer

$$\mathbf{C}_k u = \mathbf{J}k * u = u * \mathbf{J}k = \mathbf{C}_u \mathbf{J}k, \quad (3)$$

and thus

$$\mathbf{A}_k u = \mathbf{A}_u \mathbf{J}k, \quad (4)$$

which will prove helpful later-on.

While periodic boundary conditions are computationally useful, any kind of boundary treatment is artificial and may result in artifacts. For instance, boundary artifacts are expected in the boundary region with margins $l := (r - 1)/2$ where r is the diameter of the convolution kernel. To avoid these artefacts, we follow the *boundary-layer approach* (see [26] and references therein; also described by Almeida and Figueiredo in [27]) and introduce a boundary clipping (margin deletion) operator $\mathbf{B} : U \rightarrow \mathbb{R}^{m-2l}$,

$$(\mathbf{B}u)_i := u_{i+l} \quad (5)$$

that maps a large image u to a *meaningful image* $\mathbf{B}u \in \mathbb{R}^{m-2l}$ which is independent of the type of boundary conditions used for the precedent convolution. Thus, after computation of an optimal solution u^* of (1) we only use the meaningful part $\mathbf{B}u^*$.

Finally, the third component of our forward operator is the sampling $\mathbf{S} : \mathbb{R}^{m-2l} \rightarrow \mathbb{R}^n$ which mimics an integration sensor of size $s \times s$, i.e.

$$(\mathbf{S}u)_i := \frac{1}{s^2} \sum_{j \in \Delta_i} u_j \quad (6)$$

where the detector indices are given by the set $\Delta_i := \{j = (j_1, j_2) \mid i \leq j < i + s\}$, which contains s^2 elements. Consequently, for a kernel of odd dimensions the relation $m - 2l = sn$ connects the sizes of kernel, image and data.

Let us now turn to the regularization and prior knowledge on the image u and the kernel k .

2.2. Regularization functionals

2.2.1. Image regularization. A popular prior to regularize images is the *total variation* [28] as it allows the recovery of smooth images whilst preserving edges. It is defined as the global 1-norm of the discretized gradient

$$\text{TV}(u) := \sum_i \|\nabla u_i\|. \quad (7)$$

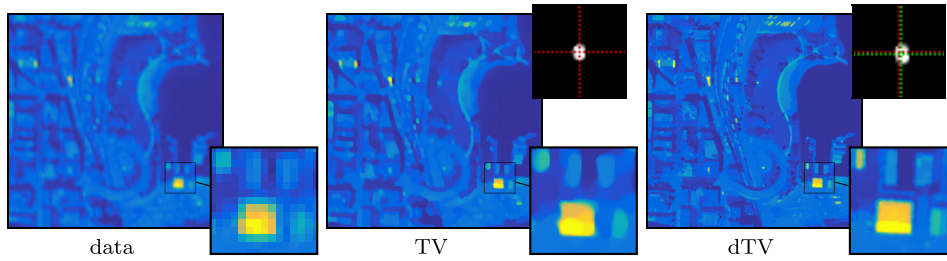


Figure 3. Comparison of total variation (TV) versus directional total variation (dTV) regularization for image fusion. From left to right: hyperspectral image (one channel, four times enlarged), results of TV and dTV together with the estimated kernels.

Due to the periodic boundary conditions needed for the forward operator, the discretized gradient (with forward differences) $\nabla u \in U^2$ of an image $u \in U$ is

$$(\nabla u)_j := u_{\text{mod}(i+e_j, m)} - u_i, \quad j \in \{1, 2\}, \quad (8)$$

where $e_1 = (1, 0)$ and $e_2 = (0, 1)$ are the standard unit vectors in \mathbb{R}^2 .

In the task at hand, the high-resolution photograph can be viewed as a source of additional *a priori* knowledge about the sought solution. As both the hyperspectral data and the high-resolution photograph depict the same scenery, it is reasonable to assume that they show the similar geometrical structures albeit having very different intensities. However, the TV functional (7) is not able to incorporate this kind of additional prior information. Instead, we will apply the so-called directional TV (dTV), which can fuse the structural information of a high-resolution image with the intensities of a low-resolution channel of a hyperspectral image. Figure 3 provides a visual comparison of reconstructions obtained from applying the TV and the dTV functional, respectively.

Definition (Directional total variation [18]). Let $\xi \in U^2$, $\|\xi_i\| \leq \gamma < 1$ be a vector-field that determines directions and denote by $\mathbf{P} \in U^{2 \times 2}$, $\mathbf{P}_i := \mathbf{I} - \xi_i \otimes \xi_i$ an associated matrix-field where \mathbf{I} is the 2×2 identity matrix and \otimes marks the outer product of vectors. Then the *directional total variation* $\text{dTV} : U \rightarrow \mathbb{R}$ is defined as as

$$\text{dTV}(u) := \sum_i \|\mathbf{P}_i \nabla u_i\|. \quad (9)$$

Note that due to the linearity of \mathbf{P}_i the directional total variation is *convex*. Let us shortly comment on the interpretation of the directional total variation. The vector-field ξ allows the definition of directions that should be less penalized and therefore get promoted. It is easy to see that the quantity $\mathbf{P}_i \nabla u_i$ can be expanded as

$$\mathbf{P}_i \nabla u_i = \nabla u_i - \langle \xi_i, \nabla u_i \rangle \xi_i \quad (10)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on \mathbb{R}^2 . This expression reduces to $(1 - \|\xi_i\|^2) \nabla u_i$ in regions where ∇u_i is collinear to ξ_i , and to ∇u_i where ∇u_i is orthogonal to ξ_i . Thus, gradients that are aligned / collinear ξ_i are favored as long as $\|\xi_i\| > 0$. In the extreme case $\|\xi_i\| = 1$, aligned gradients are not penalized any more. It is important to note, that this prior *does not enforce* gradients in the direction of ξ_i , as a gradient in the direction of ξ_i is never ‘cheaper’ than a zero gradient.

The uniform upper bound $\gamma < 1$ ensures that the directional total variation is equivalent to the total variation in a semi-norm sense, i.e.

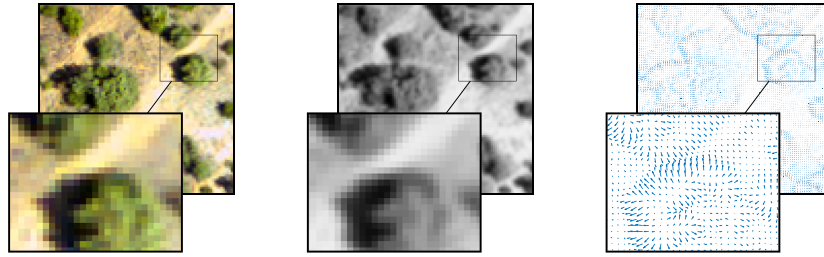


Figure 4. Side information for directional total variation. The side information is given as an RGB image (left) and converted into gray scales (center) from which a vector-field may be computed (right), e.g. with (12) (shown for $\gamma = 1, \varepsilon = 0.3$). The close-ups show that the vector-field does not only carry detailed information about edge *locations* but also about edge *directions*.

$$(1 - \gamma^2)\text{TV}(u) \leq \text{dTV}(u) \leq \text{TV}(u). \quad (11)$$

Similar versions of the directional total variation have been used before [17, 19, 29] and it is related to other anisotropic priors [14, 30, 31] and the notion of parallel level sets [32–34]. Note that the directional total variation generalizes the usual total variation (7) for $\xi = 0$ and other versions of the directional total variation [35, 36] where the direction ξ is constant and not depending in the pixel location i . With the isotropic choice $\mathbf{P}_i = w_i \mathbf{I}$, $0 \leq w_i \leq 1$ it reduces to weighted total variation [18, 29, 37–39].

Having this definition of the directional total variation at hand, the question is how to define the vector-field $\xi \in U^2$. As we want to favor images having a similar structure as the high-resolution photograph, we let $v \in U$ denote a gray-scale image which is generated from the color high-resolution image and define the vector-field

$$\xi_i := \gamma \frac{\nabla v_i}{\|\nabla v_i\|_\varepsilon}, \quad (12)$$

where $\|x\|_\varepsilon := \sqrt{\|x\|^2 + \varepsilon^2}$ and the scaling $\gamma \in [0, 1)$ ensures that $\|\xi_i\| \leq \gamma$. The parameter $\varepsilon > 0$ guarantees that the vector-field ξ is well-defined, even if $\nabla v_i = 0$. Moreover, it allows to control which gradient magnitudes $\|\nabla v_i\|$ should be considered meaningful and which are merely due to noise. Thus, both parameters ε and γ in general will depend on the given side information v but can be chosen without considering a particular data set f . Note that in practice, one chooses γ very close or even equal to 1 (see [18] for the latter) such that the side information image has sufficiently much influence. Extensive studies of this parameter are performed in [36] where the authors obtain their best results for small values of $a := 1 - \gamma^2$. An example photograph, its gray scale version and the corresponding vector-field are shown in figure 4.

In addition to the structural *a priori* knowledge given by the photograph and encoded in the directional total variation, it is reasonable to assume that the solution we are looking for is non-negative. Thus, the regularization functional for the image becomes

$$\mathcal{R}_u := \lambda_u \text{dTV} + \iota_{[0, \infty)^m} \quad (13)$$

where the characteristic function of the non-negative quadrant is defined as

$$\iota_{[0, \infty)^m}(u) := \begin{cases} 0 & \text{if } u \in [0, \infty)^m \\ \infty & \text{else} \end{cases}. \quad (14)$$

The regularization parameter for the image $\lambda_u > 0$ controls the trade-off between regularity and data fidelity.

2.3. Kernel regularization

Similarly as for the image u , it is reasonable to assume that the kernel is non-negative and regular. Motion blurs that arise from ideal diffraction-free lenses have typically a compactly supported kernel with sharp cut-off boundaries [26, p 234]. Thus, the desired regularity of the kernel may be achieved by utilizing the total variation. Of course, other smoothness priors such as the H^1 -semi-norm or total generalized variation [40] are possible, too, and will be subject of future work. In addition, blind deconvolution has the intrinsic non-uniqueness in the problem formulation that does not allow to estimate the correct global scaling factors for both the kernel and the image. To circumvent this issue, a common choice (e.g. [24–26, 41]) is to normalize the kernel with respect to the 1-norm. Thus, together with the non-negativity, we assume that the kernel is in the unit simplex

$$\mathbb{S} := \left\{ k \in K : k_i \geq 0, \sum_i k_i = 1 \right\}. \quad (15)$$

Letting $\iota_{\mathbb{S}}$ denote the characteristic function of the unit simplex, defined analogously to (14), the regularization functional for the kernel becomes

$$\mathcal{R}_k := \lambda_k \text{TV} + \iota_{\mathbb{S}} \quad (16)$$

where we again can trade-off regularity and data fidelity by the regularization parameter $\lambda_k > 0$ which in general needs to be chosen differently than the regularization parameter for the image λ_u .

3. Algorithms

Given the abstract mathematical model (1) which models the solution that we are after, this section discusses algorithms that will attempt to numerically compute this solution. For a concise presentation, we cast problem (1) in the form

$$(u^*, k^*) \in \arg \min_{u, k} \Psi(u, k) \quad (17)$$

by defining $\Psi(u, k) := \mathcal{D}(u, k) + \mathcal{R}_u(u) + \mathcal{R}_k(k)$ and abbreviating the data fidelity term in (1) by $\mathcal{D}(u, k) := \frac{1}{2} \|\mathbf{A}_k u - f\|^2$. While the function Ψ is not convex in the joint variable (u, k) , it is bi-convex, i.e. it is convex in each of the variables u and k . The latter holds true as the regularization functions \mathcal{R}_u and \mathcal{R}_k are convex and the data fidelity \mathcal{D} is bi-convex.

An abstract algorithm for finding solutions of (17) is the *proximal alternating minimization (PAM)* algorithm [42]. Given an image-kernel pair (u, k) and step sizes $\tau_u, \tau_k > 0$, it computes the next iterate (u^+, k^+) by the update scheme

$$u^+ = \arg \min_x \left\{ \frac{1}{2} \|x - u\|^2 + \tau_u \Psi(x, k) \right\}, \quad (18a)$$

$$k^+ = \arg \min_x \left\{ \frac{1}{2} \|x - k\|^2 + \tau_k \Psi(u^+, x) \right\}. \quad (18b)$$

Under reasonably mild conditions on the regularity of Ψ (see [42] for more details), PAM monotonously decreases the objective and its iterates converge to a critical point of Ψ . As the overall problem is non-convex, we cannot hope to find global solutions with this strategy.

3.1. Proximal alternating linearized minimization (PALM)

In addition, we may exploit the fact that the data term \mathcal{D} is continuously differentiable with respect to u and k and that it is not very difficult to compute the *proximal operators* with respect to the regularization functionals \mathcal{R}_u and \mathcal{R}_k . The proximal operator of a functional \mathcal{R} is defined as the solution operator to a denoising problem with regularization term \mathcal{R}

$$\text{prox}_{\mathcal{R}}(y) := \arg \min_x \left\{ \frac{1}{2} \|x - y\|^2 + \mathcal{R}(x) \right\}. \quad (19)$$

These extra information can be utilized with the *proximal alternating linearized minimization (PALM)* [43]. Linearizing the differentiable part \mathcal{D} of the objective function Ψ in the update of (18) yields the new update

$$u^+ = \text{prox}_{\tau_u \mathcal{R}_u} \left(u - \tau_u \nabla_u \mathcal{D}(u, k) \right), \quad (20a)$$

$$k^+ = \text{prox}_{\tau_k \mathcal{R}_k} \left(k - \tau_k \nabla_k \mathcal{D}(u^+, k) \right). \quad (20b)$$

It can be interpreted both as a linearized version of PAM and as an alternating form of forward-backward splitting [44, 45]. Note that PALM also monotonously decreases the objective and has the same convergence guarantees as long as the step size parameters are well-chosen which we will discuss later in this section.

3.2. Inertial PALM (iPALM)

Local methods for non-convex optimization may lead to critical points that are not global minimizers. The authors of [46] proposed an inertial variant of forward-backward splitting which next to the gradient direction gives inertia to steer the iterates in the previously chosen direction. They empirically showed on a simple two dimensional problem that inertia may help to avoid spurious critical points. An inertial version of PALM [41], called *iPALM*, follows the updates

$$u_\alpha = u + \alpha(u - u^-), \quad (21a)$$

$$u^+ = \text{prox}_{\tau_u \mathcal{R}_u} \left(u_\alpha - \tau_u \nabla_u \mathcal{D}(u_\alpha, k) \right), \quad (21b)$$

$$k_\alpha = k + \alpha(k - k^-), \quad (21c)$$

$$k^+ = \text{prox}_{\tau_k \mathcal{R}_k} \left(k_\alpha - \tau_k \nabla_k \mathcal{D}(u^+, k_\alpha) \right), \quad (21d)$$

with inertia parameter $\alpha \in [0, 1)$. Next to the current iterate (u, k) it also depends on the previous iterate (u^-, k^-) . We remark that PALM can be recovered by vanishing inertial parameter $\alpha = 0$. Also note that one may define iPALM not with one inertial parameter α but with four

different ones [41]. While iPALM is *not* guaranteed to decrease the objective Ψ , it monotonously decreases a surrogate functional that relates to Ψ [41] and its iterates are guaranteed to converge to a critical point of Ψ for well-chosen step size and inertia parameter.

3.3. Step sizes and backtracking

The convergence of PALM and iPALM depend on the step size parameters τ_u, τ_k . Following [41], we choose the step size parameters as

$$\tau_x := \frac{1 - \alpha}{1 + 2\alpha} \frac{2}{\theta L_x}, \quad x \in \{u, k\}, \quad (22)$$

where $\theta > 1$ is a global constant and L_x is the local Lipschitz constant of the derivative of the data fit \mathcal{D} with respect to x . Using the relation (4), its derivatives with respect to the image and kernel are

$$\nabla_u \mathcal{D}(u, k) = \mathbf{A}_k^* (\mathbf{A}_k u - f), \quad (23a)$$

$$\nabla_k \mathcal{D}(u, k) = \mathbf{J}^* \mathbf{A}_u^* (\mathbf{A}_u \mathbf{J} k - f) \quad (23b)$$

which shows that L_u will depend on k and L_k on u . While it is possible to conservatively estimate these constants, standard estimates turn out to be highly pessimistic and one can do significantly better by determining L_u and L_k with a backtracking scheme [46, 47]. The actual local Lipschitz constants L_u and L_k satisfy the descent inequalities (also called descent lemma) [41, 43, 46, 47]

$$\mathcal{D}(u^+, k) \leq \mathcal{D}(u_\alpha, k) + \langle \nabla_u \mathcal{D}(u_\alpha, k), u^+ - u_\alpha \rangle + \frac{L_u}{2} \|u^+ - u_\alpha\|^2, \quad (24a)$$

$$\mathcal{D}(u^+, k^+) \leq \mathcal{D}(u^+, k_\alpha) + \langle \nabla_k \mathcal{D}(u^+, k_\alpha), k^+ - k_\alpha \rangle + \frac{L_k}{2} \|k^+ - k_\alpha\|^2. \quad (24b)$$

If during the iterations these inequalities are not satisfied, we increase the Lipschitz estimates by a constant factor $\eta > 1$ and repeat the iteration.

In addition to the descent inequality, an inexact evaluation of the proximal operators (e.g. due to finite iterations) may also hinder convergence. Thus, we will evaluate the proximal operators to a precision such that the proximal descent inequalities

$$\mathcal{R}_u(u^+) \leq \mathcal{R}_u(u) + \langle \nabla_u \mathcal{D}(u_\alpha, k), u - u^+ \rangle + \frac{1}{2\tau_u} (\|u - u_\alpha\|^2 - \|u^+ - u_\alpha\|^2), \quad (25a)$$

$$\mathcal{R}_k(k^+) \leq \mathcal{R}_k(k) + \langle \nabla_k \mathcal{D}(u^+, k_\alpha), k - k^+ \rangle + \frac{1}{2\tau_k} (\|k - k_\alpha\|^2 - \|k^+ - k_\alpha\|^2) \quad (25b)$$

hold. Equations (24), (25) and the parameters L_u, L_k guarantee the monotonic descent of PALM and iPALM with respect to the objective or its surrogate [41, 43].

The pseudo code for PALM and iPALM with this backtracking can be found in algorithm 1.

Algorithm 1. PALM / iPALM with backtracking. PALM is recovered by setting the inertial parameter $\alpha = 0$. The parameters $\underline{L}, \bar{L}, \eta, \theta$ can all be chosen different for the image u and kernel k but for simplicity we chose not to.

Input: initial iterates u, k , inertial parameter $\alpha \in [0, 1)$, initial step size parameters L_u, L_k , constant for step size $\theta > 1$, bounds for step size parameters $0 < \underline{L} \leq \bar{L} < \infty$, backtracking constant $\eta > 1$

```

1: function iPALM
2:    $u^- \leftarrow u, k^- \leftarrow k$ 
3:   for  $t = 0, 1, \dots$  do
4:     % update image  $u$ 
5:      $u_\alpha \leftarrow u + \alpha(u - u^-)$ 
6:      $(u^+, L_u) \leftarrow \text{BACKTRACKING}(u_\alpha, \nabla_u \mathcal{D}(u_\alpha, k), L_u)$ 
7:      $u^- \leftarrow u, u \leftarrow u^+$ 
8:     % update kernel  $k$ 
9:      $k_\alpha \leftarrow k + \alpha(k - k^-)$ 
10:     $(k^+, L_k) \leftarrow \text{BACKTRACKING}(k_\alpha, \nabla_k \mathcal{D}(u^+, k_\alpha), L_k)$ 
11:     $k^- \leftarrow k, k \leftarrow k^+$ 
12:  return  $(u^+, k^+)$ 
13: function BACKTRACKING( $x_\alpha, g, L$ )
14:  while True do
15:    % select step size
16:     $\tau \leftarrow \frac{1-\alpha}{1+2\alpha} \frac{2}{\theta L}$ 
17:    % forward-backward step where the proximal operator
18:    % is approximated by a fixed number of warm-started iterations
19:     $x^+ \leftarrow \text{prox}_{\tau \mathcal{R}}(x_\alpha - \tau g)$ 
20:    if Descent inequality (24) does not hold then
21:      % increase Lipschitz constant and try again
22:       $L \leftarrow \min(\eta L, \bar{L})$ 
23:      continue
24:    if Proximal descent inequality (25) does hold then
25:      % continue to next iteration with larger step size
26:       $L \leftarrow \max(L/\eta, \underline{L})$ 
27:      break
28:  return  $(x^+, L)$ 

```

4. Numerical setting

4.1. Data

In this work we use data from two separate data sets. The first data set [48] is from environmental remote sensing to study vegetation in the Mediterranean woodlands in Alta Tajo natural park, northeast of Madrid, Spain [49]. It was acquired in October–November 2014 using an AISA Eagle hyperspectral sensor and a Leica RCD105 39 megapixel digital camera mounted on a Dornier 228 aircraft. The hyperspectral imagery consisting of 126 spectral bands almost equidistantly covering the wavelengths 400 nm–970 nm with a spatial resolution of $1 \text{ m} \times 1 \text{ m}$. The aerial photograph has a resolution of $0.25 \text{ m} \times 0.25 \text{ m}$ but only captures channels corresponding to the visible red, green and blue. We refer to images from this data set as `environment1`, `environment2`.

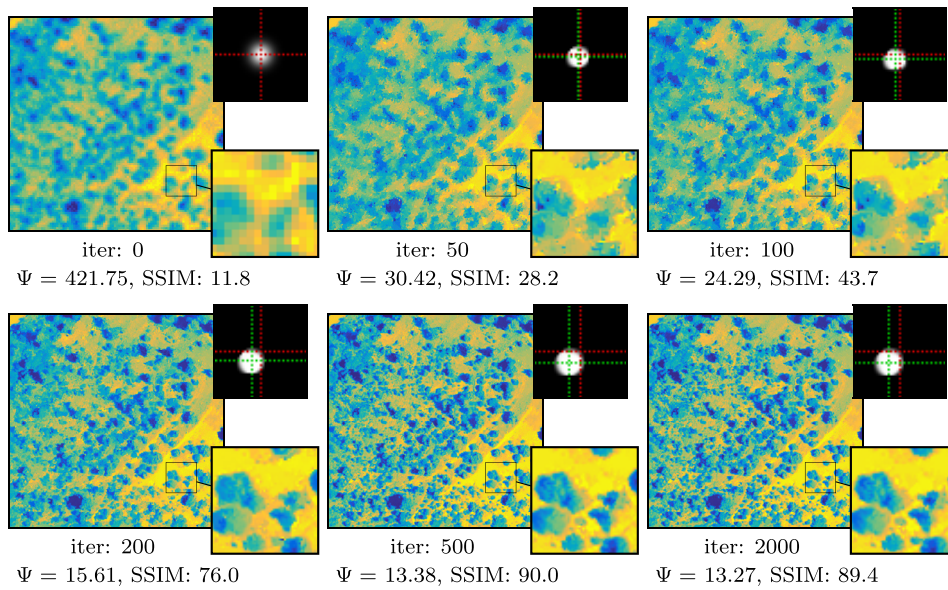


Figure 5. Iterates of PALM for data set `groundtruth_disk` with $\lambda_u = 0.1$, $\lambda_k = 10$ with objective function values Ψ and similarity to ground truth.

The second data set [50, 51] shows an urban area of Vancouver, Canada. It was acquired in March–May 2015 by the DEIMOS-2 satellite operating on a mean altitude of 620 km. The satellite carried a very highly resolving push-broom camera with one panchromatic channel of $1 \text{ m} \times 1 \text{ m}$ resolution and the four multispectral channels red, green, blue and near-infrared with a resolution of $4 \text{ m} \times 4 \text{ m}$. Images from this data set are marked `urban1`, `urban2`.

Next to the already described real data sets we simulated an environmental data set where we know the ground truth image and kernel. We test two different kernels, one which is disk-shaped and one which is an off-centered Gaussian and refer to the corresponding data sets as `groundtruth_disk` and `groundtruth_Gaussian`. The red channel of an RGB image is convolved with these two kernels, subsampled by a factor of $s = 4$ and Gaussian noise of variance 0.001 is added. The data is shown at the top left of figures 12 and 13. For the data set `groundtruth_Gaussian` we use the RGB image itself as side information and for `groundtruth_disk` we use a shifted version of it. This situation occurs frequently in applications since usually data and side information are acquired by different imaging devices which can lead to a relative translation between the observed images. Note that we used replicated instead of periodic boundary conditions for the convolution in order to avoid *inverse crime*. As the ground truth is known in these cases, we evaluate the result in terms of the *structural similarity index (SSIM)* [52] and the *Haar wavelet-based perceptual similarity index (HPSI)* [53]. As compared to measures like the peak-signal-to-noise ratio (PSNR) or the mean squared error (MSE), the SSIM and the HPSI aim at reproducing how image similarity is perceived by human viewers. The SSIM is one of the most widely used image similarity measures and obtained by comparing local statistical parameters such as local means and local variances. The HPSI is based on a simple discrete Haar wavelet transform and incorporates basic assumptions about the human visual system. The HPSI was recently shown to yield state-of-the-art correlations with human opinion scores on large benchmarking databases of differently distorted images.

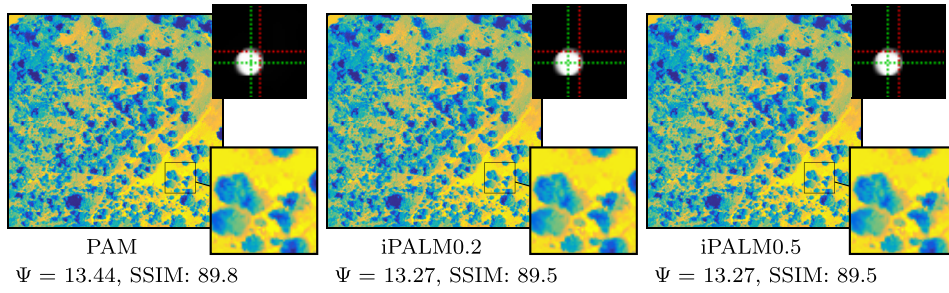


Figure 6. Reconstructions by three different algorithms—PAM and iPALM with inertial parameters $\alpha = 0.2$ and $\alpha = 0.5$ —after thousands of iterations for the same setting as figure 5. The visual impression of the images and kernels and the objective values indicate that all algorithms converge to a similar critical point.

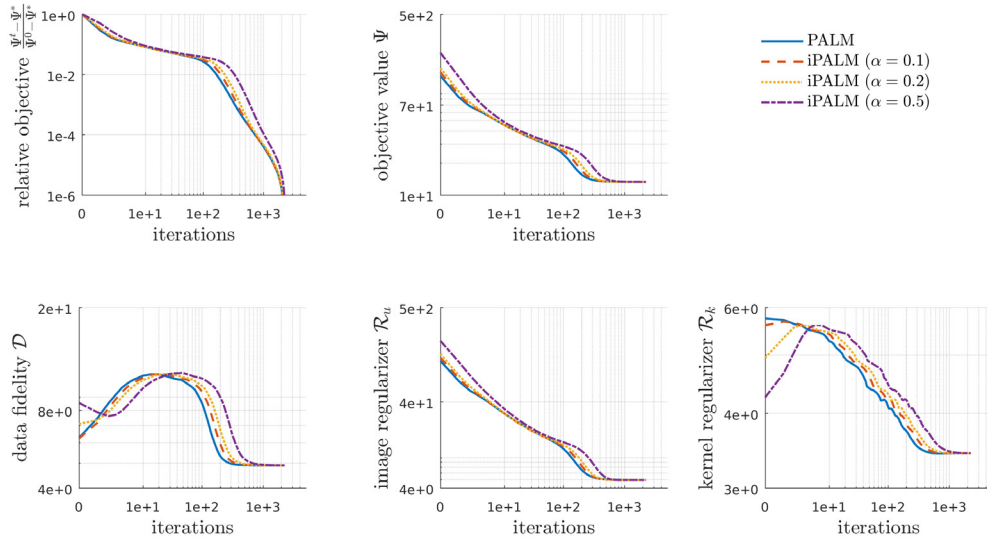


Figure 7. Algorithm statistics over the course of the iterations for the same setting as figure 5. The relative objective function values (top left) show that all algorithms have about the same speed on this example. The objective function values (top center) and its summands (bottom) indicate that also all algorithms converge to the same critical point. Overall, the plots show that more inertia leads to slower convergence.

In all numerical experiments the data f are scaled to $[0, 1]$.

4.2. Visualization

Throughout the results, the data and reconstructed images are shown with MATLAB's 'parula' colormap where yellow represents values ≥ 1 and dark blue corresponds to values ≤ 0 , respectively. The kernels are visualized with gray scales where black corresponds to the smallest value and white to the largest. In addition, we draw a red hair cross on the center, i.e. the pixel with indices $l + 1$, where l denotes the radius of the kernel. The kernel's centroid

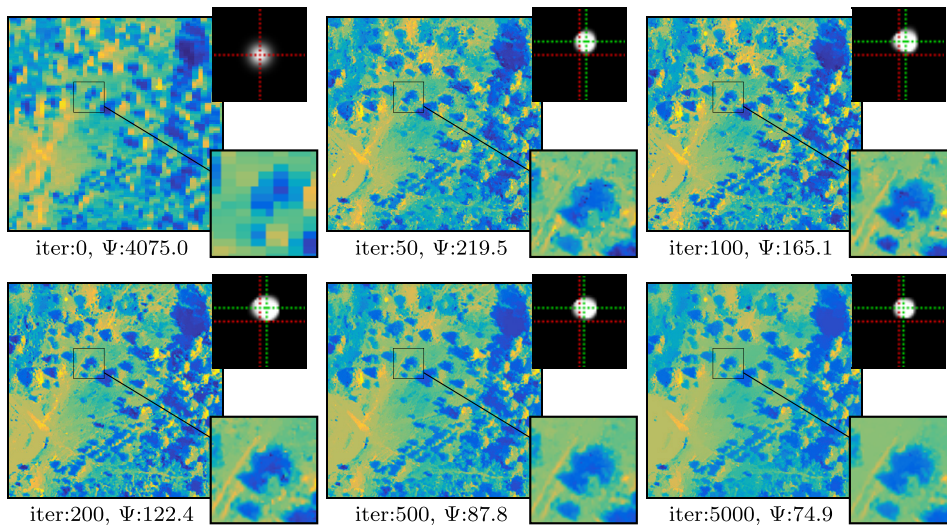


Figure 8. Iterates of PALM for the data set `environment1` with $\lambda_u = 1$ and $\lambda_k = 10$.

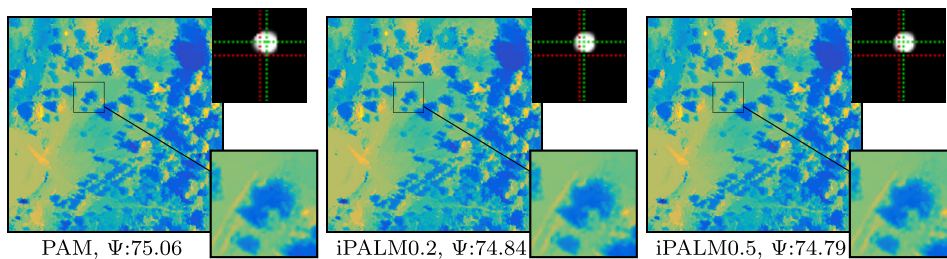


Figure 9. Reconstructions by PAM and iPALM with inertial parameters $\alpha = 0.2$ and $\alpha = 0.5$ for the setting of figure 8. Visually, the images and kernels and quantitatively, the objective function values indicate that all algorithms converge to the same critical point.

$\sum_i ik_i$ is visualized by a green hair cross in order to visualize the off-set between data and side information.

4.3. Algorithms

The numerical results will be computed with the algorithms PALM, iPALM and PAM. For PAM, the inner problems require solving a deconvolution problem with convex regularization which is implemented with ADMM [54–56] following a similar strategy as in [18]. All three—PALM, iPALM and PAM—require the computation of proximal operators for the total variation and directional total variation under a convex constraint. These are implemented with a finite number of warm started fast gradient projection/FISTA [47, 57] iterations. Within these, projections onto the constraint sets are computed by $\text{prox}_{\nu_{[0, \infty)^m}}(u) = \max(0, u)$ (to be understood componentwise) and the fast algorithm of [58] for the projection onto the unit simplex. A MATLAB implementation of the algorithms and the data itself can be found at [59].

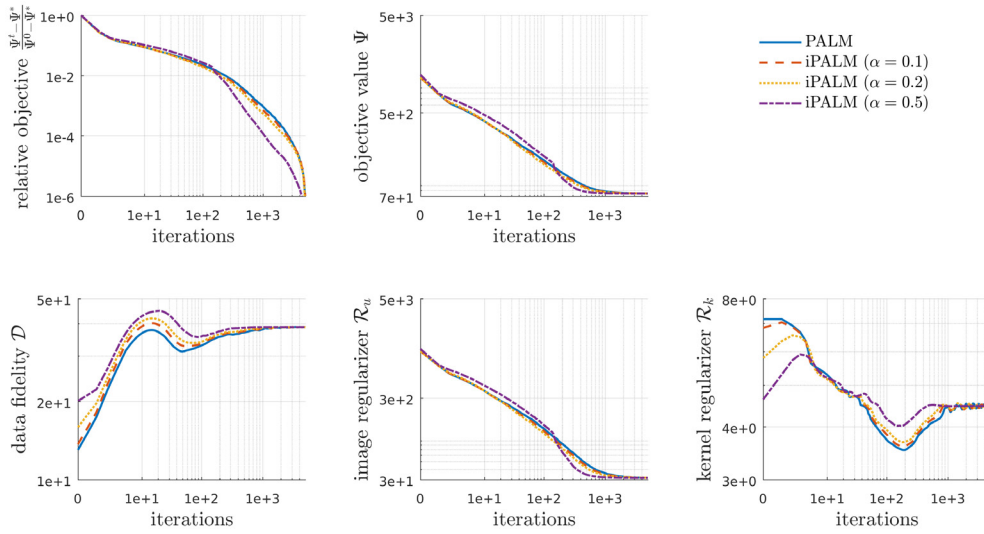


Figure 10. Algorithm statistics for PALM and iPALM with a variety of inertial parameters for the same setting as figure 8. The objective function values (top center) and its summands (bottom) indicate that all algorithms converge to the same critical point. The relative objective function values (top left) show that for this example a higher inertial parameter leads to faster convergence.

4.4. Parameters

The numerical experiments require model parameters and algorithm parameters. For the mathematical model, the parameters of the directional total variation are chosen as $\gamma = 0.9995$ and $\varepsilon = 0.003$ for all experiments, see (12) and figure A1 in the appendix. The low resolution images are of size 100×100 with an subsampling rate of $s = 4$. Since we fix the kernel size to be 41×41 , we will consequently obtain reconstructions of size 440×440 where we only visualize the meaningful part of size 400×400 .

Except for the algorithm comparison, we use PALM with 2000 iterations and a step size constant $\theta = 1.1$, see (22). The backtracking parameters in algorithm 1 are chosen as $\eta = 2$, $\underline{L} = 1$ and $\bar{L} = 10^{30}$.

4.5. Initialization

Both algorithm 1 and our implementation of PAM require initial guesses for image and kernel. While the kernel is initialized with a compactly supported Gaussian kernel, we choose the initial image to be $\mathbf{H}f$, where $\mathbf{H} : \mathbb{R}^n \rightarrow U$ is a right inverse of the operator $\mathbf{S} \circ \mathbf{B}$, i.e. \mathbf{H} is an upsampling operator (see the top-left images in figures 5 and 8).

5. Numerical results

5.1. Comparison of algorithms

Before we focus on the actual image fusion reconstructions, we compare the algorithms PALM, iPALM and PAM on two different data sets. The results are shown in figures 5, 6 and 7 for the data set `groundtruth_disk` and in figures 8, 9 and 10 for `environment1`.

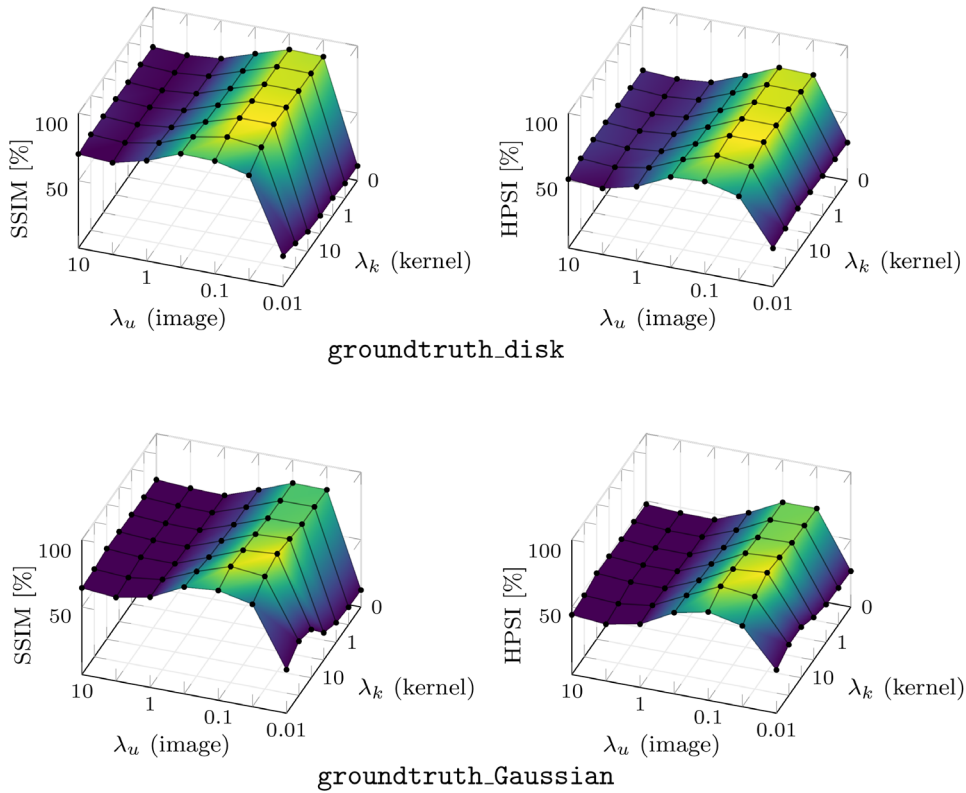


Figure 11. Similarity measures SSIM and HPSI for different values of the regularization parameters λ_u and λ_k . Top: data set `groundtruth_disk` with shifted side information. Bottom: data set `groundtruth_Gaussian`. Yellow corresponds to high and blue to low similarity. All four plots indicate that the reconstructions are not very sensitive with respect to the kernel regularization.

Please refer to the captions of each figure for some detailed observations. We would like to highlight two important observations. First, both examples show that all three algorithms converge to practically the same critical point and that a visually pleasing reconstruction is already obtained after about 500 iterations. Thus, even though problem (1) is non-convex, the algorithm itself does not seem to have much influence on the final reconstructions. Second, the examples show that inertia may slow the convergence down as for `groundtruth_disk` or may make it faster as for `environment1`. Other results—which are not shown here for brevity—indicate that this is correlated with the regularization parameter λ_u . For large values, e.g. $\lambda_u \geq 1$, we found that inertia consistently sped up convergence while for smaller values, e.g. $\lambda_u \leq 0.1$ we found that it was always slowing down the convergence. Thus, for simplicity and due to the lack of supporting examples that would justify other decisions, all further reconstructions are performed by PALM.

5.2. Simulated data

This section is dedicated to simulated data where the ground truth image is known. Figure 11 shows the similarity indices SSIM and HPSI of the reconstructed images when varying the image and kernel regularization parameters λ_u and λ_k . The results show that the proposed

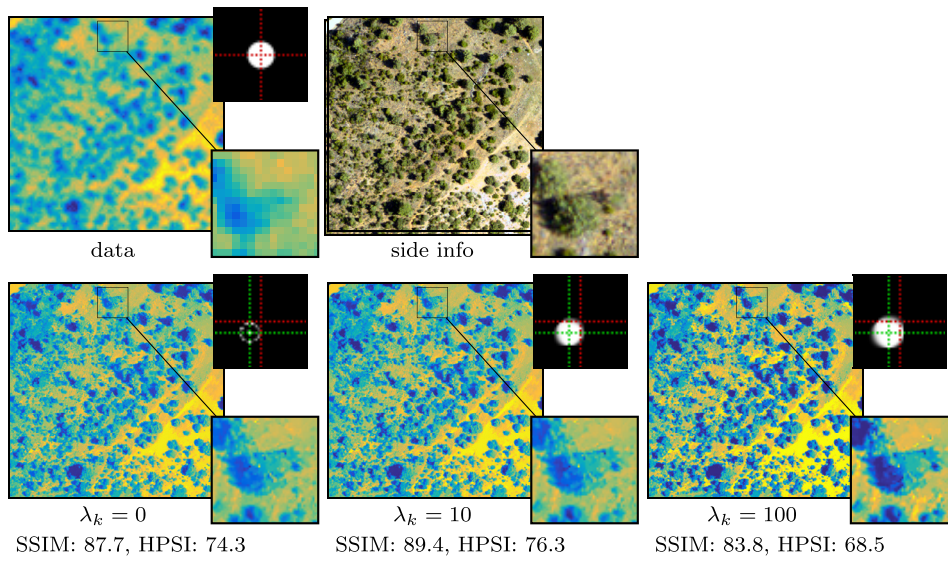


Figure 12. Reconstructions for `groundtruth_disk` with shifted side information. Top: low resolution data and side information. Bottom: reconstruction with varying image kernel regularization and image regularization chosen as $\lambda_u = 0.1$. SSIM and HPSI are given in percentages and higher values indicating better performance.

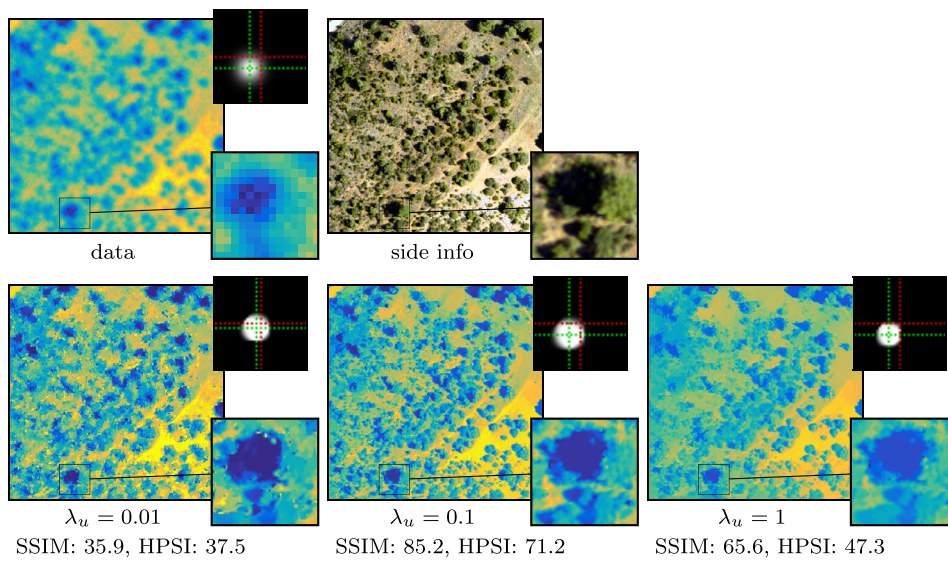


Figure 13. Reconstructions for `groundtruth_Gaussian`. Top: low resolution data and side information. Bottom: reconstruction for varying image regularization and kernel regularization being chosen as $\lambda_k = 10$. SSIM and HPSI are given in percentages and higher values indicating better performance.

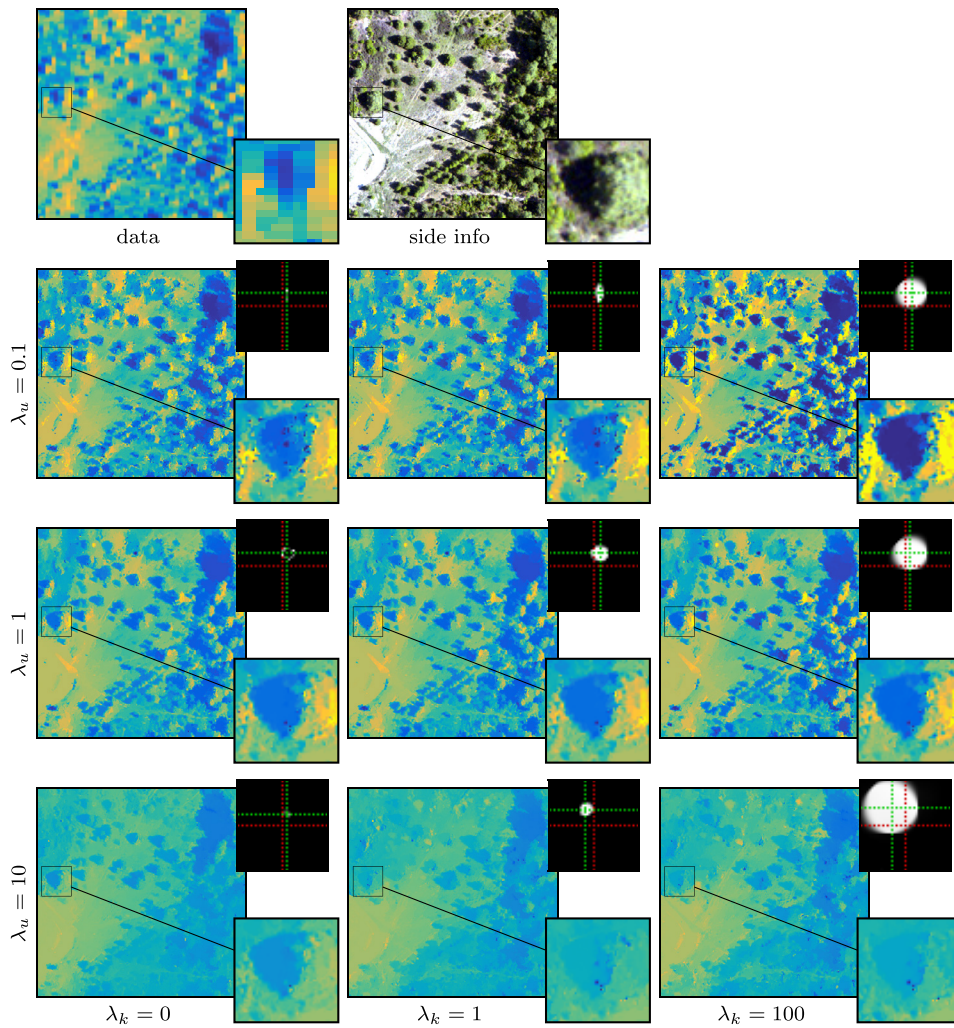


Figure 14. Varying both regularization parameters for `environment1` which corresponds to a near-infrared channel with wavelength of around 900 nm. Too small image regularization λ_u leads to point artefacts and too large λ_u to a loss of contrast and wrong estimation of the kernel. While no kernel regularization does not seem to have negative effects on the reconstruction, too large kernel regularization changes the image contrast.

model is stable with respect to the choice of parameters as long as the image regularization λ_u is not too small. It is remarkable that the success of the model depends only mildly on the choice of λ_k and a vanishing kernel regularization $\lambda_k = 0$ has only modest impact. However, it can also be seen that best reconstructions are obtained for non-zero λ_k which indicates that a mild kernel regularization should be performed.

In figure 12 we show reconstruction results for the `groundtruth_disk` test case where we fix the image regularizing parameter $\lambda_u = 0.1$ and vary the kernel regularization λ_k . Note that we complicated the situation by using a side information which is shifted by several pixels and, thus, *not aligned* to the data. The proposed approach is well-fit to deal with this

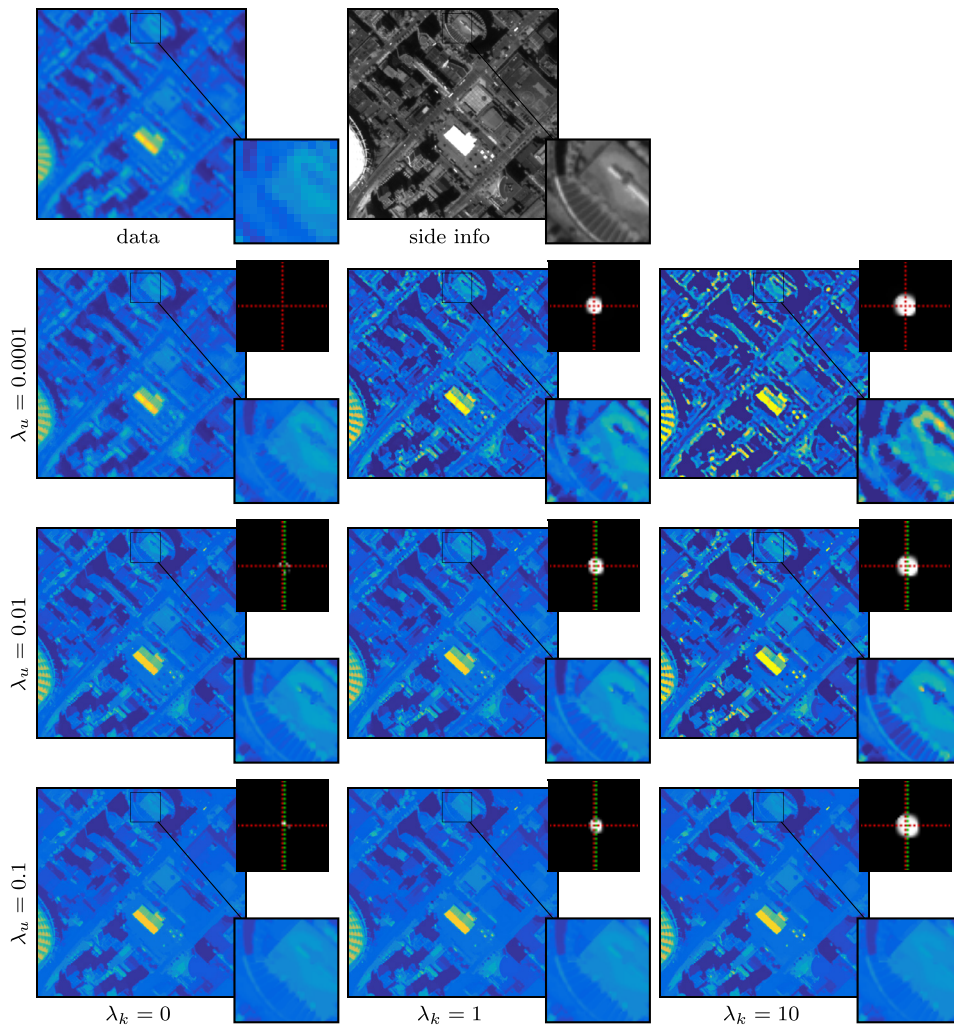


Figure 15. Results for urban1 (near-infrared) for varying image and kernel regularization. Over- and under-regularization have similar effects like in figure 14. However, due to a lower noise level of the data, the parameters can be chosen smaller.

problem since blind deconvolution can implicitly perform rigid registration of data and side information by producing a blurring kernel that compensates for the translation by having an off-centered centroid. The results show that our method is able to compensate for unaligned side information images and can achieve very good results, nevertheless. The influence of the amount of kernel regularization is noticeably small although over-regularization clearly leads to different contrasts in the reconstructed image. Note that the barycentric translations of the kernels in figure 12 correspond exactly—up to pixel accuracy—to the number of pixels that the shifted side information deviates from the un-shifted one.

Figure 13 shows in a similar way the reconstruction results for the test case `groundtruth_Gaussian`, using the un-shifted side information. In contrast to the previous example, here we fix the parameter $\lambda_k = 10$ and vary λ_u instead. Here, the effects of different regularization strengths are considerably larger; particularly, in the case $\lambda_u = 0.01$

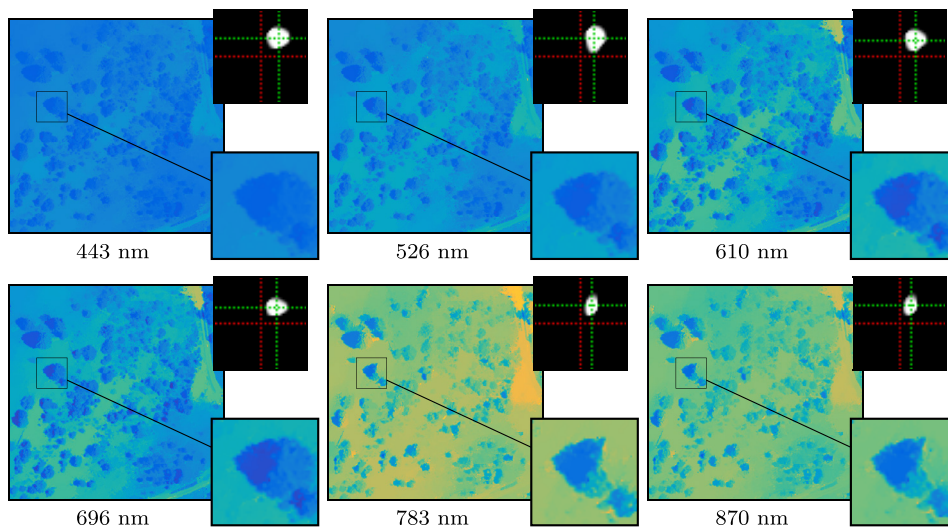


Figure 16. Reconstructions of several channels corresponding to different wavelengths given in nanometer (nm) of `environment2`. The regularization parameters for this experiment were chosen constant across the channels as $\lambda_u = 1$ and $\lambda_k = 1$. For visualization, reconstructed images are scaled back to their physical range. Note that despite the reconstruction being performed channel-by-channel, all kernels have about the same shape and the same shift.

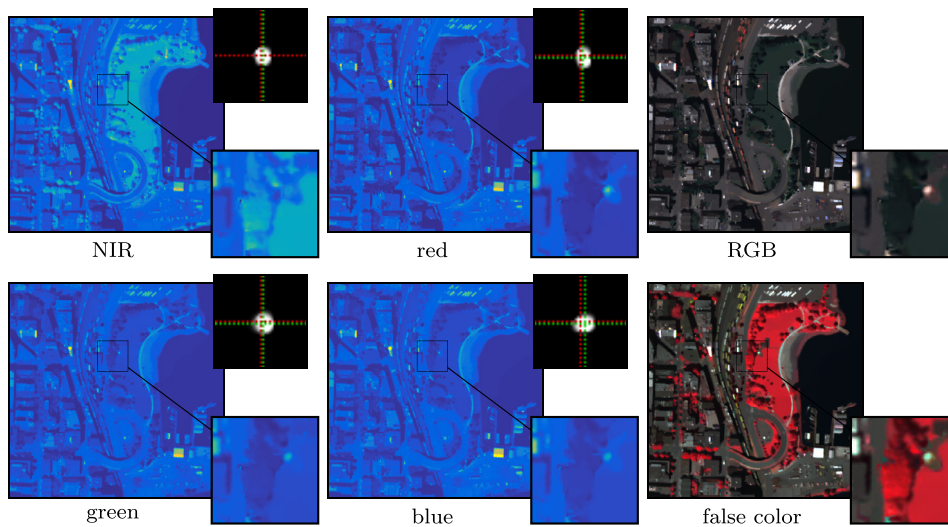


Figure 17. Reconstruction of the multispectral data set `urban2`. The regularization parameters were chosen the same for all channels as $\lambda_u = 0.01$ and $\lambda_k = 1$. Kernels are being estimated consistently over the range of the channels up to an accuracy of one pixel. The resulting RGB and false color images have well-defined boundaries and show no color smearing. (NIR = near-infrared)

the shift between data and side information—which was induced by convolving with the off-centered kernel—is not fully detected. Consequently, the resulting similarity values, being not translationally invariant, are relatively poor. In the opposite scenario of over-regularizing the image, small structures in the image are not captured well.

5.3. Hyperspectral data

In this section we apply our method to real data and—similarly to the previous section—investigate the effects of varying image and kernel regularization. To this end, we apply our method to one channel of the hyperspectral environmental and of the multispectral urban data set. Figures 14 and 15 show data, side information, and the corresponding reconstruction results for different values of the regularization parameters. Since no ground truth solutions are known in this case, we rely on our visual impression to choose a set of parameters. Figure 14, which corresponds to near-infrared light, suggests that a good compromise among data fidelity and regularization is achieved for $\lambda_u = \lambda_k = 1$. However, even if no TV regularity of the kernel is required, that is in the case $\lambda_k = 0$, the resulting images do hardly differ. Bearing in mind the results of figure 12, we will regularize the kernel slightly in the sequel. The situation is similar in figure 15, which also corresponds to near-infrared light. However, a smaller image regularization parameter, e.g. $\lambda_u = 0.01$, seems necessary in order to avoid over-smoothing.

Finally, we apply the proposed strategy to several hyperspectral channels of the data set `environment2`, using the same parameters for the reconstruction of each channel. Here, we scaled the reconstructed images back to their physical range to allow inter-channel comparisons. Figure 16 shows the reconstruction of six hyperspectral channels ranging from blue to near-infrared light. Note that the resulting kernels only change slightly with respect to the channel. We would like to point out that we were able to use the same set of parameters, namely $\lambda_u = \lambda_k = 1$, for all channels. In figure 17 we show the reconstruction of all channels of the multispectral urban data set `urban2` using $\lambda_u = 0.01$ and $\lambda_k = 1$. Additionally, we visualized the results by re-combining the four channels into an RGB and a false color image which has the channels near-infrared, red and green. We would like to highlight that the kernels are consistently estimated up to a pixel accuracy and that no color smearing is visible in the RGB and false color images despite the fact that all channels have been processed independently.

6. Conclusions and outlook

We have presented a novel model for simultaneous image fusion and blind deblurring of hyperspectral images based on the directional total variation. We showed that this approach yields sharp images for both environmental and urban data sets that are either acquired by a plane or by a satellite. In addition, the numerical comparison of different algorithms showed that—despite the non-convexity of the problem—the computed solutions are stable with respect to the algorithms. The implicitly included registration by estimating the kernel yields further robustness to imperfections in the image acquisitions. Thus, the proposed approach is appealing for automated processing of large data sets.

Future work will be devoted to multiple directions. First, the proposed approach is channel-by-channel, which makes it computationally appealing. However, a coupling of the spectral channels via their kernel and/or via spectral fingerprint preservation may yield further improvements. Second, higher order regularization such as the total generalized variation [40] has been proven useful for natural images. This naturally leads to the directional total generalized variation which combines higher order smoothness with structural *a priori* knowledge as presented in

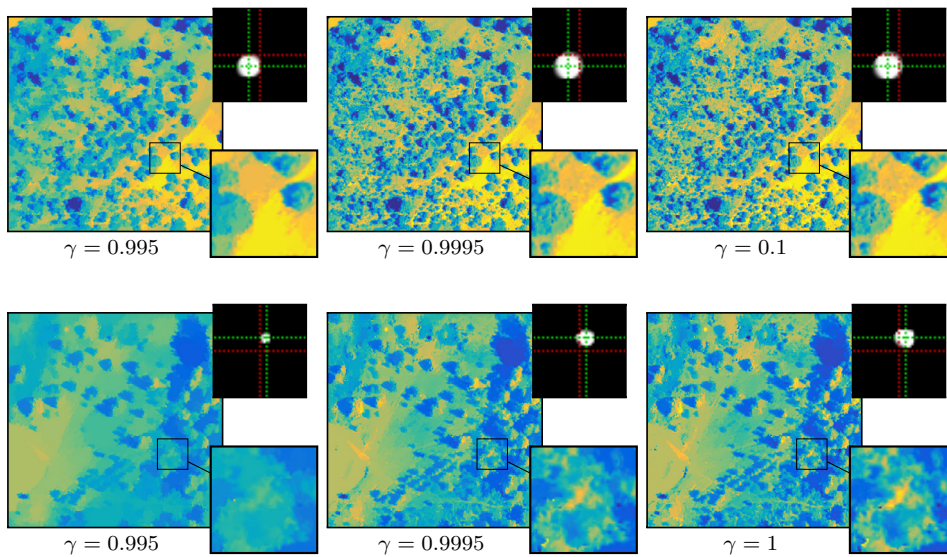


Figure A1. Reconstructions for varying vector field parameters γ . Top: results for `groundtruth_disk` with $\lambda_u = 0.1$ and $\lambda_k = 10$. Bottom: results for `environment1` with $\lambda_u = 1$, $\lambda_k = 1$.

this work, thereby generalizing the definition of [36]. Similarly, other kernel regularization functionals such as the total generalized variation or the H^1 -semi-norm will be considered in future investigations to better model the smoothness of the kernel. Finally, it is currently not clear how the computational cost of the proposed method depends on the size of the considered data. In particular, it is difficult to predict to which degree larger images also require a higher number of iterations during optimization for the method to yield satisfactory results. However, due to the fact that our approach mostly depends on operations that are of a local nature, it should be highly suitable for parallelization. The development and analysis of implementations that utilize the benefits of distributed computing architectures is also a topic for future research.

Acknowledgments

We would like to express our gratitude to Juheon Lee and Qi Wei for their help with the environmental and urban data sets, respectively. Furthermore, the authors would like to thank Deimos Imaging for acquiring and providing the data used in this study, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee.

MJE and C-BS acknowledge support from Leverhulme Trust project ‘Breaking the non-convexity barrier’, EPSRC grant ‘EP/M00483X/1’, EPSRC centre ‘EP/N014588/1’, the Cantab Capital Institute for the Mathematics of Information, and from CHiPS (Horizon 2020 RISE project grant). Moreover, C-BS is thankful for support from the Alan Turing Institute.

DAC acknowledges the support of NERC (grant number NE/K016377/1). We are grateful to NERC’s Airborne Research Facility and Data Analysis Node for conducting the airborne survey and pre-processing the environmental data collected from Alto Tajo, Spain (survey CAM11/03).

Appendix


We test the influence of parameter γ in (12) on a simulated and a real data set. Figure A1 shows the corresponding reconstruction results for $\gamma \in \{0.995, 0.9995, 1\}$ where the middle value is used in all other numerical experiments in this work. Note that the results for $\gamma = 1$ suffer from stronger point artifacts whereas those for $\gamma = 0.995$ do not capture local structures well enough.

ORCID iDs

Leon Bungert  <https://orcid.org/0000-0002-6554-9892>

David A Coomes  <https://orcid.org/0000-0002-8261-2582>

Matthias J Ehrhardt  <https://orcid.org/0000-0001-8523-353X>

Rafael Reisenhofer  <https://orcid.org/0000-0003-1652-2412>

Carola-Bibiane Schönlieb  <https://orcid.org/0000-0003-0099-6306>

References

- [1] Ustin S L, Roberts D A, Gamon J A, Asner G P and Green R O 2004 *AIBS Bull.* **54** 523–34
- [2] Homolova L, Malenovský Z, Clevers J G, García-Santos G and Schaepman M E 2013 *Ecol. Complexity* **15** 1–16
- [3] Amro I, Mateos J, Vega M, Molina R and Katsaggelos A K 2011 *EURASIP J. Adv. Signal Process.* **2011** 1–22
- [4] Loncan L et al 2015 *IEEE Geosci. Remote Sens. Mag.* **3** 27–46
- [5] Mowle E W and Dennehy C J 1991 The LANDSAT-6 satellite: an overview *National Telesystems Conf. Proc.* vol 1
- [6] Ballester C, Caselles V, Igual L, Verdera J and Rougé B 2006 *Int. J. Comput. Vision* **69** 43–58
- [7] Möller M, Wittman T and Bertozzi A L 2009 A variational approach to hyperspectral image fusion *Proc. SPIE* **7** 7334
- [8] Padwick C, Deskevich M, Pacifici F and Smallwood S 2010 WorldView-2 pansharpening *ASPRS* vol **48** pp 26–30
- [9] Möller M, Wittman T, Bertozzi A L and Burger M 2012 *SIAM J. Imaging Sci.* **5** 150–78
- [10] Simoes M, Bioucas-Dias J, Almeida L B and Chanussot J 2015 *IEEE Trans. Geosci. Remote Sens.* **53** 3373–88
- [11] Zhang Q, Cao Z, Hu Z, Jia Y and Wu X 2015 *IEEE Geosci. Remote Sens. Lett.* **12** 467–71
- [12] Hou L and Zhang X 2016 *J. Math. Imaging Vis.* **55** 36–49
- [13] Duran J, Buades A, Coll B, Sbert C and Blanchet G 2017 *ISPRS J. Photogramm. Remote Sens.* **125** 78–105
- [14] Kaipio J P, Kolehmainen V, Vauhkonen M and Somersalo E 1999 *Inverse Problems* **15** 713–29
- [15] Bowsher J E et al 2004 Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors *Proc. IEEE NSS-MIC* pp 2488–92
- [16] Vunckx K, Atre A, Baete K, Reilhac A, Deroose C M, Van Laere K and Nuyts J 2012 *IEEE Trans. Med. Imaging* **31** 599–612
- [17] Ehrhardt M J et al 2016 *IEEE Trans. Med. Imaging* **35** 2189–99
- [18] Ehrhardt M J and Betcke M M 2016 *SIAM J. Imaging Sci.* **9** 1084–106
- [19] Kolehmainen V, Ehrhardt M J and Arridge S 2017 *Inverse Problems* submitted
- [20] Mehranian A, Belzunce M A, Niccolini F, Politis M, Prieto C, Turkheimer F, Hammers A and Reader A J 2017 *Phys. Med. Biol.* **62** 5975
- [21] Lee J, Cai X, Schönlieb C B and Coomes D A 2015 *IEEE Trans. Geosci. Remote Sens.* **53** 6073–84
- [22] Amizic B, Spinoulas L, Molina R and Katsaggelos A K 2013 *IEEE Trans. Image Process.* **22** 3994–4006
- [23] Perrone D and Favaro P 2016 *Int. J. Comput. Vis.* **117** 159–72
- [24] Kundur D and Hatzinakos D 1996 *IEEE Signal Process. Mag.* **13** 43–64

- [25] Vorontsov S V and Jefferies S M 2017 *Inverse Problems* **33** 055004
- [26] Chan T F and Shen J 2005 *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods* (Philadelphia, PA: SIAM)
- [27] Almeida M S and Figueiredo M 2013 *IEEE Trans. Image Process.* **22** 3074–86
- [28] Rudin L I, Osher S and Fatemi E 1992 *Physica D* **60** 259–68
- [29] Grasmair M and Lenzen F 2010 *Appl. Math. Opt.* **62** 323–39
- [30] Estellers V, Thiran J and Bresson X 2013 *IEEE Trans. Image Process.* **22** 2611–26
- [31] Estellers V, Soatto S and Bresson X 2015 *IEEE Trans. Image Process.* **24** 1777–90
- [32] Ehrhardt M J and Arridge S R 2014 *IEEE Trans. Image Process.* **23** 9–18
- [33] Ehrhardt M J, Thielemans K, Pizarro L, Atkinson D, Ourselin S, Hutton B F and Arridge S R 2015 *Inverse Problems* **31** 015001
- [34] Ehrhardt M J 2015 Joint reconstruction for multi-modality imaging with common structure *PhD Thesis* University College London
- [35] Bayram I and Kamasak M E 2012 *IEEE Signal Process. Lett.* **19** 265–9
- [36] Kongskov R D, Dong Y and Knudsen K 2017 Directional total generalized variation regularization (arXiv:1701.02675)
- [37] Arridge S R, Kolehmainen V and Schweiger M J 2008 Reconstruction and regularisation in optical tomography *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy* ed Y Censor et al (Pisa: Scuola Normale Superiore)
- [38] Grasmair M 2009 Locally adaptive total variation regularization *Proc. SSVM* pp 331–42
- [39] Knoll F, Dong Y, Langskammer C, Hintermüller M and Stollberger R 2010 Total variation denoising with spatially dependent regularization *Proc. of ISMRM* vol 18 p 5088
- [40] Bredies K, Kunisch K and Pock T 2010 *SIAM J. Imaging Sci.* **3** 492–526
- [41] Pock T and Sabach S 2016 *SIAM J. Imaging Sci.* **9** 1756–87
- [42] Attouch H, Bolte J, Redont P and Soubeyran A 2010 *Math. Oper. Res.* **35** 438–57
- [43] Bolte J, Sabach S and Teboulle M 2014 *Math. Program.* **146** 459–94
- [44] Lions P L and Mercier B 1979 *SIAM J. Numer. Anal.* **16** 964–79
- [45] Combettes P L and Wajs V R 2005 *Multiscale Model. Simul.* **4** 1168–200
- [46] Ochs P, Chen Y, Brox T and Pock T 2014 *SIAM J. Imaging Sci.* **7** 1388–419
- [47] Beck A and Teboulle M 2009 *IEEE Trans. Image Process.* **18** 2419–34
- [48] Arnold R, Johnson J, Joseph C and Llewellyn G 2014 NERC airborne research facility <http://arsf.nerc.ac.uk>
- [49] Simonson W, Ruiz-Benito P, Valladares F and Coomes D 2016 *Biogeosciences* **13** 961–73
- [50] 2016 IEEE GRSS data fusion contest www.grss-ieee.org/community/technical-committees/data-fusion
- [51] Mou L, Zhu X, Vakalopoulou M, Karantzas K, Paragios N, Le Saux B, Moser G and Tuia D 2017 *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10** 3435–47
- [52] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 *IEEE Trans. Image Process.* **13** 600–12
- [53] Reichenhofer R, Bosse S, Kutyniok G and Wiegand T 2018 *Signal Process., Image Commun.* **61** 33–43
- [54] Gabay D and Mercier B 1976 *Comput. Math. Appl.* **2** 17–40
- [55] Boyd S P, Parikh N, Chu E, Peleato B and Eckstein J 2010 *Found. Trends Mach. Learn.* **3** 1–122
- [56] Afonso M V, Bioucas-Dias J M and Figueiredo M A T 2010 *IEEE Trans. Image Process.* **19** 2345–56
- [57] Beck A and Teboulle M 2009 *SIAM J. Imaging Sci.* **2** 183–202
- [58] Duchi J, Shalev-Shwartz S, Singer Y and Chandra T 2008 Efficient projections onto the l_1 -ball for learning in high dimensions *Proc. 25th ICML (ACM)* pp 272–9
- [59] Bungert L, Coomes D A, Ehrhardt M J, Rasch J, Reichenhofer R and Schönlieb C B 2017 Supplementary material for blind image fusion for hyperspectral imaging with the directional total variation (<https://doi.org/10.17863/CAM.17104>)