Chapter2

Total Variation Image Restoration: Overview and Recent Developments

T. Chan, S. Esedoglu, F. Park and A. Yip

Abstract

Since their introduction in a classic paper by Rudin, Osher and Fatemi [695], total variation minimizing models have become one of the most popular and successful methodology for image restoration. More recently, there has been a resurgence of interest and exciting new developments, some extending the applicabilities to inpainting, blind deconvolution and vector-valued images, while others offer improvements in better preservation of contrast, geometry and textures, in ameliorating the staircasing effect, and in exploiting the multiscale nature of the models. In addition, new computational methods have been proposed with improved computational speed and robustness. We shall review some of these recent developments.

2.1 Introduction

Variational models have been extremely successful in a wide variety of restoration problems, and remain one of the most active areas of research in mathematical image processing and computer vision. By now, their scope encompasses not only the fundamental problem of image denoising, but also other restoration tasks such as deblurring, blind deconvolution, and inpainting. Variational models exhibit the solution of these problems as minimizers of appropriately chosen functionals. The minimization technique of choice for such models routinely involves the solution of nonlinear partial differential equations (PDEs) derived as necessary optimality conditions.

Perhaps the most basic (fundamental) image restoration problem is denoising. It forms a significant preliminary step in many machine vision tasks, such as object detection and recognition. It is also one of the mathematically most intriguing problems in vision. A major concern in designing image denoising models is to preserve important image features, such as those most easily detected by the human visual system, while removing noise. One such important image feature are the edges; these are places in an image where there is a sharp change in image properties, which happens for instance at object boundaries. A great deal of research has gone into designing models for removing noise while preserving edges; recently there has also been a lot of effort in preserving other fine scale image features, such as texture. All successful denoising models take advantage of the fact that there is an inherent regularity found in natural images; this is how they attempt to tell apart noise and actual image information. Variational and PDE based models make it particularly easy to impose geometric regularity on the solutions obtained as denoised images, such as smoothness of boundaries. This is one of the main reasons behind their success.

Total variation based image restoration models were first introduced by Rudin, Osher, and Fatemi (ROF) in their pioneering work [695] on edge preserving image denoising. It is one of the earliest and best known examples of PDE based edge preserving denoising. It was designed with the explicit goal of preserving sharp discontinuities (edges) in images while removing noise and other unwanted fine scale detail. Being convex, the ROF model is one of the simplest variational models having this most desirable property. The revolutionary aspect of this model is its regularization term that allows for discontinuities but at the same time disfavors oscillations. It was originally formulated in [695] for grayscale imagery in the following form:

$$\inf_{\int_\Omega (u-f)^2\,dx=\sigma^2} \int_\Omega |\nabla u|. \tag{2.1}$$

Here, Ω denotes the image domain (for instance, the computer screen), and is usually a rectangle. The function $f(x) : \Omega \longrightarrow \mathbb{R}$ represents the given observed image, which is assumed to be corrupted by Gaussian noise of variance σ^2. The constraint of the optimization forces the minimization to take place over images that are consistent with this known noise level. The objective functional itself is called the *total variation* (TV) of the function $u(x)$; for smooth images it is equivalent to the L^1 norm of the derivative, and hence is some measure of the amount of oscillation found in the function $u(x)$. Optimization problem (2.1) is equivalent to the following *unconstrained* optimization, which was also first introduced in [695]:

$$\inf_{u\in L^2(\Omega)} \int_\Omega |\nabla u| + \lambda \int_\Omega (u-f)^2\,dx. \tag{2.2}$$

Here, $\lambda \geq 0$ is a Lagrange multiplier. The equivalence of problems (2.1) and (2.2) has been established in [162]. In the original ROF paper [695] there is an iterative numerical procedure given for choosing λ so that the solution $u(x)$ obtained solves (2.1).

We point out that total variation based energies appear, and have been previously studied in, many different areas of pure and applied mathematics. For instance, the notion of total variation of a function and functions of bounded variation appear in the theory of minimal surfaces. In applied mathematics, total variation based models and analysis appear in more classical applications such as elasticity and fluid dynamics. Due to ROF, this notion has now become central also in image processing.

Over the years, the ROF model has been extended to many other image restoration tasks, and has been modified in a variety of ways to improve its performance. In this article, we will concentrate on some recent developments in total variation based image restoration research. Some of these developments have led to new algorithms, and others to new models and theory. While we try to be comprehensive, we are of course limited to those topics and works that are of interest to us, and that we are familiar with. In particular, we aim to provide highlights of a number of new ideas that include the use of different norms in measuring fidelity, applications to new image processing tasks such as inpainting, and so on. We also hope that this article can serve as a guide to recent literature on some of these developments.

2.2 Properties and Extensions

2.2.1 BV Space and Basic Properties

The space of functions with bounded variation (BV) is an ideal choice for minimizers to the ROF model since BV provides regularity of solutions but also allows sharp discontinuities (edges). Many other spaces like the Sobolev space $W^{1,1}$ do not allow edges. Before defining the space BV, we formally state the definition of TV as:

$$\int_\Omega |\nabla f| = \sup\left\{\int_\Omega f\nabla\cdot \mathbf{g}dx \mid \mathbf{g}\in C_c^1(\Omega,\mathbb{R}^n), |\mathbf{g}(\mathbf{x})|\le 1 \forall\, \mathbf{x}\in\Omega\right\} \tag{2.3}$$

where $f \in L^1(\Omega)$ and $\Omega \subseteq \mathbb{R}^n$ is a bounded open set. We can now define the space BV as $\{f \in L^1(\Omega) \mid \int_\Omega |\nabla f| < \infty\}$. Thus, BV functions amount to L^1 functions with bounded TV semi-norm. Moreover, through the TV semi-norm there is a natural link between BV and the ROF model.

Given the choice of $BV(\Omega)$ as the appropriate space for minimizers of the ROF model (2.2), there are the basic properties of existence and uniqueness to settle. The ROF model in unconstrained form (2.2) is a strictly convex functional, hence, admits a unique minimum. Moreover, it is shown in [162] that the equality constraint $\int_\Omega (u - f)^2 d\mathbf{x} = \sigma^2$ in the non-convex ROF model (2.1) is equivalent to the convex inequality constraint $\int_\Omega (u - f)^2 d\mathbf{x} \le \sigma^2$. Hence, the non-convex minimization in (2.1) is equivalent to a convex minimization problem which under some additional assumptions is further equivalent to the above unconstrained minimization (2.2).

For BV functions there is a useful coarea formulation linking the total variation to the level sets giving some insight into the behavior of the TV norm. Given a function $f \in BV(\Omega)$ and $\gamma \in \mathbb{R}$, denote by $\{f = \gamma\}$ the set:

$\{\mathbf{x} \in \mathbb{R}^2 \mid f(\mathbf{x}) = \gamma\}$. Then, if f is regular, the TV of f can be given by:

$$\int_\Omega |\nabla f| = \int_{-\infty}^{\infty} \int_{\{f=\gamma\}} ds\, d\gamma. \tag{2.4}$$

Here, the term $\int_{\{f=\gamma\}} ds$ represents the length of the set $\{f = \gamma\}$. The formula states that the TV norm of f can be obtained by integrating along all contours of $\{f = \gamma\}$ for all values of γ. Thus, one can view TV as controlling both the size of the jumps in an image and the geometry of the level sets.

2.2.2 *Multi-channel TV*

Total variation based models can be extended to vector valued images in various ways.

An interesting generalization of TV denoising to vector valued images was proposed by Sapiro and Ringach [704]. The idea is to think of the image $u : \mathbb{R}^2 \longrightarrow \mathbb{R}^m$ as a parametrized two dimensional surface in $\mathbb{R}^m$, and to use the difference between eigenvalues of the first fundamental form as a measure of edge strength. A variational model results from integrating the square root of the magnitude of this difference as the regularization term.

Blomgren and Chan [98] generalized total variation regularization to vectorial data as the Euclidean norm of the vector of (scalar) total variations of the components. This generalization has the benefit that vector valued images defined on the line whose components are monotone functions with identical boundary conditions all have the same energy, regardless of their smoothness. This implies good edge preserving properties.

Another interesting approach generalizing edge preserving variational denoising models to vector valued images is due to Kimmel, Malladi, and Sochen [473]. They regard the given image $u(x) : \mathbb{R}^2 \longrightarrow \mathbb{R}^m$ as a surface in $\mathbb{R}^{m+2}$, and propose an area minimizing flow (which they call Beltrami flow) as a means of denoising it.

2.2.3 *Scale*

The constant λ that appears in the ROF model plays the role of a "scale parameter". By tweaking λ, a user can select the level of detail desired in the reconstructed image. In this sense, λ in (2.2) is analogous to the time variable in scale space theories for nonlinear diffusion based denoising models. The geometric interpretation of the regularization term in (2.2) given by the co-area formula suggests that λ determines which image features are kept based on, roughly speaking, their "perimeter to area" ratio.

The intuitive link between λ and scale of image features can be exactly verified in the case of an image that consists of a white disk on a black background. Strong and Chan [770] determined the solution of the ROF functional for such a given image $f(x)$. It turns out to be $(1 - \frac{1}{\lambda r})f(x)$ for $\lambda > \frac{1}{r}$. In particular, there is always a loss of contrast in the reconstruction, no matter how large the fidelity constant λ is. And when $\lambda \leq \frac{1}{r}$, the solution is identically 0, meaning that the model prefers to remove disks of radius less than $\frac{1}{\lambda}$. This simple but instructive example indicates how to relate the parameter λ to the scale of objects we desire to preserve in reconstructions. Strong and Chan's observation has been generalized to other exact solutions of the ROF model in [69].

The parameter λ can thus be used for performing multiscale decomposition of images: Image features at different scales are separated by minimizing the ROF energy using different values of λ. Recent research along these lines is described in section 2.5.3.

2.3 Caveats

While using TV-norm as regularization can reduce oscillations and regularize the geometry of level sets without penalizing discontinuities, it possesses some properties which may be undesirable under some circumstances.

Loss of contrast. The total variation of a function, defined on a bounded domain, is decreased if we re-scale it around its mean value in such a way that the difference between the maximum and minimum value (contrast) is reduced. In [770, 567], the authors showed that for any non-trivial regularization parameter, the solution to the ROF model has a contrast loss. The example of a white disk with radius R over a black background discussed in 2.2.2 is a simple illustration. In this case, the contrast loss is inversely proportional to $f(x)/r$ before the disk merges with the background. In general, reduction of the contrast of a feature by $h > 0$ would induce a decrease in the regularization term of the ROF model by $O(h)$ and an increase in the fidelity term by $O(h^2)$ only. Such scalings of the regularization and fidelity terms favors the reduction of the contrast.

Loss of geometry. The co-area formula (2.4) reveals that, in addition to loss of contrast, the TV of a function may be decreased by reducing the length of each level set. In some cases, such a property of the TV-norm may lead to distortion of the geometry of level sets when applying the ROF model. In [770], Strong and Chan show that, for circular image features, their shape is preserved at least for a small change in the regularization parameter and their location is also preserved even they are corrupted by noise of moderate level. In [69], Bellettini et al. extend Strong and Chan's results and show that the set of all bounded connected shapes C that are shape-invariant in the solution of the ROF model is precisely given by

$$\left\{ C \subset \mathbb{R}^N : C \text{ convex}, \partial C \in C^{1,1} \text{ and ess} \sup_{p \in \partial C} \kappa_{\partial C}(p) \leq |\partial C|/|C| \right\}.$$

Here, $|\partial C|$ is the perimeter of C, $|C|$ is the area of C and $\kappa_{\partial C}(p)$ is the curvature of ∂C at p. The downside of the above characterization is that the ROF model distorts the geometry of shapes that do not belong to the shape-invariant set. For instance, it has been shown in [567], if the input image is a rectangle R over a background with a different intensity, then cutting a corner (an isosceles triangle) with height h of the rectangle would induce a reduction in the TV-norm by $O(h)$ and an increment of the fitting term by $O(h^2)$, thus favoring cutting the corners.

Staircasing. This refers to the phenomenon that the denoised image may look blocky (piecewise constant). In the 1-D discrete case, there is a simple explanation to this — the preservation of monotonicity of neighboring values. Such a property requires that, for each i, if the input $f = \{f_i\}$ satisfies $f_i \leq f_{i+1}$ (resp. $\geq$), then the output must satisfy $u_i \leq u_{i+1}$ (resp. $\geq$) for any λ. In the case where f satisfies $f_{i_0-1} < f_{i_0} > f_{i_0+1} < f_{i_0+2}$ for some i_0, which often happens when the true signal is monotonically increasing around i_0 and is corrupted by noise but u satisfies $u_{i_0-1} < u_{i_0} = u_{i_0+1} < u_{i_0+2}$, then, visually, u looks like a staircase at i_0 but a monotonically increasing signal is more desirable. In the 2-D case, the monotonicity preserving property is no longer true in general, for instance, near corners of image features. However, away from the corners where the curvature of the level sets is high, staircase is often observed.

Loss of Texture. Although highly effective for denoising, the TV norm cannot preserve delicate small scale features like texture. This can be accounted for from a combination of the above mentioned geometry and contrast loss caveats of the ROF model which have the tendency to affect small scale features most severely.

2.4 Variants

Total variation based image reconstruction models have been extended in a variety of ways. Many of these are modifications of the original ROF functional (2.2), addressing the above mentioned caveats.

2.4.1 Iterated Refinement

A very interesting and innovative new perspective on the standard ROF model has been recently proposed by Osher et al. [615]. The new framework involved can be generalized to many convex reconstruction models (inverse problems) beyond TV based denoising. When applied to the ROF model in particular, this new approach fixes a number of its caveats, such as loss of contrast, and promises even further improvements in other significant aspects of reconstruction, such as preservation of textures.

The key idea is to compensate for the loss of signal in reconstructed images by minimizing the ROF model repeatedly, each time adding back the signal removed in the previous iteration. Thus, starting with a given $f_0(x) := f(x)$, repeat for $j = 1, 2, 3, \ldots$:

1. Set $u_j(x) = \text{argmin}_u$ of (2.2) using $f_j(x)$ as the given image.
2. Set $f_{j+1}(x) = f_j(x) + \big(f - u_j(x)\big)$.

When applied to the characteristic function of a disk, this algorithm recovers it perfectly after a finite number of iterations without loss of contrast.

The algorithm can be generalized to inverse problems of the form $\inf_u J(u) + H(u, f)$. Here, J is a convex regularization term, and $H(u, f)$ a fidelity term that is required to be convex in u for every f. In this setting, the iterative procedure above becomes: Start with $u_0 = 0$, repeat for $j = 1, 2, 3, \dots$

$$u_{j+1} = \arg\min_{w} H(w, f) + J(w) - J(u_j) - \langle D_u J(u_j), w - u_j \rangle. \tag{2.5}$$

Here, $D_u J(u_j)$ denotes the derivative of the functional J at the j-th iterate u_j, and $\langle \cdot, \cdot \rangle$ represents the duality pairing. If J is non-differentiable (as in the ROF model), then $D_u J(u_j)$ needs to be understood as an element of the subgradient $\partial J(u_j)$ of J at u_j. It is clear from formula (2.5) that the algorithm involves removing from the regularization term $J(u)$ its linearization at the current iterate u_j.

Formula (2.5) suggests the following definition: For $p \in \partial J(v)$, let

$$D^p(u, v) := J(u) - J(v) - \langle p, u - v \rangle$$

be the generalized *Bregman distance* associated with the functional J. It defines a notion of distance between two functions u and v because it satisfies the conditions $D^p(u, v) \geq 0$ for all u, v, and $D^p(u, u) = 0$. However, it is not a metric as it needs not be symmetric or satisfy a triangle inequality.

A number of important general theorems have been established in [615], including:

- As long as the distance of the reconstructed image u_j to the given noisy $f(x)$ remains greater than σ (the noise variance), the iteration decreases the Bregman distance of the iterates u_j to the *true* (i.e. noise-free) image.
- $H(u_j, f)$ decreases monotonically and tends to 0 as $j \longrightarrow \infty$.

In [615], further results can be found about the convergence rate of the iterates u_j to the given image f under certain regularity assumptions on f.

2.4.2 L^1 Fitting

A simple way to modify the ROF model in order to compensate for the loss of contrast is to replace the squared L^2 norm in the fidelity term in (2.2) by the L^1 norm instead. The resulting energy is

$$\int_\Omega |\nabla u| + \lambda \int_\Omega |u - f| \, dx. \tag{2.6}$$

Discrete versions of this model were studied for one dimensional signals by Alliney [14], and in higher dimensions by Nikolova [602]. In particular, it has been shown to be more effective that the standard ROF model in the presence of certain types of noise, such as salt and pepper. Recently, it has been studied in the continuous setting by Chan and Esedoglu [165].

Although the modification involved in (2.6) seems minor, it has certain desirable consequences. First and foremost, the scaling between the two terms of (2.6) is different from the one in the original ROF model (2.2), and leads to contrast invariance: If $u(x)$ is the solution of (2.6) with $f(x)$ as the given image, then $cu(x)$ is the solution of (2.6) with $cf(x)$ as the given image. This property does not hold for (2.2). A related consequence is: If the given image $f(x)$ is the characteristic function of a set Ω with smooth boundary, then the image is perfectly recovered by model (2.6) for large enough choices of the parameter λ. This is in contrast to the behavior of the ROF model, which always prefers to remove some of the original signal from the reconstructed one, and preserves a very small class of shapes. This statement can be generalized beyond original images given by characteristic functions of sets to show that a wide class of regular images are left unmodified by model (2.6) for large enough choices of the parameter λ.

In addition to having better contrast preservation properties, model (2.6) also turns out to be useful for the denoising of *shapes*. A natural variational model for denoising a shape S, which we model as a subset of $\mathbb{R}^n$, is the following: $\min_{\Sigma\subset\mathbb{R}^n} \mathrm{Per}(\Sigma) + \lambda|S \bigtriangleup \Sigma|$, where the first term in the energy represents the perimeter of the set Σ, and the second represents the volume of the symmetric difference of the sets S and Σ weighted by the scale parameter $\lambda \geq 0$. This model is exactly the one we would get if the minimization in the standard ROF model (2.2) is restricted to functions of the form $u(x) = 1_\Sigma(x)$ and $f(x) = 1_S(x)$. Unlike the standard ROF problem, however, this minimization is non-convex. In particular, standard approaches for solving it run the risk of getting stuck in local minima. The total variation model with L^1 fidelity term (2.6) turns out to be a convex formulation of the shape denoising problem given above. Indeed, the following statement has been proved in [165]: Let $u(x)$ be a minimizer of (2.6) for $f(x) = 1_S(x)$. Then, for a.e. $\mu \in [0, 1]$, the set $\Sigma(\mu) = \{x \in \mathbb{R}^N : u(x) \geq \mu\}$ is a minimizer of the shape denoising problem. Thus, in order to solve the *non-convex* shape denoising problem, it suffices to solve instead the *convex* problem (2.6) and then take (essentially) any level set of that solution.

2.4.3 Anisotropic TV

In [299], Esedoglu and Osher introduced and studied anisotropic versions of the ROF model (2.2). The motivation is to privilege certain edge directions so that they are preferred in reconstructions. This can be useful in applications in which there may be prior geometric information available about the shapes expected in the recovered image. In particular, it can be used to restore characteristic functions of convex regions having desired shapes.

The idea proposed in [299] is to replace the total variation penalty term in (2.2) with the following more general term:

$$\int_\Omega \phi(\nabla u) := \sup_{\substack{g\in C_c^1(\Omega;\mathbb{R}^n)\\ g(x)\in W_\phi \forall x\in\Omega}} \int_\Omega u(x)\mathrm{div} g(x)\, dx$$

where the function $\phi : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a convex, positively one-homogeneous function that is 0 at the origin, and the set W_ϕ is defined as follows:

$$W_\phi := \{y \in \mathbb{R}^n : x \cdot y \leq \phi(x) \ \forall x \in \mathbb{R}^n\}.$$

For example, if $\phi(x) = |x|$, then the set W_ϕ turns out to be simply the unit ball $\{y \in \mathbb{R}^N : |y| \leq 1\}$, and the definition of $\int_\Omega \phi(\nabla u)$ given above reduces to the standard definition of total variation. Another simple example in two dimensions is $\phi(x, y) = |x| + |y|$, in which case the set W_ϕ is just the closed unit square.

The set W_ϕ defined above is the *Wulff shape* associated with the function ϕ. It determines the shapes that are compatible with the anisotropy ϕ. For example, it is proved in [299] that if $f(x)$ is the characteristic function of (a scaled or translated version of) the Wulff shape W_ϕ, then the solution u is a constant multiple of $f(x)$. This result generalizes that of Strong and Chan [770] and Meyer in [567] that concern the case of a disk for the standard ROF model.

If W_ϕ is a convex polygon in two dimensions, then its sides act as preferred edge directions for the reconstructions obtained by the anisotropic ROF model. Indeed, it is proved in [299] that if $u(x) = 1_\Sigma(x)$ is a solution to the anisotropic model, and if Σ is known to be a set with piecewise smooth boundary $\partial\Sigma$, then $\partial\Sigma$ should include a line segment parallel to one of the sides of W_ϕ wherever its tangent becomes parallel to one of those sides. On the other hand, one can show that $\partial\Sigma$ can include corners that are different than the ones in ∂W_ϕ.

In addition to being of interest for applications, the results of [299] are also of theoretical interest. Indeed, these anisotropic variants of total variation constitute an infinitude of equivalent regularizations (in the sense that the semi-norms they define are equivalent), yet the properties of their minimizers have been shown to be extremely different. That suggests that in general one should not expect an image restoration model to perform quite as well as the original ROF model just because its regularization term is equivalent to total variation.

2.4.4 $H^{1,p}$ Regularization and Inf Convolution

As discussed in Section 2.3, staircasing is one of the potential caveats to watch for when using total variation based regularization. It occurs even more severely in reconstructions by functionals that have a non-convex dependence on image gradients; one famous example is the Perona-Malik scheme, which can be thought of as gradient descent for such an energy functional whose dependence on image gradients grows sublinearly at infinity. The TV model is borderline convex: its dependence on image gradients is linear at infinity. This feature, which is responsible for its ability to reconstruct images with discontinuities, is also responsible for the staircasing effect.

A natural approach to overcoming the staircasing effect is to make the reconstruction model more convex in regions of moderate gradient (away from the edges). A functional designed to accomplish this was proposed by Blomgren, Mulet, Chan, and Wong [99]. It has the form

$$\int_\Omega |\nabla u|^{P(|\nabla u|)}\,dx + \lambda \int_\Omega (u-f)^2\,dx. \tag{2.7}$$

Here, the function $P(\xi) : \mathbb{R}^+ \longrightarrow [0,2]$ is to be chosen so that it monotonically decreases from 2 to 0. A simple example is $P(\xi) = \frac{2}{1+2\xi}$.

The idea behind (2.7) is that the model automatically adapts the gradient exponent to fit the data, so that near edges it behaves exactly like the ROF model, and away from the edges it may behave more like the Dirichlet energy. This leads to much smoother reconstructions in regions of moderate gradient and thus prevents staircasing. On the other hand, unlike the ROF model, (2.7) is non-convex and difficult to analyze.

Another approach to preventing staircasing is to introduce higher order derivatives into the energy; the cost of moderately high but constant gradient regions is zero for such terms. On the other hand, a functional that depends on higher order derivatives would not maintain edges in its reconstructions. It is therefore necessary to once again allow the model to decide for itself where to use the total variation norm and where to use higher order derivative norms. One of the earliest proposals of this kind was made by Chambolle and Lions in [162], where they introduced the notion of *inf convolution* between two convex functionals. In this approach, an image u is decomposed into two parts: $u = u_1 + u_2$. The u_1 component is measured using the total variation norm, while the second component u_2 is measured using a higher order norm. The precise decomposition of u into these two components is part of the minimization problem. More precisely, one solves the following variational problem that now involves two unknowns:

$$\inf_{u_1,u_2} \int_\Omega |\nabla u_1| + \alpha |D^2 u_2| + \lambda (u_1 + u_2 - f)^2\,dx.$$

Minimizing this energy requires the discontinuous component of the image to be allocated to the u_1 component, while regions that are well approximated by moderate but nearly constant slopes get allocated to the u_2 component at very little cost. This prevents staircasing to a remarkable degree in the one dimensional examples presented in [162]. Another method that utilizes total variation and higher order derivatives to suppress staircasing is by Chan, Marquina, and Mulet in [168].

Despite the important contributions listed above, staircasing remains one of the challenges of total variation based image reconstructions.

2.5 Further Applications to Image Reconstruction

2.5.1 Deconvolution

The TV norm can also be used to regularize image deblurring problems. The forward degradation model for a blurred and noisy image can be realized as: $f = k * u + \eta$, where f is the observed (degraded) image, k a given point spread function (PSF), u the clean image, η an additive noise (often Gaussian), and $*$ denoting the convolution operator.

The task of restoring an image u under the above degradation is known as deconvolution if the PSF k is known or blind deconvolution if there is little or no known a priori information on the PSF. If we replace the u in the unconstrained ROF model (2.2) with the convolution $k * u$, then we arrive at the TV deconvolution model:

$$\min_{u \in BV} ||k * u - f||_2^2 + \lambda_u ||u||_{TV}. \tag{2.8}$$

Here, as in the ROF model (2.2), the regularization parameter λ_u is related to the statistical signal to noise ratio (SNR).

Extending the work by You and Kaveh [911], Chan and Wong introduce in [176] the TV blind deconvolution model:

$$\min_{u,k \in BV} \|k * u - f\|_2^2 + \lambda_u \|u\|_{\mathrm{TV}} + \lambda_k \|k\|_{\mathrm{TV}}. \tag{2.9}$$

where the additional parameter λ_k controls the spread of k. Moreover, solutions $\{u(\lambda_k)\}$ of (2.9) form a one parameter family corresponding to λ_k. The authors also propose an alternating minimization algorithm for minimizing the above energy (2.9) which we denote by $F(u, k)$. Here, given u^n one solves for $k^{n+1} := \arg\min_k F(u^n, k)$, then given k^{n+1}, one solves for $u^{n+1} := \arg\min_u F(u, k^{n+1})$ alternatingly. Such an alternating procedure is shown to be convergent when the TV-norm is replaced by the H^1-norm.

A key advantage of using TV regularization for blind deconvolution is that the TV norm can recover sharp edges in the PSF (e.g. motion blur or out-of-focus blur) while not penalizing smooth transitions.

2.5.2 Inpainting

Image inpainting refers to the filling-in of missing or occluded regions in an image based on information available on the observed regions. A common principle for inpainting is to complete isophotes (level sets) in a natural way — such a philosophy is also true for professional artists to restore damaged ancient paintings. To this end, several successful inpainting models have been proposed such as Masnou and Morel [553] and Bertalmio et al. [79]. We refer the reader to [171] and the references therein for other more recent models. Among these models, Chan and Shen proposed in [171] a *TV inpainting model* which uses variational methods in inpainting. The basic ingredient is to solve the boundary value problem:

$$\min_u \int_\Omega |\nabla u| \qquad \text{subject to} \qquad u = u_0 \qquad \text{in } \Omega \setminus D. \tag{2.10}$$

Here, D is the missing region to be inpainted, u_0 is the observed image whose value in D is missing. Thus, the TV inpainting method simply fills-in the missing region such that the TV in Ω is minimized. The use of TV-norm is desirable because it has the effect of extending level sets into D without smearing discontinuities along the tangential direction of the boundary of D.

With a slight modification of (2.10), simultaneous inpainting (in D) and denoising (in $\Omega \setminus D$) may be done as follows:

$$\min_u \int_\Omega |\nabla u| + \lambda \int_{\Omega \setminus D} (u - u_0)^2 dx. \tag{2.11}$$

Define a spatial varying parameter $\lambda_e(x)$ which is 0 in D and is λ in $\Omega \setminus D$. Then the Euler-Lagrange equation for (2.11) can be written as

$$-\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + 2\lambda_e(u - u_0) = 0$$

which has the same form as that for the ROF model, except the regularization is switching between 0 and λ in different regions. Thus, it is easy to modify an implementation of the ROF model to the TV inpainting model. Finally, we remark that some variants of (2.11) such as curvature-driven diffusion [172] and Euler's Elastica [167] have been proposed which complete isophotes in a smoother way.

2.5.3 *Texture and Multiscale Decompositions*

Another way of looking at denoising problems is by separating a given noisy image f into two components to form the decomposition: $f = u + v$, where u is the denoised image and $v = f - u$ the noise. In [567], Meyer adopts this view for the purpose of texture extraction where v captures not only noise but also texture. To do this, he proposed a new decomposition model:

$$\inf_u \left\{ E(u) = \int_\Omega |\nabla u| + \lambda \|v\|_*, f = u + v \right\} \tag{2.12}$$

where the $*$ norm is given by:

$$\|v\|_* = \inf_{\mathrm{g}=(g_1,g_2)} \left\{ \|\sqrt{g_1{}^2 + g_2{}^2}\|_{L^\infty} \mid v = \partial_x g_1 + \partial_y g_2 \right\} \tag{2.13}$$

and the v component lies in what is essentially the dual space of BV, the G space:

$$G = \left\{ v \mid v = \partial_x g_1 + \partial_y g_2 \ , \ g_1, g_2 \in L^\infty(\mathbb{R}^2) \right\}. \tag{2.14}$$

Here, v is an oscillatory function representing texture and the $*$ norm is designed to give small value for these functions. Thus, the main idea in (2.12) is to try to pull out texture by controlling $\|v\|_*$. Experiments in [843, 619] (discussed below) visually show that the model (2.12) extracts texture better than the standard ROF model.

In practice, the model (2.12) is difficult to implement due to the nature of the $*$ norm. Vese and Osher [843] were the first to overcome this difficulty where they devise an L^p approximation to the norm $\| \cdot \|_*$. In a later work [619], Osher et al. propose another L^p approximation based on the H^{-1} norm and introduce a resulting fourth order PDE. Both works numerically demonstrate the effectiveness of the model (2.12) for texture extraction and also give some further applications to denoising and deblurring.

In a related work, Aujol et al. [36] propose a decomposition algorithm based on Meyer's work [567] where they further decompose an image as $f = u + v + w$ where u, v, and w are cartoon, texture, and noise respectively.

Given the scale properties of the ROF model seen in section 2.2.3, it is natural to consider a multiscale decomposition based on the ROF model. Multiscale decompositions are of particular interest since one may want to extract image features of many different scales (either coarse or fine). One such multiscale decomposition is Tadmor et al. [784] and proceeds in a hierarchical manner. After choosing an initial $\lambda_0 = \lambda$ to remove the smallest oscillation in a given image f, the regularization parameters $\{\lambda_j\}$, $\lambda_j = 2^j\lambda$ induce a sequence of dyadic scales for $j = 1, \dots, k$. If we denote by u_{λ_j} the solution to the ROF model (2.2) for parameter λ_j, then f has the decomposition:

$$f = u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + \cdots + u_{\lambda_k} + v_{\lambda_k}.$$

with v_{λ_k} denoting the k-th stage residual $v_{\lambda_k} = f - (u_{\lambda_0} + u_{\lambda_1} + u_{\lambda_2} + \cdots + u_{\lambda_k})$. Furthermore, the authors show that $\|v_{\lambda_k}\|_* \to 0$ as $k \to \infty$. Hence $\|f - \sum_{i=0}^{k} u_{\lambda_i}\|_* \to 0$ as $k \to \infty$ and the decomposition converges to f in the $*$ norm. A related work based on merging dynamics of a monotonicity constrained TV model can be found in [169].

2.6 Numerical Methods

There have been numerous numerical algorithms proposed for minimizing the ROF objective. Most of them fall into the three main approaches, namely, direct optimization, solving the associated Euler-Lagrange equations and using the dual variable explicitly in the solution process to overcome some computational difficulties encountered in the primal problem. We will focus on the latter two approaches.

2.6.1 Artificial Time Marching and Fixed Point Iteration

In their original paper [695], Rudin et al. proposed the use of artificial time marching to solve the Euler-Lagrange equations which is equivalent to the steepest descent of the energy function. More precisely, consider the image as a function of space and time and seek the steady state of the equation

$$\frac{\partial u}{\partial t} = \nabla \cdot \left(\frac{\nabla u}{|\nabla u|_\beta} \right) - 2\lambda(u - f). \tag{2.15}$$

Here, $|\nabla u|_\beta := \sqrt{|\nabla u| + \beta^2}$ is a regularized version of $|\nabla u|$ to reduce degeneracies in flat regions where $|\nabla u| \approx 0$. In numerical implementation, an explicit time marching scheme with time step Δt and space step size Δx is used. Under this method, the objective value of the ROF model is guaranteed to be decreasing and the solution will tend to the unique minimizer as time increases. However, the convergence is usually slow due to the Courant-Friedrichs-Lewy (CFL) condition, $\Delta t \leq c\Delta x^2|\nabla u|$ for some constant $c > 0$ (see [546]), imposed on the size of the time step, especially in flat regions where $|\nabla u| \approx 0$. To relax the CFL condition, Marquina and Osher use, in [546], a "preconditioning" technique to cancel singularities due to the degenerate diffusion coefficient $1/|\nabla u|$:

$$\frac{\partial u}{\partial t} = |\nabla u| \left[\nabla \cdot \left(\frac{\nabla u}{|\nabla u|_\beta} \right) - 2\lambda(u - f) \right] \tag{2.16}$$

which can also be viewed as mean curvature motion with a forcing term $-2\lambda(u - f)$. Explicit schemes suggested in [546] for solving the above equation improve the CFL to $\Delta t \leq c\Delta x^2$ which is independent of $|\nabla u|$.

To completely get rid of CFL conditions, Vogel and Oman proposed in [849] a fixed point iteration scheme (FP) which solves the stationary Euler-Lagrange directly. The Euler-Lagrange equation is linearized by lagging the diffusion coefficient and thus the $(i + 1)$-th iterate is obtained by solving the sparse linear equation:

$$\nabla \cdot \left(\frac{\nabla u^{i+1}}{|\nabla u^i|_\beta} \right) - \lambda(u^{i+1} - f) = 0. \tag{2.17}$$

While this method converges only linearly, empirically, only a few iterations are needed to achieve visual accuracy. In practice, one typically employs specifically designed fast solvers to solve (2.17) in each iteration.

2.6.2 Duality-based Methods

The methods described in Section 2.5.1 are based on solving the primal Euler-Lagrange equation which is degenerate in regions where $\nabla u = 0$. Although regularization by $1/|\nabla u|_\beta$ avoids the coefficient of the parabolic term becoming arbitrarily large, the use of a large enough β for effective regularization will reduce the ability of the ROF model to preserve edges.

Chan et al. in [166], Carter in [151] and Chambolle in [160] exploit the dual formulation of the ROF model By using the identity $\|\mathbf{x}\| \equiv \sup_{\|\mathbf{g}\| \leq 1} \mathbf{x} \cdot \mathbf{g}$ for vectors in Euclidean spaces and treating $\mathbf{g}$ as the dual variable, one arrives at the dual formulation:

$$\sup_{\mathbf{g} \in C_c^1(\Omega, B^2)} \int_\Omega f\nabla \cdot \mathbf{g} dx - \frac{1}{2\lambda} \int_\Omega (\nabla \cdot \mathbf{g})^2 dx \tag{2.18}$$

where B^2 is the unit disk in $\mathbb{R}^2$. Once $\mathbf{g}$ is obtained, the primal variable can be recovered by $u = f - \lambda^{-1}\nabla \cdot \mathbf{g}$. A promise of the dual formulation is that the objective function is differentiable in $\mathbf{g}$, unlike the primal problem which is badly behaved when $\nabla u = 0$. However, the optimization problem becomes a constrained one which requires additional complexity to solve.

The approach used in [166] solves for u and g simultaneously. Its derivation starts by treating the term $\nabla u/|\nabla u|$ in the primal Euler-Lagrange equation as an independent variable g, leading to the system:

$$-\nabla \cdot \mathrm{g} + \lambda(u - f) = 0, \qquad \mathrm{g}|\nabla u|_\beta - \nabla u = 0.$$

The above system of nonlinear equations is solved by Newton's method and quadratic convergence rate is almost always achieved. In the Newton updates, one may combine the two equations to eliminate the need to update g, thus the cost per iteration is as cheap as the fixed point iteration (2.17). Empirically, this primal-dual method is much more robust than applying Newton's method directly to the primal problem in u only.

In [160], Chambolle devised an efficient algorithm solely based on the dual formulation (2.18). By carefully looking at the Euler-Lagrange equation for (2.18) and eliminating the associated Lagrange multipliers, one arrives at solving $H(\mathrm{g}) - |H(\mathrm{g})| = 0$ where $H(\mathrm{g}) = -\nabla(f - \lambda^{-1}\nabla \cdot \mathrm{g})$ is the negative of the gradient of the primal variable u. The update formula for g used in [160] is a simple relaxation $\mathrm{g}^{n+1} = \frac{\mathrm{g}^n + \tau H(\mathrm{g}^n)}{1+\tau|H(\mathrm{g}^n)|}$ where $\tau > 0$ is chosen to be small enough so that the iteration converges.