Joint Image Reconstruction and Segmentation Using the Potts Model

Martin Storath¹, Andreas Weinmann^{2,4}, Jürgen Frikel^{3,4}, Michael Unser¹

 $^1\mathrm{Biomedical}$ Imaging Group, École polytechnique fédérale de Lausanne, Switzerland

²Department of Mathematics, Technische Universität München, Germany ³Department of Mathematics, Tufts University, USA

⁴Research Group Fast Algorithms for Biomedical Imaging, Helmholtz Zentrum München, Germany

E-mail: martin.storath@epfl.ch, andreas.weinmann@tum.de, juergen.frikel@helmholtz-muenchen.de, michael.unser@epfl.ch

Abstract. We propose a new algorithmic approach to the non-smooth and nonconvex Potts problem (also called piecewise-constant Mumford-Shah problem) for inverse imaging problems. We derive a suitable splitting into specific subproblems that can all be solved efficiently. Our method does not require a priori knowledge on the gray levels nor on the number of segments of the reconstruction. Further, it avoids anisotropic artifacts such as geometric staircasing. We demonstrate the suitability of our method for joint image reconstruction and segmentation from limited data in x-ray and photoacoustic tomography. For instance, our method is able to reconstruct the Shepp-Logan phantom from 7 angular views only. We demonstrate the practical applicability in an experiment with real PET data.

Keywords: Potts model, piecewise-constant Mumford-Shah model, regularization of ill-posed problems, image segmentation, Radon transform, spherical Radon transform, computed tomography, photoacoustic tomography, deconvolution.

1. Introduction

In this paper, we consider ill-posed imaging problems with incomplete data. Incomplete data are often due to technical restrictions or design issues of the imaging modality, like in the case of freehand SPECT [80] or limited-angle tomography [18, 61, 76]. Also health-related considerations lead to incomplete data. For example, sparse-angle setups are used to reduce radiation doses in x-ray tomography [5]. In addition to incompleteness, the data is usually corrupted by noise and may also suffer from blur [40, 51, 71, 83]. Altogether, this makes the reconstruction problems severely ill-posed which means that small perturbations in the data potentially lead to large errors in the reconstruction [30, 56, 57, 61]. It may even happen that certain features (singularities) are invisible from the incomplete data [38, 46, 68]. Consequently, the quality of reconstruction decreases significantly and one tends to lose fine details. Nevertheless, one can still try to reconstruct the object at a coarser scale. This is often of particular interest in medical imaging; for example, the locations of inner organs might be needed for surgery planning [69].

Classical reconstruction methods perform poorly in limited-data situations. Better approaches incorporate specific prior assumptions on the reconstruction. They are typically stated in terms of the minimization of some cost function. A popular representative is the convex total variation [25, 72]. A recent trend is to use non-convex regularizers [9, 27, 64, 69]. Although analytically and computationally more demanding, they give more freedom in the modeling and often yield better reconstructions [27, 28]. In their seminal work [60], Mumford and Shah introduced a cost functional on the piecewise-constant functions – now called the piecewise-constant Mumford-Shah functional – where the length of the discontinuity set is penalized. This functional has shown good performance, especially for the recovery of geometric macrostructures from poor data [50, 69, 70]. The piecewise-constant Mumford-Shah model also appears in statistics and image processing where it is often called *Potts model* [12, 13, 14, 66, 81]. The variational formulation of the Potts model is given by

$$\arg\min_{u} \gamma \|\nabla u\|_{0} + \|Au - f\|_{2}^{2}.$$
 (1)

Here, A is a linear operator (e.g., the Radon transform) and f is an element of the data space (e.g., a sinogram). A mathematically precise definition of the jump term $\|\nabla u\|_0$ is rather technical in a spatially continuous setting. However, if u is piecewise-constant and the discontinuity set of u is sufficiently regular, say, a union of C^1 curves, then $\|\nabla u\|_0$ is just the total arc length of this union. In general, the gradient ∇u is given in the distributional sense and the boundary length is expressed in terms of the (d-1)-dimensional Hausdorff measure. When u is not piecewise-constant, the jump penalty is infinite [70]. The second term measures the fidelity of a solution u to the data f. The parameter $\gamma > 0$ controls the balance between data fidelity and jump penalty. The Potts model can be interpreted in two ways. On the one hand, if the imaged object is (approximately) piecewise-constant, then the solution is an (approximate) reconstruction of the imaged object. On the other hand, since a piecewise-constant solution directly induces a partitioning of the image domain, it can be seen as joint reconstruction and segmentation. Executing reconstruction and segmentation jointly typically leads to better results than performing the two steps successively [50, 69].

The Potts problem is NP-hard [12]. Thus, there is no hope to find a global minimizer in reasonable time. Nevertheless, due to its importance in image reconstruction and segmentation, several approximative strategies have been proposed.

Kim et al. [47] consider an Ambrosio-Tortorelli-type approximation. Bar et al. [4] use a level-set based active contour method for deconvolution. Ramlau and Ring [69] employ a related level-set approach for the joint reconstruction and segmentation of x-ray tomographic images; further applications are electron tomography [48] and SPECT [50]. The latter authors were the first to investigate the regularizing properties of such functionals [69, 70]. We elaborate further on Potts regularization in inverse problems and on existing algorithmic approaches in the sections 1.1 and 1.2, respectively.

In this paper, we first discretize the Potts problem as

$$u^* = \underset{u \in \mathbb{R}^{m \times n}}{\arg\min} \gamma \sum_{s=1}^{S} \omega_s \|\nabla_{p_s} u\|_0 + \|Au - f\|_2^2.$$
(2)

The symbol ∇_p denotes finite differences with respect to the displacement vector p so that $\nabla_p u = u(\cdot + p) - u$, where $p \in \mathbb{Z}^2 \setminus \{0\}$. The symbol $\|\nabla_{p_s} u\|_0$ denotes the number of nonzero entries of $\nabla_{p_s} u$. The displacement vectors p belong to a neighborhood system $\mathcal{N} = \{p_1, ..., p_S\}$ and $\omega_1, ..., \omega_S$ are nonnegative weights. The simplest neighborhood system is made up of the two canonical basis vectors of \mathbb{R}^2 along with unit weights. Unfortunately, when refining the grid, this discretization converges to a limit that measures the boundary in terms of the ℓ^1 analogue of the Hausdorff measure [22]. The practical consequences are unwanted block artifacts in the reconstruction (geometric staircasing). The addition of diagonal or "knight-move" finite differences (referring to the moves of a knight in chess) mildens such anisotropy effects [23]. We provide a general scheme for the proper choice of finite-difference systems and accompanying weights which allows for arbitrarily good approximations of the Euclidean length.

Based on (2), we propose a new minimization strategy for the Potts problem. Our key contribution is a particularly suitable splitting of the Potts problem (2) into specific subproblems. The first subproblem is a classical Tikhonov-regularized problem with a solution that reduces to a linear system of equations. All the remaining subproblems can be solved efficiently by dynamic programming [22, 37, 60, 73]. We prove that our algorithm converges. A major advantage of our method is that neither the number of segments nor the gray-values of a solution have to be fixed a priori. Further, the method does not require any initial guess of the solution. Last but not least, it is highly parallelizable and easy to implement.

We demonstrate the suitability of our method for joint image reconstruction and segmentation in several setups. We reconstruct from Radon data, which appear in x-ray tomography (CT) and in positron emission tomography (PET), and from spherical Radon data, which appear in photoacoustic tomography (PAT). In noise-free conditions, we achieve almost perfect reconstruction of the Shepp-Logan phantom from only seven projections (see Figure 1). In the presence of noise, our method provides high-quality reconstructions of piecewise-constant images from a small number of projections. On PET data of a physical phantom, we obtain a reliable segmentation of the anatomic structures. Eventually, we apply our method to the deconvolution of vector-valued data.

1.1. Potts Regularization in Inverse Problems

In many imaging problems, the imaging operator A is not boundedly invertible. Examples are the Radon transform and the spherical Radon transform when viewed as operators on the corresponding L^2 spaces. Because of unboundedness, a direct



(a) Original (256×256)

(b) Rec. from 7 projections of Radon data (PSNR: 45.2, MSSIM: 0.996)

(c) Rec. from 7 projections of spherical Radon data (PSNR: 51.5, MSSIM: 0.999)

Figure 1 The proposed method yields high quality reconstructions of piecewise-constant images from highly undersampled Radon as well as spherical Radon data.

inversion (if possible at all) may amplify small perturbations in the data. In order to attenuate such effects and produce more stable reconstructions, regularization is needed. A popular approach for regularization is by means of minimizing an energy functional of the form $\gamma J(u) + ||Au - f||_2^2$. Here, the second term is the data-fidelity term while the first term – called the regularizing term – is a penalty that incorporates a priori knowledge on the solution. Classical regularizations are Besov or Sobolev seminorms which associate certain smoothness classes with the solution. Less classical choices are the TV seminorm which leads to solutions of bounded variation. In the context of sparsity regularization, non-convex functionals are also used as regularizing terms [16]. The Potts functional (1) fits into this framework by letting J be the jump penalty. It is non-convex and may be seen as a term that enforces a sparse gradient. We note that Mumford-Shah approaches (which include Potts functionals) also regularize the boundaries of the discontinuity set of the underlying signal [44].

The existence of minimizers of (1) is not guaranteed in a continuous domain setting [36, 35, 70, 74]. For example, if A is a compact convolution operator originating from a smooth function and data given by the impulse response $f = A\delta$, then the associated Potts functional does not have a minimizer [74]. In order to ensure the existence of minimizers, additional penalty terms such as an L^p (1 term of $the form <math>||u||_p^p$ [69, 70] or pointwise boundedness constraints [44] have been considered. We note that the existence of minimizers is guaranteed in the discrete domain setup [36, 74].

It is important to verify that the Potts model is a regularization method in the sense of inverse problems. The first work dealing with this task is [70]. Ramlau and Ring [70] assume that the solution space consists of non-degenerate piecewise-constant functions with at most k (arbitrary, but fixed) different values which are additionally bounded. Under relatively mild assumptions on the operator A, they show stability. In addition, they give a parameter choice rule and show that the solutions corresponding to the chosen parameters (which depend on the level of noise) converge to a noise-free solution as the noise vanishes. This means that the method is a regularizer in the sense of inverse problems. Related references are [49, 50] and the recent publication [44] which includes (non-piecewise-constant) Mumford-Shah functionals.

1.2. Existing Algorithmic Approaches to Potts and Related Problems

The name "Potts model" for (1) has been retained in honor of R.B. Potts [67] who, as early as 1952, considered related jump penalties in his work in the field of statistical mechanics. The classical Potts model (A = id) was first introduced in image processing by Geman and Geman [39] in a statistical framework. Their algorithmic approach is based on simulated annealing. From a variational-calculus point of view, the problem was first considered by Mumford and Shah [60]. Ambrosio and Tortorelli [1] proposed an approximation by elliptic functionals. Currently, popular algorithmic approaches for the classical case A = id are based on active contours [26, 29], graph cuts [12], convex relaxations [66, 24], and semi-global matching [43].

The general case of A being a linear operator has been investigated to a lesser extent. Yet, strategies based on active contours emerge as an important class. The idea is to parameterize the jump set of u by a set of contours which evolve according to a deformation force. Active contours are used for A being a convolution operator in [4] and for A being the Radon transform in [69]. Both use level sets to parameterize the active contours. A similar method has been applied to joint segmentation of SPECT/CT data [49, 50]. Level-set methods have also been applied to stabilize sparseangle tomography [52]. In [53], the authors use explicitly parameterized contours for the application to bioluminescence tomography. In general, active contours are quite flexible as the deformation force can be easily adjusted. Their main disadvantages are that they require a good guess on the initial contour and a good guess on the expected number of gray values of the solution.

Graph cuts [12] are a popular strategy to address the classical Potts problem with A = id. In [74], the inverse problem for a general A is approached by iteratively using a graph-cut strategy on a classical Potts problem, followed by Tikhonov regularization. There, the authors apply their algorithm to deconvolution. A drawback of this approach is that graph cuts need an (a priori) discretization of the codomain of u. Thus, one needs either a good initial guess on the values that u may take, or a very fine (and expensive) discretization of the codomain of u.

To circumvent NP hardness [12], the jump penalty is often replaced by the total variation $\|\nabla u\|_1$; see [6, 15, 17, 20, 31, 84] and the references therein. TV minimization is theoretically and algorithmically easier to access since it is a convex problem. The theory of compressed sensing gives conditions when the minimizers of the relaxed problem coincide with those of the original problem [20, 62]. However, the conditions are relatively restrictive and fail to apply to some problems of practical importance. In limited-data situations, TV minimizers typically lack sharply localized boundaries [27, 28]. To sharpen the results of total variation minimization, various techniques such as iterative reweighting [21], simplex constraints [55], or iterative thresholding [19] have been proposed.

In order to come closer to the solution of the initial Potts problem, many authors propose the use of non-convex priors. Chartrand [27] uses priors based on the ℓ^p norm of the gradient, 0 , for the reconstruction of MRI-type images. Logarithmicpriors are considered by Bostan et al. [9] for the reconstruction of biomedical images.Nikolova et al. [63, 64] propose a whole class of non-convex regularizers which aretreated using a graduated non-convexity approach.

Another approach is to transform the Potts problem to an ℓ^0 problem [3, 28, 36]. The resulting problem is separable which allows for the application of iterative hard-thresholding-type algorithms [3, 7, 8, 36]. However, using this transformation comes

with constraints in form of discrete Schwarz conditions [36] as well as a data term of the form $||ABu - f||_2^2$ with a full triangular matrix B. While the initial system matrix A is typically sparse, the modified matrix AB is not so in general.

1.3. Organization of the Article

In Section 2, we present our splitting approach to the Potts problem. We start by explaining the basic approach using an anisotropic discretization of (1). Next, we discuss strategies to get more isotropic discretizations, thus attenuating the unwanted geometric staircasing effect resulting from an anisotropic discretization. Then, we present our general algorithm. We briefly discuss more general data terms and prove the convergence of our algorithm. In Section 3, we apply our method to ill-posed imaging problems. In particular, we consider Radon data as well as spherical Radon data. Furthermore, we apply our technique to real PET data. Eventually, we apply our method to deconvolution problems.

2. A Splitting Approach for the Potts Problem

In this section, we present our splitting approach for the discrete-domain Potts functional (2). It seems instructive to first describe the basic idea of the splitting in the simplest case, which is the anisotropic discretization of the length term. This discretization turns out to be anisotropic as it measures the length of the discrete boundary in the Manhattan metric, which is the metric induced by the ℓ^1 norm on \mathbb{R}^2 . This typically leads to block artifacts in the reconstruction. To avoid this we derive appropriate neighborhood systems and corresponding weights such that the discrete length term in (2) becomes more isotropic. Based on this discretization, we formulate our general splitting for the Potts problem. The problem reduces to smaller tractable subproblems that we briefly describe. We conclude the section with a convergence result.

2.1. Basic Splitting Algorithm for an Anisotropic Discretization

In the simplest case, the discretization of the regularizing term $\|\nabla u\|_0$ uses only finite differences with respect to the coordinate axes. Thus, $p_1 = (1,0)$, $p_2 = (0,1)$, and the weights ω_1, ω_2 are equal to 1. Then, the regularizing term reads

$$\|\nabla u\|_{0} = \|\nabla_{p_{1}}u\|_{0} + \|\nabla_{p_{2}}u\|_{0} = |\{(i,j): u_{ij} \neq u_{i+1,j}\}| + |\{(i,j): u_{ij} \neq u_{i,j+1}\}|.$$

Plugging this discretization into (1), we rewrite the Potts problem as the constrained optimization problem

minimize
$$\gamma(\|\nabla_{p_1}u_1\|_0 + \|\nabla_{p_2}u_2\|_0) + \|Av - f\|_2^2$$

subject to $v - u_1 = 0, \quad v - u_2 = 0, \quad u_1 - u_2 = 0.$ (3)

The augmented Lagrangian of this optimization problem reads

$$L = \gamma \left(\|\nabla_{p_1} u_1\|_0 + \|\nabla_{p_2} u_2\|_0 \right) + \|Av - f\|_2^2 + \langle \lambda_1, v - u_1 \rangle + \frac{\mu}{2} \|v - u_1\|_2^2 + \langle \lambda_2, v - u_2 \rangle + \frac{\mu}{2} \|v - u_2\|_2^2 + \langle \varrho, u_1 - u_2 \rangle + \frac{\nu}{2} \|u_1 - u_2\|_2^2.$$
(4)

The constraints are now part of the (multivariate) target functional L. The parameter $\nu > 0$ controls how strong the split variables u_1, u_2 are tied to each other and $\mu > 0$

controls their coupling to v. The variables λ_1 , λ_2 , and ρ are $(m \times n)$ -dimensional arrays of Lagrange multipliers. The inner product is defined as $\langle x, y \rangle = \sum_{i,j} x_{ij} y_{ij}$. Completing the squares in (4), we reformulate L in the convenient form

$$L = \gamma \left(\|\nabla_{p_1} u_1\|_0 + \|\nabla_{p_2} u_2\|_0 \right) + \|Av - f\|_2^2 + \frac{\mu}{2} \|v - u_1 + \frac{\lambda_1}{\mu}\|_2^2 - \frac{\mu}{2} \|\frac{\lambda_1}{\mu}\|_2^2 + \frac{\mu}{2} \|v - u_2 + \frac{\lambda_2}{\mu}\|_2^2 - \frac{\mu}{2} \|\frac{\lambda_2}{\mu}\|_2^2 + \frac{\nu}{2} \|u_1 - u_2 + \frac{\varrho}{\nu}\|_2^2 - \frac{\nu}{2} \|\frac{\varrho}{\nu}\|_2^2.$$
(5)

We now use the alternating direction method of multipliers (ADMM). The basic idea of ADMM is to minimize the augmented Lagrangian L with respect to u_1 , u_2 , and v separately and to perform gradient ascent steps with respect to the Lagrange multipliers. (We refer to [10] for a detailed treatment on optimization strategies based on ADMM.) To simplify the expressions for $\arg \min_{u_s} L$, s = 1, 2, and $\arg \min_v L$, we will use the following lemma.

Lemma 2.1. For $a, b_1, ..., b_N \in \mathbb{R}$ and $x_1, ..., x_N > 0$, we have that

$$\sum_{i} x_{i}(a - b_{i})^{2} = (\sum_{i} x_{i})(a - \frac{\sum_{i} b_{i} x_{i}}{\sum_{i} x_{i}})^{2} + C$$

where $C \in \mathbb{R}$ is a constant that does not depend on a.

Proof. We calculate

$$\begin{split} \sum_{i} x_{i}(a-b_{i})^{2} &= a^{2}(\sum_{i} x_{i}) - 2a(\sum_{i} b_{i}x_{i}) + \sum_{i} b_{i}^{2}x_{i} \\ &= (\sum_{i} x_{i}) \left(a^{2} - 2a\frac{\sum_{i} b_{i}x_{i}}{\sum_{i} x_{i}} + \frac{\sum_{i} b_{i}^{2}x_{i}}{\sum_{i} x_{i}}\right) \\ &= (\sum_{i} x_{i}) \left(a^{2} - 2a\frac{\sum_{i} b_{i}x_{i}}{\sum_{i} x_{i}} + \left(\frac{\sum_{i} b_{i}x_{i}}{\sum_{i} x_{i}}\right)^{2} - \left(\frac{\sum_{i} b_{i}x_{i}}{\sum_{i} x_{i}}\right)^{2} - \frac{\sum_{i} b_{i}^{2}x_{i}}{\sum_{i} x_{i}}\right) \\ &= (\sum_{i} x_{i}) \left(\left(a - \frac{\sum_{i} b_{i}x_{i}}{\sum_{i} x_{i}}\right)^{2} - \left(\frac{\sum_{i} b_{i}x_{i}}{\sum_{i} x_{i}}\right)^{2} - \frac{\sum_{i} b_{i}^{2}x_{i}}{\sum_{i} x_{i}}\right). \end{split}$$

The last two terms do not depend on a, which shows the assertion.

Using Lemma 2.1, we can rearrange the quadratic summands of $\arg \min_{u_s} L$ for s = 1, 2 and those of $\arg \min_v L$. Doing so, we get the iteration

$$\begin{cases} u_{1}^{k+1} = \arg \min_{u_{1}} \frac{2\gamma}{\mu_{k}+\nu_{k}} \|\nabla_{p_{1}}u_{1}\|_{0} + \|u_{1} - \frac{1}{\mu_{k}+\nu_{k}}(\mu_{k}v^{k} + \nu_{k}u_{2}^{k} + \lambda_{1}^{k} - \rho^{k})\|_{2}^{2}, \\ u_{2}^{k+1} = \arg \min_{u_{2}} \frac{2\gamma}{\mu_{k}+\nu_{k}} \|\nabla_{p_{2}}u_{2}\|_{0} + \|u_{2} - \frac{1}{\mu_{k}+\nu_{k}}(\mu_{k}v^{k} + \nu_{k}u_{1}^{k+1} + \lambda_{2}^{k} + \rho^{k})\|_{2}^{2}, \\ v^{k+1} = \arg \min_{v} \|Av - f\|_{2}^{2} + \frac{\mu_{k}+\nu_{k}}{2} \|v - \frac{1}{2\mu_{k}}(\mu_{k}u_{1}^{k+1} + \mu_{k}u_{2}^{k+1} - \lambda_{1}^{k} - \lambda_{2}^{k})\|_{2}^{2}, \\ \lambda_{1}^{k+1} = \lambda_{1}^{k} + \mu_{k}(v^{k+1} - u_{1}^{k+1}), \\ \lambda_{2}^{k+1} = \lambda_{2}^{k} + \mu_{k}(v^{k+1} - u_{2}^{k+1}), \\ \rho^{k+1} = \rho^{k} + \nu_{k}(u_{1}^{k+1} - u_{2}^{k+1}). \end{cases}$$

$$(6)$$

As coupling parameter, we use an increasing sequence $(\mu_k)_{k \in \mathbb{N}}$. This is a slight refinement of the standard ADMM [65].

The crucial observation is that we can solve all of the subproblems of (6) efficiently. The first line decomposes into univariate Potts problems of the form

$$(u_1^{k+1})_{:,j} = \underset{g \in \mathbb{R}^m}{\arg\min} \frac{2\gamma}{\mu_k + \nu_k} \|\nabla g\|_0 + \|g - \frac{1}{\mu_k + \nu_k} (\mu_k v_{:,j}^k + \nu_k (u_2^k)_{:,j} + (\lambda_1^k)_{:,j} - \varrho_{:,j}^k)\|_2^2.$$
(7)



Figure 2 Reconstruction of geometric shapes from Radon data with 4 angles and noise level 0.01. The discretization \mathcal{N}_0 is anisotropic and tends to produce geometric staircasing. The neighborhood \mathcal{N}_1 improves the result significantly. The knight-move system \mathcal{N}_2 gives the most accurate reconstruction of the geometric shapes.

where we use the subscript notation $x_{:,j}$ to denote the *j*-th row of the $(m \times n)$ -image x, that is, $x_{:,j} = (x_{ij})_{i=1,...,m}$. Analogously, we get a decomposition for the second line of (6) into the problems

$$(u_{2}^{k+1})_{i,:} = \operatorname*{arg\,min}_{g \in \mathbb{R}^{n}} \frac{2\gamma}{\mu_{k} + \nu_{k}} \|\nabla g\|_{0} + \|g - \frac{1}{\mu_{k} + \nu_{k}} (\mu_{k} v_{i,:}^{k} + \nu_{k} (u_{1}^{k+1})_{i,:} + (\lambda_{2}^{k})_{i,:} + \varrho_{i,:}^{k})\|_{2}^{2}.$$
(8)

The third line of (6) is a classical L^2 Tikhonov regularization. The last three lines are simple gradient-ascent steps in the Lagrange multipliers. We briefly describe in Section 2.4 the strategies to solve these subproblems.

2.2. Design of Isotropic Discretizations

The anisotropic discretization of Section 2.1 measures the length of the jump set in the anisotropic Manhattan metric [22]. This leads to geometric staircasing in the reconstructions illustrated in Figure 2. The Euclidean length can be approximated better when complementing the neighborhood system with finite-difference vectors, for example, diagonal directions or "knight-move" directions [23]. We now present a general scheme to construct appropriate neighborhood systems.

The starting point is the anisotropic neighborhood system

$$\mathcal{N}_0 = \{(1,0), (0,1)\}.$$

The vectors in this system have the (formal) slopes 0 and ∞ . We add a new finitedifference vector $(x, y) \in \mathbb{Z}^2$ to the system only if its slope y/x is not yet contained in the system. For example, we can add the vector (1, 1) with slope 1. For reasons of symmetry, we also add the orthogonal vector (1, -1). Thus, we get the neighborhood system

$$\mathcal{N}_1 = \{(1,0), (0,1), (1,1), (1,-1)\}.$$
(9)

The next vectors to include in the neighborhood system are the four knight move vectors $(\pm 2, 1)$, $(1, \pm 2)$ which leads to the system

$$\mathcal{N}_2 = \{(1,0), (0,1), (1,1), (1,-1), (-2,-1), (-2,1), (2,1), (2,-1)\}.$$
(10)

The general scheme of adding new vectors corresponds to the standard enumeration of the rational numbers.

Appropriate weights can be derived as follows. Let us assume that u is a binary $(n \times n)$ image with an edge along the direction $(x, y) \in \mathcal{N}$. We first look at lines with a slope y/x between (-1) and 1 going from the left to the right boundary of the image. (If the slope of (x, y) is not in the interval [-1, 1] then we look at the $\pi/2$ -rotated image and exchange the roles of x and y.) The Euclidean length of such a line is given by $n\sqrt{x^2 + y^2}/x$. Since we want that the total jump length of this image equals that Euclidean length, we get a condition on the weights that takes the form

$$\sum_{s=1}^{S} \omega_s \, \|\nabla_{p_s} u\|_0 = n \frac{\sqrt{x^2 + y^2}}{x}.$$
(11)

It remains to evaluate the left-hand side of (11) for the binary image u. This can be done either manually for small neighborhood systems or with the help of a computer program for larger neighborhood systems. When counting the non-zero entries of $\nabla_{p_s} u$ we assume n to be large so that boundary effects are negligible. We end up with a system of S equations for the S unknowns. For the diagonal neighborhood system \mathcal{N}_1 , (11) yields the conditions

$$\begin{split} \omega_1 &+ \omega_3 + \omega_4 = 1, \\ \omega_2 + \omega_3 + \omega_4 = 1, \\ \omega_1 + \omega_2 + 2\omega_3 &= \sqrt{2}, \\ \omega_1 + \omega_2 &+ 2\omega_4 = \sqrt{2}. \end{split}$$

Solving this linear system, we get the weights

$$\omega_1 = \omega_2 = \sqrt{2} - 1$$
 and $\omega_3 = \omega_4 = 1 - \frac{\sqrt{2}}{2}$

For the knight-move neighborhood system \mathcal{N}_2 , we get an analogous system of equations in S = 8 unknowns which gives us the weights

$$\omega_s = \begin{cases} \sqrt{5} - 2, & \text{for } s = 1, 2, \\ \sqrt{5} - \frac{3}{2}\sqrt{2}, & \text{for } s = 3, 4, \\ \frac{1}{2}(1 + \sqrt{2} - \sqrt{5}), & \text{for } s = 5, ..., 8. \end{cases}$$

We now turn to the question of how well we approximate the Euclidean length with the above discretizations. The neighborhood systems give rise to a norm $\|\cdot\|_{\mathcal{N}}$ defined for $p \in \mathbb{R}^2$ by

$$\|p\|_{\mathcal{N}} = \sum_{s=1}^{S} \omega_s |\langle p, p_s \rangle|$$

By construction, the length $||p_s||_{\mathcal{N}}$ coincides with the Euclidean length $||p_s||_2$ for all vectors p_s in the neighborhood system as illustrated in Figure 3. In [23], it is proposed to measure the isotropy of a finite-difference system by the ratio E between the longest and the shortest unit vector with respect to that length, compiled as

$$E = \max_{\|p\|_2=1} \|p\|_{\mathcal{N}} / \min_{\|p\|_2=1} \|p\|_{\mathcal{N}}.$$



(c) System with knight moves \mathcal{N}_2 using the weights of [23] $(E'_2 \approx 1.05)$.

(d) System with knight moves \mathcal{N}_2 using our weights $(E_2 \approx 1.03)$.

Figure 3 The solid line represents the length of a Euclidean unit vector as measured in the finite-difference systems (as a function of the angle). The red dots identify the normalized vectors of the neighborhood system $p_s/||p_s||_2$ and the dashed line is the Euclidean unit circle. The isotropy increases significantly when passing from \mathcal{N}_0 to \mathcal{N}_1 . The increase in isotropy is less substantial when passing from \mathcal{N}_1 to \mathcal{N}_2 .

The closer the quantity E is to one the higher is the isotropy. For the anisotropic system \mathcal{N}_0 , we get the value $E_0 = \sqrt{2} \approx 1.41$. The introduction of diagonal directions reduces this value significantly to $E_1 \approx 1.08$. If we include also the knight-move differences (\mathcal{N}_2), then the ratio improves further to $E_2 \approx 1.03$. We note that the weights for the system \mathcal{N}_1 coincide with those proposed in [23], up to a normalization factor. For \mathcal{N}_2 , our weights are more isotropic than the weights of [23].

2.3. Splitting Scheme for General Discretizations

We now derive a minimization strategy for the general discretization (2). Let us denote the neighborhood system by $\mathcal{N} = \{p_1, ..., p_S\}$ and let $\omega_1, ..., \omega_S > 0$ where $S \ge 2$. We first rewrite (2) as the constrained optimization problem

minimize
$$\gamma \sum_{s=1}^{S} \omega_s \|\nabla_{p_s} u_s\|_0 + \|Av - f\|_2^2$$

subject to $u_r - u_t = 0$, for all $1 \le r < t \le S$,
 $v - u_s = 0$, for all $1 \le s \le S$. (12)

The augmented Lagrangian of this optimization problem reads

$$L = \gamma \sum_{s=1}^{S} \omega_{s} \|\nabla_{p_{s}} u_{s}\|_{0} + \frac{\mu}{2} \|v - u_{s} + \frac{\lambda_{s}}{\mu}\|_{2}^{2} + \frac{\nu}{2} \sum_{1 \le r < t \le S} \|u_{r} - u_{t} + \frac{\varrho_{r,t}}{\nu}\|_{2}^{2} + \|Av - f\|_{2}^{2}.$$
 (13)

The parameter $\nu > 0$ controls how strong the split variables $u_1, ..., u_S$ are tied to each other and $\mu > 0$ controls their coupling to v. The variables $\lambda_s, \varrho_{r,t} \in \mathbb{R}^{n \times m}$ are Lagrange multipliers. In the ADMM iteration, we minimize L sequentially with respect to $v, u_1, ..., u_S$. The minimization of L with respect to u_s reads

$$\underset{u_s}{\operatorname{arg\,min}\,} L = \underset{u_s}{\operatorname{arg\,min}\,} \gamma \omega_s \|\nabla_{p_s} u_s\|_0 + \frac{\mu}{2} \|v - u_s + \frac{\lambda_s}{\mu}\|_2^2 + \frac{\nu}{2} \sum_{1 \le r < s} \|u_r - u_s + \frac{\varrho_{r,s}}{\nu}\|_2^2 + \frac{\nu}{2} \sum_{s < t \le S} \|u_s - u_t + \frac{\varrho_{s,t}}{\nu}\|_2^2.$$
(14)

We rearrange this expression using Lemma 2.1 to

$$\underset{u_s}{\arg\min} L = \underset{u_s}{\arg\min} \frac{2\gamma\omega_s}{\mu + \nu(S-1)} \|\nabla_{p_s} u_s\|_0 + \|u_s - w_s\|_2^2$$

with

$$w_s = \frac{\mu v + \lambda_s + \sum_{1 \le r < s} (\nu u_r + \varrho_{r,s}) + \sum_{s < t \le S} (\nu u_t - \varrho_{s,t})}{\mu + \nu(S-1)}.$$

In a similar way, we consider the minimizer with respect to v as

$$\underset{v}{\arg\min} L = \underset{v}{\arg\min} ||Av - f||_{2}^{2} + \sum_{s=1}^{S} \frac{\mu}{2} ||v - u_{s} + \frac{\lambda_{s}}{\mu}||_{2}^{2}$$

which we rewrite using Lemma 2.1 as

$$\underset{v}{\arg\min} L = \underset{v}{\arg\min} ||Av - f||_{2}^{2} + \frac{\mu S}{2} ||v - \frac{1}{S} \sum_{s=1}^{S} (u_{s} - \frac{\lambda_{s}}{\mu})||_{2}^{2}.$$

Having computed explicit expressions for the minimization with respect to each variable, we obtain the ADMM iteration

$$\begin{cases}
 u_{1}^{k+1} = \arg \min_{u_{1}} \frac{2\gamma\omega_{1}}{\mu_{k}+\nu_{k}(S-1)} \|\nabla_{p_{1}}u_{1}\|_{0} + \|u_{1}-w_{1}^{k}\|_{2}^{2}, \\
 \vdots \\
 u_{S}^{k+1} = \arg \min_{u_{S}} \frac{2\gamma\omega_{S}}{\mu_{k}+\nu_{k}(S-1)} \|\nabla_{p_{S}}u_{S}\|_{0} + \|u_{S}-w_{S}^{k}\|_{2}^{2}, \\
 v_{s}^{k+1} = \arg \min_{v} \|Av - f\|_{2}^{2} + \frac{\mu_{k}S}{2} \|v - z^{k}\|_{2}^{2}, \\
 \lambda_{s}^{k+1} = \lambda_{s}^{k} + \mu_{k}(v^{k+1} - u_{s}^{k+1}), \text{ for all } 1 \leq s \leq S, \\
 \varrho_{r,t}^{k+1} = \varrho_{r,t}^{k} + \nu_{k}(u_{r}^{k+1} - u_{t}^{k+1}), \text{ for all } 1 \leq r < t \leq S.
 \end{cases}$$
(15)

a

Here, w_s^k is given by

$$w_s^k = \frac{\mu_k v^k + \lambda_s^k + \sum_{1 \le r < s} (\nu_k u_r^{k+1} - \varrho_{r,s}^k) + \sum_{s < t \le S} (\nu_k u_t^k + \varrho_{s,t}^k)}{\mu_k + \nu_k (S-1)}$$

and z^k by

$$z^{k} = \frac{1}{S} \sum_{s=1}^{S} \left(u_{s}^{k+1} - \frac{\lambda_{s}^{k}}{\mu_{k}} \right).$$

The key observation every of the subproblem in the ADMM iteration can be solved efficiently. The minimization problems in $u_1, ..., u_S$ decompose into univariate Potts problems with respect to the directions induced by the finite difference vectors p_s . To fix ideas, consider the finite-difference vector $p_s = (1, 1)$. Then, the solution u_s^{k+1} is given by solving one univariate Potts problem for each diagonal line of the two-dimensional array w_s^k . The last subproblem, like in the anisotropic case, is a classical Tikhonov-type regularization.

We eventually remark that the anisotropic splitting (6) is a special case of the general form (15) if we choose the anisotropic finite-difference system \mathcal{N}_0 .

2.4. Solution of the Subproblems

Solution methods for the univariate Potts and classical Tikhonov problems are well studied. Since they are the important building blocks of our iteration (15), we briefly recall the idea of the algorithms.

The classical univariate Potts problem is given by

$$P_{\gamma}(g) = \gamma \|\nabla g\|_{0} + \|g - f\|_{2}^{2} \to \min,$$
(16)

where $g, f \in \mathbb{R}^n$ and $\|\nabla g\|_0 = |\{i : g_i \neq g_{i+1}\}|$ denotes the number of jumps of g. This can be solved exactly by dynamic programming [22, 37, 59, 60, 73, 79, 82]. The basic idea is that a minimizer of the Potts functional for data $(f_1, ..., f_r)$ can be computed in polynomial time provided that minimizers of the partial data $(f_1), (f_1, f_2), ..., (f_1, ..., f_{r-1})$ are known. We denote the respective minimizers by $g^1, g^2, ..., g^{r-1}$. In order to compute a minimizer for data $(f_1, ..., f_r)$, we first create a set of r minimizer candidates $h^1, ..., h^r$, each of length r. These minimizer candidates are given by

$$h^{\ell} = (g^{\ell-1}, \underbrace{\mu_{[\ell,r]}, \dots, \mu_{[\ell,r]}}_{\text{Length}}),$$
(17)

where g^0 is the empty vector and $\mu_{[\ell,r]}$ denotes the mean value of data $f_{[\ell,r]} = (f_{\ell}, ..., f_r)$. Among the candidates h^{ℓ} , one with the least Potts functional value is a minimizer for the data $f_{[1,r]}$.

In [37], Friedrich et al. proposed the following $\mathcal{O}(n^2)$ time and $\mathcal{O}(n)$ space algorithm. They observed that the functional values of a minimizer $P_{\gamma}(g^r)$ for data $f_{[1,r]}$ can be computed directly from the functional values $P_{\gamma}(g^1), \ldots, P_{\gamma}(g^{r-1})$ and the squared mean deviations of the data $f_{[1,r]}, \ldots, f_{[r,r]}$. Indeed, using (17), the functional value of the minimizer g^r is given (setting $P_{\gamma}(g^0) = -\gamma$) by

$$P_{\gamma}(g^{r}) = \min_{\ell=1,\dots,r} P_{\gamma}(g^{\ell-1}) + \gamma + d_{[\ell,r]},$$
(18)

where $d_{[\ell,r]}$ denotes the squared deviation from the mean value

$$d_{[\ell,r]} = \min_{y \in \mathbb{R}} \|y - f_{[\ell,r]}\|_2^2 = \|(\mu_{[\ell,r]}, ..., \mu_{[\ell,r]}) - f_{[\ell,r]}\|_2^2$$

The evaluation of (18) is $\mathcal{O}(n)$ if we precompute the first and second moments of data $f_{[\ell,r]}$. If ℓ^* denotes the minimizing argument in (18), then $(\ell^* - 1)$ indicates the rightmost jump location at step r, which is stored as J(r). The jump locations of a solution g^r are thus J(r), J(J(r)), J(J(J(r))),...; the values of g^r between two consecutive jumps are given by the mean of f on this interval. Note that we only have to compute and store the jump locations J(r) and the minimal Potts functional value $P_{\gamma}(g^r)$ in each iteration. The reconstruction of the minimizer from the jump locations only has to be done once for g^n at the end; it is thus not time-critical.

The algorithm for solving (16) consists of two nested loops for r = 1, ..., n and $\ell = 1, ..., r$, which amounts to n(n + 1)/2 iterations in total. Typically, a significant amount of configurations are unreachable and thus can be skipped [73, Theorem 2]. The time complexity is still $\mathcal{O}(n^2)$, but the practical runtime is improved by a fourfold to fivefold factor. Refer to [73] for the complete flow diagram of the accelerated algorithm.

Our second subproblem is the solution of a classical L^2 Tikhonov regularization

$$v^{k+1} = \arg\min_{v} \|Av - f\|_2^2 + \frac{\mu}{2} \|v - z\|_2^2$$
(19)

with some $z \in \mathbb{R}^{m \times n}$. The unique minimizer of this problem is given by the solution of the normal equation

$$(A^*A + \frac{\mu}{2}I)v = A^*f + \frac{\mu}{2}z.$$
 (20)

Here, A^* is the adjoint of A. This linear system can be solved using, for example, the conjugate-gradient method. In some cases we can exploit the structure of A for more efficient solution methods. This is the case when A is the Radon transform or a convolution operator (see Section 3).

2.5. General Data Terms

Inspecting the ADMM iteration, we observe that the data term only appears in the first line. That line consists of a classical Tikhonov regularization with the L^2 data term $||Av - f||_2^2$. Minimizers of that problem can be solved efficiently for many other data terms d(u, f), such as L^p data terms, $p \ge 1$, of the form

$$d(v,f) = \|Av - f\|_p^p$$

or a Huber data term which is a hybrid between L^1 and L^2 data terms [25, 78]. In general, our algorithm is applicable whenever the proximal point operator $\operatorname{prox}_{d(\cdot,f)/\mu}$ of $d(\cdot, f)$, defined by

$$\operatorname{prox}_{d(\cdot,f)/\mu}(z) = \arg\min_{v} d(v,f) + \frac{\mu}{2} \|v - z\|_{2}^{2},$$

can be evaluated efficiently. This is often the case when $d(\cdot, f)$ is a convex functional. We refer to [11] for an extensive overview on strategies for convex optimization.

2.6. Convergence

In this section, we show that Algorithm (15) converges in the prototypical case $\nu_k = 0$ for all k. We leave a convergence proof for Algorithm (15) with general ν_k as an open problem. For the proof, we use methods developed in [74].

Theorem 2.2. Let the sequence $(\mu_k)_{k \in \mathbb{N}}$ be increasing and satisfy $\sum_k \mu_k^{-1/2} < \infty$. Further, let $\nu_k = 0$ for all k. Then, the iteration (15) converges in the sense that

$$(u_1^k, \dots, u_S^k, v^k) \to (u_1^*, \dots, u_S^*, v^*) \quad with \quad u_1^* = \dots = u_S^* = v^*,$$
$$\frac{\lambda_s^k}{\mu_k} \to 0 \quad for \ all \quad s \in \{1, \dots, S\}.$$
(21)

Proof. We denote the S functionals appearing in the first S lines of (15) by F_s^k , i.e.,

$$F_s^k(u_s) = \frac{2\gamma\omega_s}{\mu_k} \|\nabla_{p_s} u_s\|_0 + \|u_s - (v^k + \frac{\lambda_s^k}{\mu_k})\|_2^2$$

Using this notation, we rewrite the first S lines of (15) as $u_s^{k+1} \in \arg\min_u F_s^k(u_s)$ for all $s \in \{1, \ldots, S\}$. We first estimate the distance $||u_s^{k+1} - (v^k - \frac{\lambda_s^k}{\mu_k})||_2$. To that end, we note that $F_s^k(u_s^{k+1}) \leq F_s^k\left(v^k - \frac{\lambda_s^k}{\mu_k}\right)$ which holds true since u_s^{k+1} minimizes F_s^k . Applying the definition of F_s^k , we get

$$\gamma \|\nabla_{p_s} u_s^{k+1}\|_0 + \frac{\mu_k}{2} \|u_s^{k+1} - \left(v^k - \frac{\lambda_s^k}{\mu_k}\right)\|_2^2 \le \gamma \|\nabla_{p_s} \left(v^k - \frac{\lambda_s^k}{\mu_k}\right)\|_0 \le \gamma L,$$

where $L = \max(N, M)$ is the maximal side length of the considered $(N \times M)$ image. Since the first summand on the left-hand side is nonnegative, we get that

$$\|u_s^{k+1} - \left(v^k - \frac{\lambda_s^k}{\mu_k}\right)\|_2^2 \le \frac{\gamma L}{\mu_k}.$$
(22)

In particular, for all $s \in \{1, \ldots, S\}$, we obtain that

$$\lim_{k \to \infty} u_s^{k+1} - \left(v^k - \frac{\lambda_s^k}{\mu_k} \right) = 0.$$
(23)

We now draw our attention to the (S + 1)th line of (15). We denote the corresponding functional by G^k , i.e.,

$$G^{k}(v) = \|Av - f\|_{2}^{2} + \frac{\mu_{k}S}{2}\|v - \frac{1}{S}\sum_{s=1}^{S}(u_{s}^{k+1} - \frac{\lambda_{s}^{k}}{\mu_{k}})\|_{2}^{2}.$$

The minimality of v^{k+1} implies the inequality

$$G^{k}(v^{k+1}) \leq G^{k}\left(\frac{1}{S}\sum_{s=1}^{S}(u_{s}^{k+1}-\frac{\lambda_{s}^{k}}{\mu_{k}})\right).$$

We now apply the definition of G^k to estimate

$$\begin{split} \|Av^{k+1} - f\|_{2}^{2} + \frac{\mu_{k}S}{2} \|v^{k+1} - \left(\frac{1}{S}\sum_{s=1}^{S}(u^{k+1}_{s} - \frac{\lambda^{k}_{s}}{\mu_{k}})\right)\|_{2}^{2} \\ \leq \|A\left(\frac{1}{S}\sum_{s=1}^{S}(u^{k+1}_{s} - \frac{\lambda^{k}_{s}}{\mu_{k}})\right) - f\|_{2}^{2} \\ \leq \|A\left(\frac{1}{S}\sum_{s=1}^{S}(u^{k+1}_{s} - \frac{\lambda^{k}_{s}}{\mu_{k}} - v^{k})\right) + Av^{k} - f\|_{2}^{2} \\ \leq \left(\|A\|\|\frac{1}{S}\sum_{s=1}^{S}(u^{k+1}_{s} - \frac{\lambda^{k}_{s}}{\mu_{k}} - v^{k})\|_{2} + \|Av^{k} - f\|_{2}\right)^{2}. \end{split}$$
(24)

Here, ||A|| is the operator norm of A acting on ℓ^2 . We now combine (24) and (22) to estimate the magnitude of the residuals $Av^{k+1} - f$. We get

$$||Av^{k+1} - f|| \le \frac{C}{\sqrt{\mu_k}} + ||Av^k - f||,$$

where C > 0 is a constant that only depends on γ , L, and ||A||. Solving this recursion yields

$$||Av^{k+1} - f|| \le C \sum_{j=1}^{k} \frac{1}{\sqrt{\mu_j}} + ||Av^0 - f||_p,$$

which shows that the sequence of residuals $(Av^{k+1} - f)_{k \in \mathbb{N}}$ is bounded. Together with (24), this implies

$$\begin{aligned} \frac{\mu_k S}{2} \|v^{k+1} - \frac{1}{S} \sum_{s=1}^S (u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k) \|^2 &\leq (\|A\| \| \frac{1}{S} \sum_{s=1}^S (u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k) \| + C')^2 \\ &\leq (\|A\| \frac{1}{S} \sum_{s=1}^S \| (u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k) \| + C')^2. \end{aligned}$$

Here, C' is a positive constant independent of k. We now use (23) to conclude that the sequence (with respect to k)

$$\mu_k \|v^{k+1} - \frac{1}{S} \sum_{s=1}^S (u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k)\|_2^2 \text{ is bounded.}$$
(25)

We use this fact to establish the convergence of the sequence v^k by showing that it is a Cauchy sequence. We estimate

$$\|v^{k+1} - v^k\| \le \|v^{k+1} - \frac{1}{S}\sum_{s=1}^S (u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k)\| + \|\frac{1}{S}\sum_{s=1}^S (u_s^{k+1} - \frac{\lambda_s^k}{\mu_k} - v^k) - v^k\|.$$

We now apply (22) as well as (25) to the summands on the right-hand side to obtain

$$||v^{k+1} - v^k|| \le \frac{C''}{\sqrt{\mu_k}}$$

for some constant C'' > 0 which is again independent of k. The assumption on the sequence μ_k guarantees that v^k is a Cauchy sequence and hence that v^k converges to some v^* .

To establish the last statement in (21), we rewrite each of the last S lines in (15) to obtain the identity

$$\frac{\lambda_s^{k+1}}{\mu_k} = \left(\frac{\lambda_s^k}{\mu_k} + u_s^{k+1} - v^k\right) + (v^k - v^{k+1}).$$
(26)

By (23) and (25), each term in parenthesis converges to 0. Hence,

$$\lim_{k\to\infty}\frac{\lambda_s^{k+1}}{\mu_k}=0.$$

Since we assume that the sequence μ_k is nondecreasing, we have that $\mu_k/\mu_{k+1} \leq 1$ and thus, for all $s = 1, \ldots, S$,

$$\lim_{k \to \infty} \frac{\lambda_s^k}{\mu_k} = 0.$$

This shows the last statement in (21). Finally, we rewrite the third line of (15) as $u_s^{k+1} - v^{k+1} = (\lambda_s^{k+1} - \lambda_s^k)/\mu_k$ to obtain the inequality

$$||u_s^{k+1} - v^{k+1}|| \le \frac{||\lambda_s^{k+1}||}{\mu_k} + \frac{||\lambda_s^k||}{\mu_k} \to 0.$$

This means that $u_s^k - v^k \to 0$ for all $s = 1, \ldots, S$ and, since v^k converges, also each u_s^k converges and the corresponding limit u_s^* equals v^* , which completes the proof. \Box

3. Applications

We now demonstrate the applicability of our method to several imaging problems. We first consider tomographic problems that are modeled by the classical Radon transform and by closely related operators; in particular, x-ray computed tomography and positron emission tomography [61]. Further, we look at thermoacoustic and photoacoustic tomography which are frequently modeled by the spherical Radon transform [76]. Furthermore, we apply our method to deconvolution problems. If the imaged object is (approximately) piecewise constant then the result can be seen as reconstruction of the imaged object. Otherwise, the results have to be understood as a joint reconstruction and segmentation.

Unless stated otherwise, the setup for the numerical experiments is as follows: We use the ADMM iteration (15) with the coupling sequence $\mu_k = 10^{-6} \cdot k^{\tau}$ with $\tau = 2.01$ and we choose ν_k identically zero. This choice satisfies the hypothesis of Theorem 2.2. For the experiments involving the Shepp-Logan phantom, we observed the best results using the neighborhood system \mathcal{N}_1 of (9) while the neighborhood \mathcal{N}_2 of (10) gave better result for the more realistic images (Figures 6, 8, and 9); see also the comparison in Figure 2. For the solution of the Tikhonov type problem (19), we use Matlab's conjugate-gradient method on the normal equation (20). We use a "warm start", which means that we use the solution of the previous iteration v^k as initial guess for v^{k+1} . Then the conjugate-gradient iteration converges typically in few steps. The split variables v, u_s for $s \geq 1$, and the Lagrange multipliers are all initialized with 0. We stop the iteration when the relative L^2 error of u_1 and u_2 falls below 10^{-3} .



Figure 4 Our method (bottom row) is able to recover the large geometric structures from Radon data (360 projections), even in the presence of extremely high noise. For comparison, the results of the filtered backprojection are shown in the top row. (See Figure 1a for the original image.)

In our simulated experiments, the noise is Gaussian with standard deviation $\sigma = \text{noiselevel} \cdot ||f||_{\infty}$. The model parameter γ was adjusted empirically. The quality of a reconstruction u is measured in terms of the peak signal-to-noise-ratio given by $\text{PSNR}(u) = 10 \log_{10} \left(\frac{mn ||g||_{\infty}^2}{||g-u||_2^2} \right)$ where $g \in \mathbb{R}^{m \times n}$ is the ground truth. We further use the mean structural similarity index (MSSIM) [77] which is better suited to perceived visual quality than the PSNR as it takes the similarity of local structures into account. The MSSIM is bounded from above by 1 and a higher MSSIM means a better reconstruction quality. Here, we use Matlab's function ssim with standard parameters for the computation of the MSSIM.

3.1. Radon Data

The Radon transform is defined by

$$\mathcal{R}u(\theta, s) = \int_{-\infty}^{\infty} u(s\theta + t\theta^{\perp}) \,\mathrm{d}t, \qquad (27)$$

where $s \in \mathbb{R}$, $\theta \in S^1$, and θ^{\perp} is the unit vector $\pi/2$ radians counterclockwise from θ . It serves as the mathematical model of x-ray tomography and positron emission tomography. The Radon transform is a bounded linear operator from $L^2(\Omega)$ to $L^2(S^1 \times \mathbb{R})$, where Ω denotes the closed unit ball in \mathbb{R}^2 . Reconstruction from Radon data is special in our approach because we can solve the Tikhonov problem (19)



Figure 5 Reconstruction of the Shepp-Logan phantom from highly undersampled Radon data with noise. The fine details vanish but the large geometric structures are still are recovered by our method (bottom row). Filtered backprojection results are shown for comparison (top row).

very efficiently using a filtered-backprojection-type algorithm. We first recall that the backprojection operator \mathcal{R}^* is defined via

$$\mathcal{R}^* f(x) = \int_{S^1} f(\theta, x \cdot \theta) \,\mathrm{d}\theta$$

Now we can formulate a regularized variant of the classical filtered backprojection. Since we have not found the statement in this form in the literature, we provide a short proof.

Lemma 3.1. Let $f = \mathcal{R}w$ for some $w \in L^2(\Omega)$ and $\alpha > 0$. Then,

$$\underset{v \in L^{2}(\Omega)}{\arg\min} \|\mathcal{R}v - f\|_{2}^{2} + \alpha \|v - z\|_{2}^{2} = \mathcal{R}^{*}H_{\alpha}(f - \mathcal{R}z),$$
(28)

where the filtering operator is defined via

$$H_{\alpha}f = \mathcal{F}_s^{-1}(h_{\alpha}\mathcal{F}_sf)$$

with the filter function

$$h_{\alpha}(r) = \frac{|r|}{4\pi + \alpha \, |r|}.$$

Above, \mathcal{F}_s (and \mathcal{F}_s^{-1}) denotes the one dimensional Fourier transform (and its inverse) of a function $f(\theta, s)$ with respect to the parameter s.

Proof. Setting u = v - z and $g = f - \mathcal{R}z$ we first rewrite the cost functional as

$$\|\mathcal{R}v - f\|_{2}^{2} + \alpha \|v - z\|_{2}^{2} = \|\mathcal{R}u - g\|_{2}^{2} + \alpha \|u\|_{2}^{2}.$$
 (29)

In what follows, we denote by u_{α} the minimizer of the right-hand side in (29). We first note that u_{α} satisfies the normal equation $\mathcal{R}^*\mathcal{R}u_{\alpha} + \alpha u_{\alpha} = \mathcal{R}^*g$ (see (20)). Further, we get from [61, Theorem II.1.5] that $\mathcal{R}^*\mathcal{R}u_{\alpha} = 2(\|\cdot\|_2^{-1} * u_{\alpha})$. Therefore, by taking the Fourier transform on both sides of the normal equation and applying the convolution theorem together with [42, Ch.V, Lemma 5.2], we obtain the following relation for the Fourier transform of u_{α} :

$$\hat{u}_{\alpha}(\xi) = \frac{\|\xi\|_2}{4\pi + \alpha \, \|\xi\|_2} \widehat{\mathcal{R}^*g}(\xi).$$
(30)

Formula (30) is a special case of a formula derived in [45, Sec. 7]. Now, since $g = \mathcal{R}w'$ with w' = w - z, a similar argument shows that $\widehat{\mathcal{R}^*\mathcal{R}w'}(\xi) = 4\pi \|\xi\|_2^{-1} \widehat{w'}(\xi)$ and, therefore, that

$$\hat{u}_{\alpha}(\xi) = \frac{4\pi}{4\pi + \alpha \left\|\xi\right\|_2} \widehat{w'}(\xi).$$
(31)

Next, we make use of the Fourier-slice theorem [61, Theorem II.1.1]. It states that $\widehat{\mathcal{R}_{\theta}w'}(r) = \sqrt{2\pi} \widehat{w'}(r\theta)$, where $\mathcal{R}_{\theta}w' = \mathcal{R}w'(\theta, \cdot)$. Using (31) we obtain

$$\begin{split} u_{\alpha}(x) &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \widehat{u}_{\alpha}(\xi) e^{ix \cdot \xi} \, \mathrm{d}\xi \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^2} \frac{4\pi}{4\pi + \alpha \left\|\xi\right\|_2} \widehat{w'}(\xi) e^{ix \cdot \xi} \, \mathrm{d}\xi \\ &= \int_{S^1} 2 \int_0^\infty \frac{1}{4\pi + \alpha \left|r\right|} \widehat{w'}(r\theta) e^{irx \cdot \theta} r \, \mathrm{d}r \, \mathrm{d}\theta \\ &= \int_{S^1} \int_{-\infty}^\infty \frac{|r|}{4\pi + \alpha \left|r\right|} \widehat{w'}(r\theta) e^{irx \cdot \theta} \, \mathrm{d}r \, \mathrm{d}\theta \\ &= \int_{S^1} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \frac{|r|}{4\pi + \alpha \left|r\right|} \widehat{\mathcal{R}_{\theta} w'}(r) e^{irx \cdot \theta} \, \mathrm{d}r \, \mathrm{d}\theta \\ &= \int_{S^1} H_{\alpha} \mathcal{R} w'(\theta, x \cdot \theta) \, \mathrm{d}\theta \\ &= \mathcal{R}^* H_{\alpha} \mathcal{R} w'(x). \end{split}$$

Since $\mathcal{R}w' = \mathcal{R}w - \mathcal{R}z = f - \mathcal{R}z$, the assertion follows.

We show in Figure 4 the result of our method for the reconstruction of the Shepp-Logan phantom from Radon data with dense angular sampling (360 angles). For very high noise levels the small details vanish but the larger geometric structures are still reconstructed.

Using (28) is efficient but it is strongly affected by angular undersampling. Whenever the undersampling is too pronounced, using (28) leads to large reconstruction errors. In such cases, we resort to the standard conjugate-gradient method on the normal equation (20). With this strategy, we obtain a high-quality reconstruction of the Shepp-Logan phantom from only 7 projections of noise-free data as illustrated in Figure 1 (PSNR above 40 dB, MMSIM above 0.99). In Figure 5, we

Joint Image Reconstruction and Segmentation Using the Potts Model



(a) FBP using 192 projections.

(b) Proposed method using 192 projections.

(c) FBP using 24 projections.

(d) Proposed method using 24 projections.

Figure 6 Joint reconstruction and segmentation from undersampled PET data of a physical thorax phantom [32]. The proposed method segments the anatomic structures (lung, spine, and thorax body). Filtered backprojection results are shown for comparison.

perform similar experiments, this time with noise added and a few more measurements. We observe that, while the small structures are lost in the reconstruction, the geometric structures at large scale are well recovered.

Next, we apply our method to PET data. The PET model is closely related to the CT model. The PET data is given as a product of the Radon transform with a function depending on the attenuation. Under the assumption that this function is known, the remaining problem is the computation of u from its line integrals, which is exactly the tomography problem. In Figure 6, we see the results for a PET data of a physical thorax phantom [32]. Our method is able to jointly reconstruct and to segment the anatomic structures (lung, spine, and thorax body) from highly undersampled data.

3.2. Spherical Radon Data

The spherical mean Radon transform \mathcal{M} is given by

$$\mathcal{M}u(\theta, t) = \int_{S^1} u(\theta + t\zeta) \, \mathrm{d}\zeta.$$
(32)

The value $\mathcal{M}u(\theta, t)$ is the average of u over the sphere centered at $\theta \in \mathbb{R}^2$ with radius t > 0. Data of this form appear naturally in various models of photoacoustic tomography. PAT is a hybrid imaging modality that uses laser pulses for the generation of pressure waves (acoustic signal) inside of the imaged object. These acoustic waves propagate through the tissue and beyond. They are finally recorded by ultrasonic detectors (transducers) which are distributed around the object [75, 76].

The corresponding image reconstruction problem consists of recovering a function $u : \mathbb{R}^2 \to \mathbb{R}$ from the knowledge of certain spherical integrals $\mathcal{M}u(\theta, t)$. Here, the centers of the circles are placed on an acquisition curve $\Xi \subset \mathbb{R}^2$ that surrounds the object. The radii of the circles vary in $[0, \infty)$. The problem of the inversion of \mathcal{M} has been intensively investigated in recent years and still attracts much attention [2, 18, 33, 54, 76]. In many practical PAT setups, the acquisition curve is a circle so that we assume $\Xi = S^1$ from now on. We consider a compact set $\Omega \subset \mathbb{R}^2$ that is contained inside the interior of the ball bounded by S^1 . Then, the spherical mean Radon transform is a bounded linear operator from $L^2(\Omega)$ to $L^2(S^1 \times [0, 2]), \, \mathrm{d}\theta r \, \mathrm{d}r)$. For the solution of the Tikhonov problem (19), we need the adjoint of \mathcal{M} . This adjoint



Figure 7 Reconstruction of the Shepp-Logan phantom from few projections of the spherical Radon transform (Original: Figure 1a). The filtered backprojection-type reconstructions suffer heavily from artifacts (top row). Although some fine structures get lost, our reconstruction recovers the large-scale structures (bottom row).

is given by the backprojection operator \mathcal{M}^* which is defined by

$$\mathcal{M}^* f(x) = \int_{S^1} f(\theta, |x - \theta|) \,\mathrm{d}\theta.$$
(33)

The integral (33) is well-defined for sufficiently regular functions. Using density arguments, the operator is continuously extended to L^2 .

We solve the Tikhonov problem using the standard conjugate gradients on the normal equation. To this end, we implemented the projection operator (32) and the backprojection operator (33) using numerical quadrature. The exact projection centers are $\{k \cdot \Delta \theta_M : k = 1, ..., M\}$, where $\Delta \theta_M = 360^{\circ}/M$ and M denotes the number of centers.

We compare our approach with filtered-backprojection-type reconstruction algorithms as proposed in [2, 34, 41]. In this work, we use R. Seyfried's implementation of the algorithm in [2, 34]. We briefly recall the basic idea: The reconstruction is computed approximatively by using summability kernels $K_{\varepsilon} : \Omega \times \Omega \to \mathbb{R}, \ \varepsilon \in (0, 1)$, of the form

$$K_{\varepsilon}(x,y) = \int_{S^1} h_{\varepsilon}(x,\zeta,|y-\zeta|) \,\mathrm{d}\zeta.$$

Then, an approximation to u can be computed from spherical mean values of u via

$$u_{\varepsilon}(x) = \int_{\Omega} K_{\varepsilon}(x, y) u(y) \, \mathrm{d}y = \int_{0}^{2} \int_{S^{1}} h_{\varepsilon}(x, \theta, t) \mathcal{M}u(\theta, t) \, \mathrm{d}\theta \, t \, \mathrm{d}t.$$



Figure 8 Joint reconstruction and segmentation of an image blurred by a Gaussian kernel ($\sigma = 10$) and corrupted by extreme Gaussian noise of level 0.5.

Under certain smoothness assumptions on h_{ε} , one obtains pointwise or even uniform convergence $u_{\varepsilon}(x) \to u(x)$.

We observe that the FBP-type reconstruction suffers from severe artifacts when only few data are available. In contrast, our method gives a high-quality reconstruction from only 7 projections in the noise-free case (Figure 1). For noisy data, we still can recover the macro structures using our method as shown in Figure 7.

3.3. Further Applications

We finally demonstrate the applicability of our method to deblurring problems. Here, the operator A is a convolution operator. Hence, the normal equation (20) can be solved efficiently by fast Fourier transform. The experiments illustrate in particular that the method can also be applied to vector-valued data such as color images. We follow the splitting strategy proposed in the present paper and extend the univariate Potts problem to vector-valued data as found in [73]. It is worth mentioning that the computational effort grows only linearly in the dimension of the vectorial data. For example, the cost for processing a color image is about three times the cost of processing a gray-value image. We show in Figure 8 the joint reconstruction and segmentation of an image blurred by a Gaussian kernel. (The image was taken from the Berkeley Segmentation Dataset [58]). In Figure 9, we see the restoration of a traffic sign from simulated motion blur. Motion blur is modeled as a one-dimensional convolution along the direction $v \in \mathbb{R}^2$. Here, we use a moving average with respect to the horizontal direction. This experiment also illustrates that a positive couplingparameter sequence ν_k can improve the result.

4. Conclusion

In this paper, we have developed a new splitting approach for the Potts model (or piecewise-constant Mumford-Shah model) for ill-posed inverse problems in imaging. We have presented a general discretization scheme which permits near-isotropic approximations of the length terms. This discretization allowed us to split the Potts problem into specific subproblems that can be solved with efficient algorithms. We have demonstrated the capability of our method in various imaging applications. In particular, our algorithm has reconstructed the Shepp-Logan phantom from only seven projections of Radon and spherical Radon data, respectively. Further, we have obtained high-quality reconstructions of piecewise-constant images from highly



Figure 9 Restoration and segmentation of an image from a simulated horizontal motion blur of 65 pixel length with Gaussian noise of level 0.1. The letters are recovered almost perfectly (c). The result improves if we use the algorithm of (15) with $\nu_k = \mu_k/S$ instead of $\nu_k = 0$ (d).

incomplete and noisy data. Finally, we have demonstrated that is applicable for joint reconstruction and segmentation of real tomographic data.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 267439. The second and the third author acknowledge support by the Helmholtz Association within the young investigator group VH-NG-526. We would like to thank Eric Todd Quinto for valuable discussions on the topic of incomplete data tomography, Ruben Seyfried for providing us the implementation of his summability method for PAT, and Jeffrey Fessler for making his PET dataset publicly available.

References

- L. Ambrosio and V. M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via Γ-convergence. Communications on Pure and Applied Mathematics, 43(8):999– 1036, 1990.
- [2] M. Ansorg, F. Filbir, W. R. Madych, and R. Seyfried. Summability kernels for circular and spherical mean data. *Inverse Problems*, 29(1):015002, 2012.
- [3] M. Artina, M. Fornasier, and F. Solombrino. Linearly constrained nonsmooth and nonconvex minimization. SIAM Journal on Optimization, 23(3):1904–1937, 2013.
- [4] L. Bar, N. Sochen, and N. Kiryati. Variational pairing of image segmentation and blind restoration. In ECCV 2004, pages 166–177. Springer, 2004.
- [5] O. Barkan, A. Averbuch, S. Dekel, and Y. Tenzer. A mathematical model for extremely low dose adaptive computed tomography acquisition. In M. Floater, T. Lyche, M.-L. Mazure, K. Mørken, and L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces*, volume 8177 of *Lecture Notes in Computer Science*, pages 13–33. Springer, 2014.
- [6] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2(1):183–202, 2009.
- [7] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. Journal of Fourier Analysis and Applications, 14(5-6):629–654, 2008.
- [8] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. Applied and Computational Harmonic Analysis, 27(3):265–274, 2009.
- [9] E. Bostan, U. Kamilov, M. Nilchian, and M. Unser. Sparse stochastic processes and discretization of linear inverse problems. *IEEE Transactions on Image Processing*, 22(7):2699–2710, 2013.

- [10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- [11] S. Boyd and L. Vandenberghe. Convex Optimization. Cambridge University Press, 2004.
- [12] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(11):1222–1239, 2001.
- [13] L. Boysen, S. Bruns, and A. Munk. Jump estimation in inverse regression. *Electronic Journal of Statistics*, 3:1322–1359, 2009.
- [14] L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics*, 37(1):157– 183, 2009.
- [15] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. Journal of Inverse and Ill-posed Problems, 2014. To appear.
- [16] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011, 2009.
- [17] M. Bronstein, A. Bronstein, M. Zibulevsky, and H. Azhari. Reconstruction in diffraction ultrasound tomography using nonuniform FFT. *IEEE Transactions on Medical Imaging*, 21(11):1395–1401, 2002.
- [18] A. Buehler, A. Rosenthal, T. Jetzfellner, A. Dima, D. Razansky, and V. Ntziachristos. Modelbased optoacoustic inversions with incomplete projection data. *Medical Physics*, 38(3):1694, 2011.
- [19] X. Cai and G. Steidl. Multiclass segmentation by iterated ROF thresholding. In Energy Minimization Methods in Computer Vision and Pattern Recognition, pages 237–250. Springer, 2013.
- [20] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.
- [21] E. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ¹ minimization. Journal of Fourier Analysis and Applications, 14(5):877–905, 2008.
- [22] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. SIAM Journal on Applied Mathematics, 55(3):827–863, 1995.
- [23] A. Chambolle. Finite-differences discretizations of the Mumford-Shah functional. ESAIM: Mathematical Modelling and Numerical Analysis, 33(02):261–288, 1999.
- [24] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. SIAM Journal on Imaging Sciences, 5(4):1113–1158, 2012.
- [25] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
- [26] T. Chan and L. Vese. Active contours without edges. IEEE Transactions on Image Processing, 10(2):266–277, 2001.
- [27] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Processing Letters, 14(10):707–710, 2007.
- [28] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *IEEE International Symposium on Biomedical Imaging: From Nano to* Macro, pages 262–265, 2009.
- [29] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. International Journal of Computer Vision, 72(2):195–215, 2007.
- [30] M. E. Davison. The ill-conditioned nature of the limited angle tomography problem. SIAM Journal on Applied Mathematics, 43(2):428–448, 1983.
- [31] M. Defrise, C. Vanhove, and X. Liu. An algorithm for total variation regularization in highdimensional linear problems. *Inverse Problems*, 27(6):065002, 2011.
- [32] J. Fessler, E. Ficaro, N. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Transactions on Medical Imaging*, 16(2):166–175, 1997. http://web.eecs.umich.edu/~fessler/result/tr/pet,trans,2d, sino/.
- [33] F. Filbir, R. Hielscher, and W. R. Madych. Reconstruction from circular and spherical mean data. Applied and Computational Harmonic Analysis, 29(1):111–120, 2010.
- [34] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. *Preprint*, 2014.
- [35] M. Fornasier, R. March, and F. Solombrino. Existence of minimizers of the Mumford-Shah functional with singular operators and unbounded data. Annali di Matematica Pura ed

Applicata, 192(3):361-391, 2013.

- [36] M. Fornasier and R. Ward. Iterative thresholding meets free-discontinuity problems. Foundations of Computational Mathematics, 10(5):527–567, 2010.
- [37] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler. Complexity penalized M-estimation. Journal of Computational and Graphical Statistics, 17(1):201–224, 2008.
- [38] J. Frikel and E. T. Quinto. Characterization and reduction of artifacts in limited angle tomography. *Inverse Problems*, 29(12):125007, 2013.
- [39] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721-741, 1984.
- [40] B. Hahn. Reconstruction of dynamic objects with affine deformations in computerized tomography. Journal of Inverse and Ill-Posed Problems, 2014. To appear.
- [41] M. Haltmeier, T. Schuster, and O. Scherzer. Filtered backprojection for thermoacoustic computed tomography in spherical geometry. *Mathematical Methods in the Applied Sciences*, 28(16):1919–1937, 2005.
- [42] S. Helgason. The Radon Transform. Birkhäuser, Boston, 1999.
- [43] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):328–341, 2008.
- [44] M. Jiang, P. Maass, and T. Page. Regularizing properties of the Mumford-Shah functional for imaging applications. *Inverse Problems*, 30(3):035007, 2014.
- [45] P. Jonas and A. K. Louis. A Sobolev space analysis of linear regularization methods for ill-posed problems. Journal of Inverse and Ill-Posed Problems, 9(1):59–74, 2001.
- [46] A. I. Katsevich. Local tomography for the limited-angle problem. Journal of Mathematical Analysis and Applications, 213(1):160–182, 1997.
- [47] J. Kim, A. Tsai, M. Cetin, and A. S. Willsky. A curve evolution-based variational approach to simultaneous image restoration and segmentation. In *Proceedings of the IEEE International Conference on Image Processing*, volume 1, pages I–109. IEEE, 2002.
- [48] E. Klann. A Mumford-Shah-like method for limited data tomography with an application to electron tomography. SIAM Journal on Imaging Sciences, 4(4):1029–1048, 2011.
- [49] E. Klann and R. Ramlau. Regularization properties of Mumford-Shah-type functionals with perimeter and norm constraints for linear ill-posed problems. *SIAM Journal on Imaging Sciences*, 6(1):413–436, 2013.
- [50] E. Klann, R. Ramlau, and W. Ring. A Mumford-Shah level-set approach for the inversion and segmentation of SPECT/CT data. *Inverse Problems and Imaging*, 5(1):137–166, 2011.
- [51] H. Kohr and A. K. Louis. Fast and high-quality reconstruction in electron tomography based on an enhanced linear forward model. *Inverse Problems*, 27(4):045008, 2011.
- [52] V. Kolehmainen, M. Lassas, and S. Siltanen. Limited data x-ray tomography using nonlinear evolution equations. SIAM Journal on Scientific Computing, 30(3):1413–1429, 2008.
- [53] T. Kreutzmann and A. Rieder. Geometric reconstruction in bioluminescence tomography. Inverse Problems and Imaging, 8(1):173–197, 2014.
- [54] P. Kuchment and L. Kunyansky. Mathematics of photoacoustic and thermoacoustic tomography. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 817–866. Springer, 2010.
- [55] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. Scale Space and Variational Methods in Computer Vision, pages 150–162, 2009.
- [56] A. K. Louis. The limited angle problem in computerized tomography. Technical Report 5, Technische Universität Kaiserslautern, 1984.
- [57] A. K. Louis. Incomplete data problems in x-ray computerized tomography. I. Singular value decomposition of the limited angle transform. *Numerische Mathematik*, 48(3):251–262, 1986.
- [58] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the 8th International Conference on Computer Vision, pages 416–423, 2001. http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/.
- [59] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In IEEE Conference on Computer Vision and Pattern Recognition, volume 17, pages 137–154, 1985.
- [60] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.
- [61] F. Natterer. The Mathematics of Computerized Tomography. Teubner, Stuttgart, 1986.
- [62] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. SIAM

Journal on Imaging Sciences, 6(2):1035-1058, 2013.

- [63] M. Nikolova, M. Ng, and C.-P. Tam. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Transactions on Image Processing*, 19(12):3073–3088, 2010.
- [64] M. Nikolova, M. Ng, S. Zhang, and W. Ching. Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. SIAM Journal on Imaging Sciences, 1(1):2–25, 2008.
- [65] J. Nocedal and S. Wright. Numerical Optimization. Springer, New York, 2006.
- [66] T. Pock, A. Chambolle, D. Cremers, and H. Bischof. A convex relaxation approach for computing minimal partitions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 810–817. IEEE, 2009.
- [67] R. Potts. Some generalized order-disorder transformations. Mathematical Proceedings of the Cambridge Philosophical Society, 48(1):106–109, 1952.
- [68] E. T. Quinto. Singularities of the X-ray transform and limited data tomography in ℝ² and ℝ³. SIAM Journal on Mathematical Analysis, 24(5):1215–1225, 1993.
- [69] R. Ramlau and W. Ring. A Mumford–Shah level-set approach for the inversion and segmentation of X-ray tomography data. Journal of Computational Physics, 221(2):539–557, 2007.
- [70] R. Ramlau and W. Ring. Regularization of ill-posed Mumford-Shah models with perimeter penalization. *Inverse Problems*, 26(11):115001, 2010.
- [71] A. Rosenthal, V. Ntziachristos, and D. Razansky. Model-based optoacoustic inversion with arbitrary-shape detectors. *Medical Physics*, 38(7):4285, 2011.
- [72] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [73] M. Storath and A. Weinmann. Fast partitioning of vector-valued images. Accepted for SIAM Journal on Imaging Sciences, 2014. Preprint at http://bigwww.epfl.ch/preprints/ storath1402p.pdf.
- [74] M. Storath, A. Weinmann, and L. Demaret. Jump-sparse and sparse recovery using Potts functionals. *IEEE Transactions on Signal Processing*, pages 1–13, 2014. [Online] doi:10.1109/TSP.2014.2329263.
- [75] K. Wang and M. Anastasio. Photoacoustic and thermoacoustic tomography: Image formation principles. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 781–816. Springer, 2010.
- [76] L. V. Wang. Photoacoustic Imaging and Spectroscopy (Optical Science and Engineering). CRC Press, 2009.
- [77] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [78] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. Submitted, preprint arXiv:1312.7710, 2013.
- [79] A. Weinmann, M. Storath, and L. Demaret. The L¹-Potts functional for robust jump-sparse reconstruction. Submitted, preprint arXiv:1207.4642, 2012.
- [80] T. Wendler, K. Herrmann, A. Schnelzer, T. Lasser, J. Traub, O. Kutter, A. Ehlerding, K. Scheidhauer, T. Schuster, M. Kiechle, M. Schwaiger, N. Navab, S. I. Ziegler, and A. K. Buck. First demonstration of 3-D lymphatic mapping in breast cancer using freehand SPECT. *European Journal of Nuclear Medicine and Molecular Imaging*, 37(8):1452–1461, 2010.
- [81] G. Winkler. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction. Springer, Berlin, 2003.
- [82] G. Winkler and V. Liebscher. Smoothers for discontinuous signals. Journal of Nonparametric Statistics, 14(1-2):203–222, 2002.
- [83] P. A. Wolf, J. S. Jørgensen, T. G. Schmidt, and E. Y. Sidky. Few-view single photon emission computed tomography (SPECT) reconstruction based on a blurred piecewise constant object model. *Physics in Medicine and Biology*, 58(16):5629–5652, Aug. 2013.
- [84] M. Yan and L. A. Vese. Expectation maximization and total variation-based model for computed tomography reconstruction from undersampled data. In SPIE Medical Imaging, pages 79612X-79612X, 2011.