

Connecting Non-Quadratic Variational Models and MRFs

Kevin Schelten Stefan Roth

Department of Computer Science, TU Darmstadt

Abstract

Spatially-discrete Markov random fields (MRFs) and spatially-continuous variational approaches are ubiquitous in low-level vision, including image restoration, segmentation, optical flow, and stereo. Even though both families of approaches are fairly similar on an intuitive level, they are frequently seen as being technically rather distinct since they operate on different domains. In this paper we explore their connections and develop a direct, rigorous link with a particular emphasis on first-order regularizers. By representing spatially-continuous functions as linear combinations of finite elements with local support and performing explicit integration of the variational objective, we derive MRF potentials that make the resulting MRF energy equivalent to the variational energy functional. In contrast to previous attempts, we provide an explicit connection for modern non-quadratic regularizers and also integrate the data term. The established connection opens certain classes of MRFs to spatially-continuous interpretations and variational formulations to a broad range of probabilistic learning and inference algorithms.

1. Introduction

Many vision problems, particularly in low-level vision require some form of regularization and are posed as energy minimization problems. Two basic approaches dominate the literature: Spatially-continuous variational approaches¹ on one hand [21, 26], and spatially-discrete Markov random field (MRF) approaches on the other hand [4, 9]. Variational approaches originate in the calculus of variations and assume that, while camera sensors may only yield spatially-discrete measurements, the underlying quantity of interest, such as the intensity of incident light on the sensor, is ultimately spatially continuous. MRF approaches take a different route: They assume that in order to be realizable on a digital computer, a finite representation needs to be ultimately used. Hence, the desired output as well as any in-

¹We take “variational” to mean formulations based on spatially-continuous energy functionals, which is not to be confused with variational methods for approximate inference in MRFs.

termediate representation are taken to be spatially discrete. Beyond that, variational models of low-level vision are typically seen as deterministic, while MRFs admit a probabilistic view and regard the measurements and the output (at least in the Bayesian case) as inherently uncertain.

An important question is thus whether and how these approaches can be related. Superficially, both formulations are clearly similar. Yet, variational models and MRFs seem to operate on different levels: Variational approaches involve functions on the entire domain, whereas MRFs impose local constraints through cliques of pixels. The question of their relationship has been posed since the early days of regularization [27], but explicit links have been demonstrated only for a few cases [11], such as for quadratic regularizers [27]. Hence, their technical relationship has remained somewhat nebulous, and even nowadays most work in low-level vision is firmly rooted in one or the other paradigm. In this paper we show that this need not be the case.

In particular, we address the widely perceived chasm between modern first-order variational and MRF approaches. Starting from a standard variational formulation for image restoration [21], we show that if we minimize the spatially-continuous energy functional over a *finite element* (FE) space of a certain degree, an explicit connection to an equivalent spatially-discrete MRF can be established without having to approximate the spatially-continuous energy functional. The nodes in the MRF are still continuous valued, however. The degree of the underlying FE representation in the variational formulation is directly linked to the size of the cliques in the equivalent MRF. In contrast to common FE approaches to variational problems, we do not perform additional discretization steps, but instead carry out an *explicit*, and usually *analytical integration* of the domain variable. Compared to previous work [27], our approach not only applies to quadratic, but to a wide range of modern first-order regularizers. This link also demonstrates that certain MRFs have an explicit spatially-continuous interpretation, which further tightens the connection between statistical and variational approaches [22]. Moreover, this opens variational approaches to direct and explicit probabilistic interpretations and to standardized inference techniques, such as graph cuts [5], belief propagation [35], or more advanced

variants [12, 31]. As an example, we demonstrate sampling the corresponding MRF prior and posterior.

Beyond the main goal of establishing such a connection, our approach also compares favorably to the common *finite difference* (FD) method: Conceptually, our approach does not require to approximate the original spatially-continuous energy functional; rather the functional is directly evaluated and minimized over a finite dimensional function space. Practically, our method improves the performance in an image restoration application.

2. Background

Variational approaches. Variational approaches to low-level vision aim to minimize an energy functional that is comprised of a data fidelity term and a spatial term, *e.g.*:

$$\begin{aligned} \mathcal{E}(f; u) &= \mathcal{E}_D(f; u) + \lambda \mathcal{E}_S(f) & (1) \\ &= \int_{\Omega} (f - u)^2 dx + \lambda \int_{\Omega} \varphi(\nabla f) dx. & (2) \end{aligned}$$

This example is a slight generalization of the standard Rudin-Osher-Fatemi (ROF) formulation for image restoration [21]. The data fidelity term $\mathcal{E}_D(f; u)$ ensures that the restored image $f : \Omega \rightarrow \mathbb{R}$, a spatially-continuous function, is close to the noisy input $u : \Omega \rightarrow \mathbb{R}$, which is also assumed to be spatially continuous. Similar data fidelity terms can be used for other problems, such as optical flow [15]. The spatial term $\mathcal{E}_S(f)$ regularizes the problem by encouraging spatial smoothness, and depending on the choice of $\varphi(\cdot)$ allows for discontinuities; λ controls the amount of regularization. Here we take image restoration as a representative example, but note that many variational formulations for optical flow, stereo, and other problems can be treated similarly.

There is a large variety of different approaches for solving variational problems. The traditionally most popular approach is to analytically derive the *Euler-Lagrange* partial differential equations and to subsequently discretize and solve them numerically [*e.g.* 15, 21]. As noted by Pock *et al.* [16], the derivation and solution of these equations can be tedious, and the resulting algorithm may not have an interpretation in terms of energy minimization. To address this, [16] discretizes the energy functional using *finite differences* and proposes minimization algorithms via algorithmic differentiation. Such *finite difference methods* (FDM) have found widespread use [*e.g.* 6].

The third main category of solution methods, which we will rely on here, is based on finite element functions [2]. The *finite element method* (FEM) uses local, usually piecewise polynomial representations for the sought after function. Like FD discretizations, FEMs approximately solve the original problem, but do so by restricting the space of functions over which the functional is minimized rather than approximating the functional itself, which can be intuitively

viewed as a discretization of the function space. An advantage of FEMs is that different implementations can be compared in terms of the *original* spatially-continuous energy. A variety of FE basis functions have been used, including standard finite element polynomials [2], as well as B-splines [3, 18] and radial basis functions [8, 14]. While not as popular as FDMs, FEMs are also common in low-level vision. For example, [33] uses discontinuous FEs to perform image denoising while explicitly preserving edges. [30] relies on a *mixed* FEM to derive a numerically stable discretization of a variational model for fluid flow estimation. In contrast to using the FE approach solely for the purpose of solving variational problems, the focus of this paper lies on linking modern variational models to MRFs.

Markov random field approaches. Most MRF approaches to low-level vision rely on a Bayesian formulation. Again using image restoration as an example [9], the posterior of the restored image $\mathbf{o} \in \mathbb{R}^{N \times N}$ given the noisy input $\mathbf{i} \in \mathbb{R}^{N \times N}$ is formulated in terms of a likelihood and a prior:

$$p(\mathbf{o}|\mathbf{i}) \propto p(\mathbf{i}|\mathbf{o}) \cdot p(\mathbf{o}). \quad (3)$$

The likelihood $p(\mathbf{i}|\mathbf{o})$ is an observation model and intuitively takes the role of the data term from Eq. (1). The prior $p(\mathbf{o})$ models the a-priori spatial regularity and intuitively corresponds to the spatial term from before. Note, however, that input \mathbf{i} and output \mathbf{o} are now spatially discrete and represented as arrays of $N \times N$ pixels; depending on the context, we also treat them as vectors. Please note that the arrays \mathbf{i} and \mathbf{o} denote the discrete analogs of the spatially continuous functions u and f ; we use the different notation to explicitly illustrate that point.

The intuitive connection to variational models becomes even more apparent if we invoke Boltzmann’s law:

$$\begin{aligned} p(\mathbf{o}|\mathbf{i}; T) &= \frac{1}{Z(\mathbf{i}, T)} \exp \left\{ -\frac{1}{T} E(\mathbf{o}; \mathbf{i}) \right\} \\ &\propto \exp \left\{ -\frac{1}{T} (E_D(\mathbf{i}; \mathbf{o}) + E_S(\mathbf{o})) \right\}, \end{aligned} \quad (4)$$

where $E(\mathbf{o}; \mathbf{i})$ is the posterior MRF energy, and E_D and E_S are the respective MRF energies for the likelihood and the prior. The “temperature” T controls the shape of the distribution (for notational simplicity, we assume $T = 1$ in the following); the partition function $Z(\mathbf{i}, T)$ ensures normalization. The goal of this paper is to develop an explicit formal connection between the variational energy functional $\mathcal{E}(f; u)$ from Eq. (1) and the posterior MRF energy $E(\mathbf{o}; \mathbf{i})$.

The key property of MRFs, *e.g.* of the prior $p(\mathbf{o})$, is that a node (here, pixel) is conditionally independent of the rest (of the image) given its Markov blanket. By virtue of the Hammersley-Clifford theorem [4], the distributions factor-

ize into products over the cliques \mathcal{C} of the underlying graph

$$p(\mathbf{o}) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{o}_{(c)}) = \frac{1}{Z} \exp \left\{ - \sum_{c \in \mathcal{C}} \rho_c(\mathbf{o}_{(c)}) \right\}, \quad (5)$$

where $\mathbf{o}_{(c)}$ denotes all pixels belonging to clique c , $\psi_c(\cdot)$ are the factors, and $\rho_c(\cdot)$ are the corresponding potential functions. In the most frequent case of pairwise MRFs, the cliques consist of spatially neighboring pairs of pixels [9]. In high-order MRFs [e.g. 20], cliques consist of larger spatial neighborhoods of pixels.

A wide variety of different algorithms have been applied to MRF inference. Most often, the maximum a-posteriori (MAP) estimate

$$\mathbf{o}^* = \arg \max_{\mathbf{o}} p(\mathbf{o}|\mathbf{i}) = \arg \min_{\mathbf{o}} E(\mathbf{o}; \mathbf{i}) \quad (6)$$

is computed. In the continuous-valued case, gradient methods are frequently used, which again bears similarities to the variational approach². In the discrete-valued case, belief propagation [35], graph cuts [5], and more advanced variants [12, 31] have enjoyed enormous popularity in recent years. As an alternative to MAP, it is possible to compute the Bayesian minimum mean squared error (MMSE) estimate, which has been shown to outperform standard MAP estimation in the context of generatively trained models [25]. Finally, we note that a statistical MRF formulation enables learning of model parameters [20], which is paramount in complex models with many parameters.

3. Connecting Variational Models and MRFs

Previous work. The first connections between statistical models, such as MRFs, and variational problems have been developed early on. Szeliski [27] described a close relation between quadratic regularization and Gaussian MRFs using FEs. Starting from a first-order smoothness functional, the membrane model [28]

$$\mathcal{E}_S(f) = \frac{1}{2} \iint (f_{x_1}^2 + f_{x_2}^2) d\mathbf{x}, \quad (7)$$

[27] uses linear FEs to derive the discretization

$$E_S(\mathbf{o}) = \frac{1}{2} \sum_{k,l} (o_{k+1,l} - o_{k,l})^2 + (o_{k,l+1} - o_{k,l})^2, \quad (8)$$

consisting only of forward differences. For image restoration [27], moreover, proposes the data term

$$\mathcal{E}_D(\mathbf{i}; f) = \frac{1}{2} \sum_m a_m (f(x_{1,m}, x_{2,m}) - i_m)^2, \quad (9)$$

²Note that while MRFs are spatially discrete, they may well be continuous-valued, as in this paper.

which is discretized as

$$E_D(\mathbf{i}; \mathbf{o}) = \frac{1}{2} \sum_{k,l} a_{k,l} (o_{k,l} - i_{k,l})^2 \quad (10)$$

$$= \frac{1}{2} (\mathbf{o} - \mathbf{i})^T A (\mathbf{o} - \mathbf{i}).$$

The diagonal matrix $A = \text{diag}\{a_{k,l}|k,l\}$ contains optional weights. Combining the discretized data and smoothness functionals results in a quadratic form $E(\mathbf{o}) = \frac{1}{2} \mathbf{o}^T B \mathbf{o} - \mathbf{o}^T \mathbf{d} + e$. Invoking Boltzmann's law (cf. Eq. (4)) yields the correspondence of the above functionals to a Gaussian MRF. These well-studied MRFs result in an elegant framework for computing estimates as well as their uncertainty from the posterior. Beyond this case of global quadratic regularizers, Kulkarni *et al.* [11] furthermore relate MRFs and certain segmentation functionals.

While Szeliski [27] noted that non-quadratic regularizers can, in principle, be related to probability densities through Gibbs distributions, no explicit formulas for the potentials nor study of feasibility was provided. However, modern variational techniques tend to rely on non-quadratic penalties, which are robust toward outliers [1, 21]. Our key contribution is to revisit this relation and generalize it to the considerably more involved case of modern, non-quadratic regularizers, as well as to provide concise formulas for the equivalent MRF potentials.

Finite difference approximations. An approximate connection between variational models and MRFs, which quite surprisingly is rarely made, exists through finite difference discretizations. To be implemented on a computer, energy functionals have to be discretized somehow, which frequently happens using FDs [6, 16]. A standard FD discretization for the ROF model (Eq. (1) with $\varphi(\mathbf{y}) = |\mathbf{y}|$) is to use forward differences. E.g., Chambolle [6] considers

$$E_{\text{ROF}}(\mathbf{o}; \mathbf{i}) = \sum_{k,l} (o_{k,l} - i_{k,l})^2 \quad (11)$$

$$+ \lambda \sum_{k,l} \sqrt{(o_{k+1,l} - o_{k,l})^2 + (o_{k,l+1} - o_{k,l})^2}.$$

By defining appropriate prior potentials $\psi_c^S(\mathbf{o}_{(c)}) = \exp\{-\lambda \sqrt{(o_{k+1,l} - o_{k,l})^2 + (o_{k,l+1} - o_{k,l})^2}\}$ with triple cliques $c = \{(k,l), (k+1,l), (k,l+1)\}$ as well as per-pixel likelihood potentials $\psi_{k,l}^D(i_{k,l}; o_{k,l}) = \exp\{-(o_{k,l} - i_{k,l})^2\}$ an MRF whose energy is equal to Eq. (11) is given as

$$p_{\text{ROF}}(\mathbf{o}|\mathbf{i}) \propto \prod_{k,l} \psi_{k,l}^D(i_{k,l}; o_{k,l}) \cdot \prod_c \psi_c^S(\mathbf{o}_{(c)}). \quad (12)$$

Note that this MRF is not pairwise, since the prior potentials ψ_c^S involve 3 variables. The corresponding triple cliques are illustrated in Fig. 1(a). One drawback of this FD approach

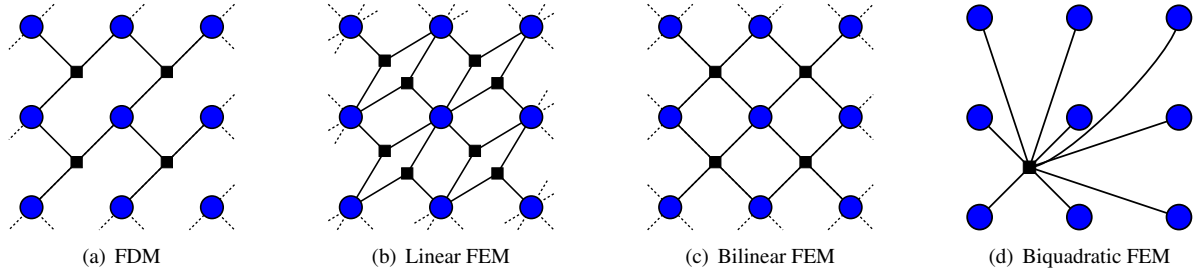


Figure 1. *Factor graph structure of variational models.* (a) Triple cliques of a standard FD discretization of the ROF model [6]; (b) Triple cliques of linear FEs; (c) 2×2 cliques of bilinear FEs; (d) 3×3 cliques of biquadratic FEs (only one clique shown for clarity).

is that different discretizations lead to different approximative functions, and thus cannot be compared in terms of the original variational objective. An implementation using finite elements with analytic integration, as is pursued here, guarantees exact energy values in the sense that the *original* energy functional is evaluated instead of an approximation thereof.

Finite element (Ritz) approach. Variational problems as in Eq. (1) are typically formulated for infinite-dimensional function spaces \mathcal{F} , e.g. the space of bounded variation [1]. To simplify the following treatment, we assume a generic variational model

$$\mathcal{E}(f; u) = \int_{\Omega} \mathcal{L}(f, u) \, d\mathbf{x} \quad \rightarrow \quad \min_{f \in \mathcal{F}}. \quad (13)$$

To be implemented in practice, either the energy needs to be approximated in a finite-dimensional way (cf. Eq. (11)), or the function space \mathcal{F} over which the functional is minimized needs to be restricted. Such a smaller, restricted space of solutions is typically defined in terms of a finite set of basis functions $b_k(\mathbf{x})$; the functions are consequently determined by their basis parameters. In contrast to typical approaches [18], we identify the basis parameters directly with the pixels, which means that the spatially-discrete input and output images, \mathbf{i} and \mathbf{o} , relate to the spatially-continuous input and output functions, f and u , as

$$u(\mathbf{x}; \mathbf{i}) = \sum_{k=1}^K i_k b_k(\mathbf{x}) \quad \text{and} \quad f(\mathbf{x}; \mathbf{o}) = \sum_{k=1}^K o_k b_k(\mathbf{x}). \quad (14)$$

In this setting, the energy functional from Eq. (13) is equivalent to a spatially-discrete energy $E(\mathbf{o}; \mathbf{i})$

$$\mathcal{E}(f; u) = \int_{\Omega} \mathcal{L}(f(\mathbf{x}; \mathbf{o}), u(\mathbf{x}; \mathbf{i})) \, d\mathbf{x} = E(\mathbf{o}; \mathbf{i}), \quad (15)$$

since the value of the functional only depends on the input and the output image. Such explicit integration is a classical approach and was described in an early work by Ritz [19].

Connection to MRFs. To complete the connection to MRFs, we opt for basis representations with local support, such as FE representations [2], which have been used in image restoration for a long time [29]³. We decompose the domain $\Omega \subset \mathbb{R}^2$ into disjoint regions Ω_k (with $\Omega = \cup_k \Omega_k$ and $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$), such that each region is overlapped with the same number of local basis functions (note that the regions Ω_k do not represent the support of the bases b_k). Then the values of the function on each region Ω_k are governed by those basis parameters that control the non-zero basis elements; we denote their indices as $c(k) \subset \{1, \dots, K\}$. Consequently, we have that

$$\mathcal{E}(f; u) = \int_{\Omega} \mathcal{L}(f(\mathbf{x}; \mathbf{o}), u(\mathbf{x}; \mathbf{i})) \, d\mathbf{x} \quad (16)$$

$$= \sum_{k=1}^K \int_{\Omega_k} \mathcal{L}(f(\mathbf{x}; \mathbf{o}), u(\mathbf{x}; \mathbf{i})) \, d\mathbf{x} \quad (17)$$

$$= \sum_{k=1}^K \int_{\Omega_k} \mathcal{L}(f(\mathbf{x}; \mathbf{o}_{c(k)}), u(\mathbf{x}; \mathbf{i}_{c(k)})) \, d\mathbf{x} \quad (18)$$

$$= \sum_{k=1}^K \rho(\mathbf{o}_{c(k)}, \mathbf{i}_{c(k)}). \quad (19)$$

This means that if we restrict ourselves to bases with local support the energy functional $\mathcal{E}(f; u)$ is equivalent to an MRF energy with cliques $\mathcal{C} = \{c(k) | k = 1, \dots, K\}$ and potential functions

$$\rho(\mathbf{o}_{c(k)}, \mathbf{i}_{c(k)}) = \int_{\Omega_k} \mathcal{L}(f(\mathbf{x}; \mathbf{o}_{c(k)}), u(\mathbf{x}; \mathbf{i}_{c(k)})) \, d\mathbf{x}, \quad (20)$$

where the cliques exert *local control* over the values on each region Ω_k . Fig. 1 illustrates the factor graph structure of the equivalent MRF for linear, bilinear, and biquadratic FE bases. Fig. 2 shows the connection between the integration region Ω_k and the respective cliques in the equivalent MRF.

³ The assumption of a basis representation with local control deserves some discussion: On one hand real optical systems pose limitations that make (piecewise) smooth functions appear very reasonable. On the other hand, variational problems with sub-linear regularizing terms have been shown to be ill-posed [1] in the space of functions with bounded variation, yet work well in practice when restricting the function space as done here.

A key characteristic of our approach lies in the rigorous realization of the *Ritz method*: We aim to eliminate the domain variable \mathbf{x} in Eq. (20) by explicit integration in closed form, rather than making discrete approximations as is frequently done [18]. We rely on FE bases of low degree for this purpose, which allow closed form integration in a number of important cases including non-quadratic regularizers. This not only yields closed form expressions for the MRF potentials in such cases, but demonstrates that a direct equivalence between large classes of variational models and MRFs can be established assuming an implementation with localized basis representations. The crux is of course, whether the integral required to eliminate the domain variable can be computed.

4. Implementation

To illustrate the connection between variational approaches and MRFs by use of the FEM in combination with explicit integration, we discuss three different implementations of the basic image restoration model from Eq. (1).

Linear finite elements. Linear FEs rely on a triangular subdivision of the domain (Fig. 2(a)), where each triangle supports a planar surface that interpolates the pixel values on three vertices. In contrast to [27], where the data term was defined in a discrete fashion and only the spatial term was integrated, we assume that both output and input are spatially continuous and carry out an analytic integration. A discretization of the function space by linear FEs and subsequent analytic integration of the squared difference data term from Eq. (1) yields a quadratic form $\frac{1}{2}(\mathbf{o} - \mathbf{i})^T A(\mathbf{o} - \mathbf{i})$, where A is a sparse, positive definite, symmetric matrix⁴. The matrix A is a non-diagonal band matrix, which reflects the pixel correlation induced by the FEM. This is in contrast to [27], where A is diagonal (*cf.* Eq. (10)). We conclude that the variational data term thus corresponds to a Gaussian MRF likelihood $p(\mathbf{i}|\mathbf{o})$ with correlated pixels, and show in Sec. 5 that this data term leads to increased performance.

Linear FEs are particularly advantageous since any type of first-order spatial term (with an arbitrary penalty $\varphi(\cdot)$) becomes analytically integrable. This is because the partial derivatives on the triangle elements are constant. As shown in [24], explicit integration yields

$$E_S(\mathbf{o}) = \frac{1}{2} \sum_{i,j < N} \varphi((o_{i+1,j} - o_{i,j}, o_{i,j+1} - o_{i,j})^T) + \frac{1}{2} \sum_{i,j > 1} \varphi((o_{i,j} - o_{i-1,j}, o_{i,j} - o_{i,j-1})^T), \quad (21)$$

which may seem unorthodox compared to standard FDs, since the corresponding MRF energy contains a combination of forward and backward differences. Note that

⁴This and the following are derived in detail in [24].

this does not become apparent in the previously considered quadratic case (*cf.* Eq. (8)). It is also important to note that in contrast to Eq. (8) the non-quadratic case results in high-order cliques; the triple cliques of the corresponding MRF prior are shown in Fig. 1(b). These properties are a rigorous consequence of FE calculus, and the experimental results in Sec. 5 further indicate a well-founded effect.

Since the main focus of this work is on bridging variational and MRF approaches to low-level vision, we focus on regularizers common in the context of variational models. Due to its convexity, the total variation (TV) regularizer enjoys particular popularity [1, 6, 16, 21]. We therefore choose it as main model for our implementation. The TV model penalizes discontinuities by use of the Euclidean norm $\varphi(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$. To cope with the non-differentiability at $(0, 0)^T$, a common relaxation is the function $\sqrt{x_1^2 + x_2^2 + \epsilon^2}$ [7]. We will refer to this TV relaxation as TV + ϵ . Furthermore, we also consider an anisotropic variant of TV (+ ϵ), in which each partial derivative is penalized separately.

If we integrate TV + ϵ using linear FEs as in Eq. (21), the MAP estimate of the resulting MRF is readily obtained by gradient descent. In the non-differentiable case of TV, an effective method for inference is much less obvious. We opt for an adaption of Chambolle's fast duality-based algorithm [6] to the FEM case: First, the spatial term consisting of forward and backward differences gives rise to a discrete gradient operator $\nabla : X \rightarrow Y$ for spaces of matrices $X = \mathbb{R}^{N \times N}$ and $Y = \mathbb{R}^{N \times 2N} \times \mathbb{R}^{N \times 2N}$. Intuitively, the first N columns of Y accommodate the forward differences of ∇ and the second N the backward differences. The divergence operator $\text{div} : Y \rightarrow X$ is defined as the negative adjoint of the gradient operator (the exact formulas for the gradient and divergence operators are cumbersome but straightforward; see [24] for details). Furthermore, the FEM-induced data term leads to the Euler equation

$$\mathbf{0} \in A(\mathbf{o} - \mathbf{i}) + \lambda \partial E_S(\mathbf{o}), \quad (22)$$

where $E_S(\mathbf{o})$ is Eq. (21) for $\varphi(\mathbf{x}) = |\mathbf{x}|$, and ∂E_S is the subdifferential of E_S . By introducing the relation $\mathbf{v} = \frac{1}{\lambda} A(\mathbf{i} - \mathbf{o})$, the solution can be obtained from minimizing

$$\mathbf{v}^T A^{-1} \mathbf{v} - \mathbf{v}^T \left(\frac{2\mathbf{i}}{\lambda} \right) + \frac{2}{\lambda} E_S^*(\mathbf{v}), \quad (23)$$

where E_S^* denotes the *Fenchel* conjugate of E_S . The characteristic function property of E_S^* permits the solution of Eq. (23) by optimization of the quadratic objective

$$\mathbf{v}^T A^{-1} \mathbf{v} - \mathbf{v}^T \left(\frac{2\mathbf{i}}{\lambda} \right) \quad \text{subject to} \quad (24)$$

$$\mathbf{v} \in \{\text{div}(\mathbf{p}) : \mathbf{p} \in Y, |\mathbf{p}_{i,j}| \leq 1\}. \quad (25)$$

A fixed point algorithm for the completion of this task can be derived via the *Karush-Kuhn-Tucker* conditions.

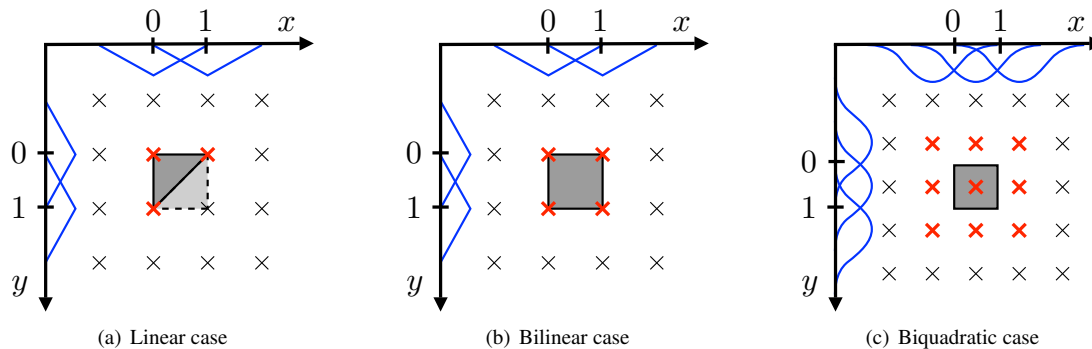


Figure 2. FE implementation with local support and MRF cliques. One local integration area Ω_k is shaded. The relevant control points, which also form one of the cliques of the equivalent high-order MRF are shown in bold red. Localized basis functions are shown in blue.

Bilinear finite elements. Beyond linear FEs, we can also use bilinear FEs with square integration regions (cf. Fig. 2(b)). A bilinear dependence on the domain variables relies on the patch-wise representation $f(\mathbf{x}; \mathbf{o}) = \sum_{k,j=0}^1 c_{k,j} x_1^k x_2^j$, where the $c_{k,j}$ are linear combinations of the 4 adjacent pixels $\mathbf{o}_{(c)}$, which exert local control and give rise to 2×2 cliques in the equivalent MRF (cf. Fig. 1(c)). Bilinear finite elements enable exact integration in the case of anisotropic TV + ϵ . Explicit integration of the variational model leads to infinitely differentiable MRF potential functions [24]. An anisotropic model with the non-convex Lorentzian penalty $\log(1 + x^2)$ [23] can also be explicitly integrated to yield an equivalent, smooth MRF.

Biquadratic tensor product B-splines. Higher-degree polynomial representations are also possible. Biquadratic tensor product B-splines have the patch-wise form $f(\mathbf{x}; \mathbf{o}) = \sum_{k,j=0}^2 c_{k,j} x_1^k x_2^j$, where the coefficients $c_{k,j}$ are linear combinations of the 3×3 clique of pixels $\mathbf{o}_{(c)}$, which exert local control (cf. Figs. 1(d) and 2(c)). For this choice of FEs, the anisotropic TV + ϵ model can be integrated semi-numerically, *i.e.* exactly in one dimension and approximately in the second. This approximation is mathematically rigorous and converges to the true integral. For each step size of quadrature, the potential functions of the corresponding MRF are smooth and allow inference by gradient descent [24]. This derivation is also feasible for the anisotropic Lorentzian model.

5. Experimentation

To demonstrate that the derived explicit connection between non-quadratic variational approaches and MRFs is not only of theoretical interest, but also provides practical advantages, we perform a series of image denoising experiments comparing standard finite difference discretizations and the MRFs obtained from explicitly integrated FEs. In both cases, the MRFs are derived from first-order variational models, however in case of FDs the equivalence is only approximate, whereas it is exact for our FE ap-

proach. As is typical in the literature [e.g. 17], we assume i.i.d. Gaussian noise with known variance. We use and compare two regularizers: (1) the differentiable TV + ϵ penalty ($\epsilon = 0.01$, except in case of bilinear FEs where $\epsilon = 1$ is required for numerical stability), and (2) the non-differentiable TV penalty. Unless otherwise mentioned, we use an isotropic (*i.e.* gradient-magnitude based) regularizer. Restoration performance is evaluated using both peak signal-to-noise ratio (PSNR) and the more advanced, perceptually motivated structural similarity index (SSIM) [32]. For each permutation of method, noise level and metric, the regularization parameter λ was determined using an exhaustive grid search on 20 separate training images by maximizing the performance of the respective metric. In case of differentiable penalties, we compute the MAP estimate of the posterior MRF using conjugate gradients [CG]. For the MRFs derived using the non-differentiable TV model, we rely on the popular duality-based method [DM] of Chambolle [6], or a variation thereof (cf. Sec. 4). Note that for the convex penalties considered here, applying inference algorithms such as belief propagation or graph cuts will likely not be beneficial. Nonetheless, the connection between variational approaches and MRFs developed here allows to apply such inference techniques to variational problems with non-convex regularizers, or to move beyond MAP estimation. We leave this for future work.

We evaluate the average image restoration performance for a variety of models and methods on 68 images (as used, *e.g.*, in [20]). Tbl. 1 summarizes the results; Fig. 3 gives a visual example. We make three main observations: (1) When comparing the same variational model, the proposed FE-MRF performs at least as well as the FD-MRF and in most cases substantially better (approx. 0.1dB). (2) The TV + ϵ penalty leads to consistently better results than standard TV regularization, which is in line with findings in the context of optical flow [34], likely because staircasing is avoided by a near-quadratic regularization near zero. The performance of the bilinear FE-MRF versus a corresponding FD-MRF is mixed, which may also be due to an ob-

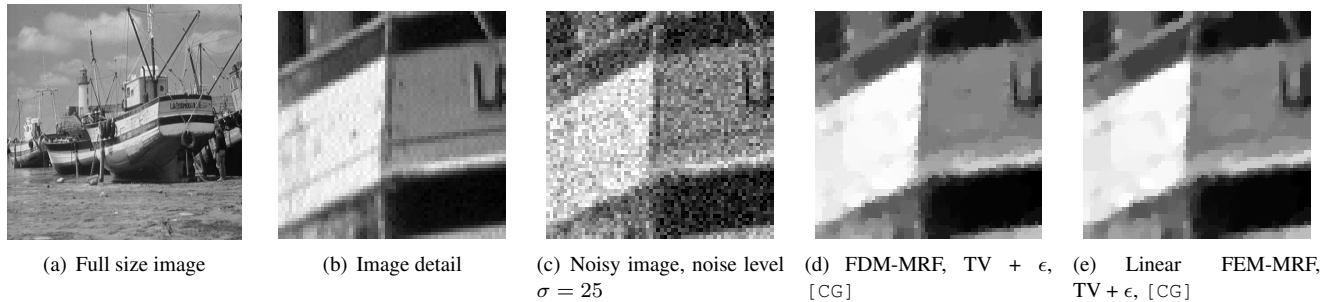


Figure 3. *Image denoising with a non-quadratic variational model and the corresponding MRF realizations.* The FEM-MRF result exhibits fewer speckle artifacts (best viewed on screen).

served numerical instability of the former. (3) By separating the contribution of the FE spatial term and the FE data term, we find that each improves upon the pure FD-MRF. The correlated data term contributes the strongest. For comparison, we also report results for two competitive denoising approaches [17, 20], for which code is publicly available. While we note that the standard first-order variational model does not perform at the level of these more sophisticated techniques (or more recent ones [10, 13]), it is simpler and enjoys widespread use to date.

One advantage of the derived equivalence of variational models and MRFs is that we can apply probabilistic methods to variational models. Of the many possibilities, we here consider sampling. Fig. 4 demonstrates standard Metropolis sampling from the probabilistic model induced by a variational formulation with a $TV + \epsilon$ regularizer and using the proposed linear FE-MRF. Samples from the posterior can, for example, be used to estimate the posterior mean (MMSE estimate) and the marginal variance (*cf.* Fig. 4(c),(d)). In the future, posterior samples may also be used to infer MAP estimates of non-convex regularizers via simulated annealing. Fig. 4(e)-(h) shows how sampling the equivalent MRF can also give insights into the generative properties of the underlying variational model.

6. Conclusions

In this paper we investigated the connections of modern, non-quadratic first-order variational models and Markov random fields. Based on finite elements and explicit integration, we derived localized potential functions of an equivalent MRF, which provided a rigorous connection between the two approaches. We demonstrated the feasibility of the approach with several examples, and adapted a duality-based inference algorithm to MAP inference in MRFs derived from FE implementations of TV-based variational models. Moreover, we gave an experimental analysis of the derived connection in an image denoising application, where we found improved performance compared to standard discretization schemes for variational models. Finally, we illustrated the connection to probabilistic models

by sampling the equivalent MRF prior and posterior. Future work should consider applying probabilistic learning and inference to the MRFs equivalent to the variational formulation, such as learning the penalty function, performing MAP estimation for non-convex penalties, or carrying out MMSE estimation, *e.g.* by variational inference. In addition, the established connection allows to investigate the generative properties of variational models (*cf.* [25]). Future work may also consider second-order priors and other problem domains, such as correspondence problems or segmentation.

Acknowledgments: We wish to thank Qi Gao, Arjan Kuijper, Bernt Schiele, Konrad Schindler and Uwe Schmidt for discussions.

References

- [1] G. Aubert, A. E. Hamidi, C. Ghannam, and M. Ménéard. On a class of ill-posed minimization problems in image processing. *J. Math. Anal. Appl.*, 352(1):380–399, Apr. 2009.
- [2] K.-J. Bathe. *Finite Element Procedures*. Prentice Hall, 1995.
- [3] O. Bernard, D. Friboulet, P. Thévenaz, and M. Unser. Variational B-spline level-set method for fast image segmentation. In *ISBI*, pages 177–180, 2008.
- [4] J. Besag. Spatial interaction and the statistical analysis of lattices. *J. Roy. Stat. Soc. B*, 36(2):192–236, 1974.
- [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, Nov. 2001.
- [6] A. Chambolle. An algorithm for total variation minimization and applications. *JMIV*, 20(1–2):89–97, Jan. 2004.
- [7] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE TIP*, 6(2):298–311, Feb. 1997.
- [8] A. Gelas, O. Bernard, D. Friboulet, and R. Prost. Compactly supported radial basis functions based collocation method for level-set evolution in image segmentation. *IEEE TIP*, 16(7):1873–1887, 2007.
- [9] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *PAMI*, 6:721–741, Nov. 1984.
- [10] V. Jain and H. S. Seung. Natural image denoising with convolutional networks. In *NIPS*2008*.
- [11] S. R. Kulkarni, S. K. Mitter, T. J. Richardson, and J. N. Tsitsiklis. Local versus nonlocal computation of length of digitized curves. *PAMI*, 16(7):711–718, July 1994.
- [12] M. P. Kumar, V. Kolmogorov, and P. H. S. Torr. An analysis of convex relaxations for MAP estimation of discrete MRFs. *JMLR*, 10:71–106, Jan. 2009.

Table 1. Average restoration performance on 68 test images (from [20]).

MRF / variational model / algorithm	$\sigma = 15$		$\sigma = 25$	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
FD-MRF, TV, [DM]	29.63	.829	27.08	.747
Linear FE-MRF, TV, [DM]	29.63	.832	27.19	.750
FD-MRF, TV + ϵ , [CG]	29.69	.833	27.22	.751
FD-MRF w/ FE data term, TV + ϵ , [CG]	29.75	.836	27.29	.756
FD-MRF w/ FE spatial term, TV + ϵ , [CG]	29.72	.834	27.24	.754
Linear FE-MRF, TV + ϵ , [CG]	29.80	.837	27.32	.757
FD-MRF, anisotropic TV + ϵ , [CG]	29.57	.829	27.12	.747
Bilinear FE-MRF, anisotropic TV + ϵ , [CG]	29.73	.826	27.11	.751
FoE [20]			27.44	.746
BLS-GSM [17]			28.02	.789

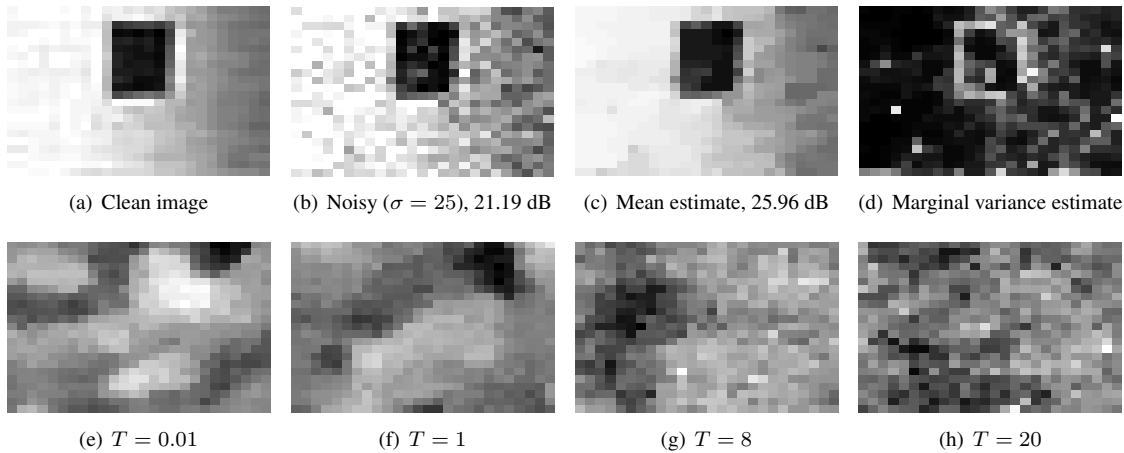


Figure 4. *Metropolis sampling* (Gaussian proposals, 100.000 iterations burn in) for the linear FEM, TV + ϵ model: (*top*) First and second order posterior estimates from samples at $T = 1$; (*bottom*) Sampling the prior at different temperatures T .

- [13] S. Lyu and E. P. Simoncelli. Statistical modeling of images with fields of Gaussian scale mixtures. In *NIPS*2006*.
- [14] B. S. Morse, W. Liu, T. S. Yoo, and K. Subramanian. Active contours using a constraint-based implicit representation. In *CVPR 2005*, volume 1, pages 285–292.
- [15] N. Papenberger, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *IJCV*, 67(2):141–158, Apr. 2006.
- [16] T. Pock, M. Pock, and H. Bischof. Algorithmic differentiation: Application to variational problems in computer vision. *PAMI*, 29(7):1180–1193, July 2007.
- [17] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE TIP*, 12(11):1338–1351, Nov. 2003.
- [18] S. Ramani, P. Thévenaz, and M. Unser. Regularized interpolation for noisy data. In *ISBI*, pages 612–615, 2007.
- [19] W. Ritz. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. *Journal für die reine und angewandte Mathematik*, 135:1–61, 1909.
- [20] S. Roth and M. J. Black. Fields of experts: A framework for learning image priors. In *CVPR 2005*, volume 2, pages 860–867.
- [21] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, Nov. 1992.
- [22] H. Schar, M. J. Black, and H. W. Haussecker. Image statistics and anisotropic diffusion. In *ICCV 2003*, volume 2, pages 840–847.
- [23] H. Schar, M. J. Black, and H. W. Haussecker. Anisotropic diffusion using algorithm-specific image statistics. Journal draft, never submitted, May 2003.
- [24] K. Schelten and S. Roth. Connecting non-quadratic variational models and MRFs. Technical Report TUD-CS-2011-0103, TU Darmstadt, Department of Computer Science, 2011.
- [25] U. Schmidt, Q. Gao, and S. Roth. A generative perspective on MRFs in low-level vision. In *CVPR 2010*.
- [26] C. Schnörr, R. Sprengel, and B. Neumann. A variational approach to the design of early vision algorithms. *Computing Supplement*, 11:149–165, 1996.
- [27] R. Szeliski. Bayesian modeling of uncertainty in low-level vision. *IJCV*, 5(3):271–301, Dec. 1990.
- [28] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *PAMI*, 8(4):413–424, July 1986.
- [29] D. Terzopoulos. The computation of visible-surface representations. *PAMI*, 10(4):417–438, July 1988.
- [30] A. Vlasenko and C. Schnörr. Superresolution and denoising of 3d fluid flow estimates. In *Pattern Recognition (DAGM) 2009*, pages 482–491.
- [31] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. MAP estimation via agreement on (hyper)trees: Message passing and linear programming approaches. *IEEE T. Info. Theory*, 51(11):3697–3717, Nov. 2005.
- [32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, Apr. 2004.
- [33] Z. Wang, F. Qi, and F. Zhou. A discontinuous finite element method for image denoising. *Image Analysis and Recognition*, pages 116–125, 2006.
- [34] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic Huber-L1 optical flow. In *BMVC 2009*.
- [35] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In *Exploring Artificial Intelligence in the New Millennium*, chapter 8, pages 239–236. 2003.