University of California
Los Angeles

Variational Models for Fine Structures

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Hayden Kyler Schaeffer

2013

Abstract of the Dissertation

Variational Models for Fine Structures

by

Hayden Kyler Schaeffer

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2013

Professor Stanley Osher, Co-chair

Professor Luminita Vese, Co-chair

Mathematical models in imaging science attempt to understand and analyze the underlying quantitative structure of images. The most popular mathematical techniques tend to center around a variational principle. In general, variational methods are formulated by specifying an energy whose minimizers contain properties associated with an ideal image. Thus far, variational models have been successful in addressing the classical problems in imaging, namely the problems of denoising, deblurring, segmentation, and inpainting. Most work has concentrated around reconstructing homogeneous intensity regions with jump discontinuities (*i.e.* edges) – one type of fine structure. More recent work has included models which incorporate tools for texture recovery. In practice, the most challenging components to recover from images are those which reside on fine-scales, namely the jumps and textures. This thesis focuses on the recovery and understanding of fine-scale information.

In many image segmentation methods, the edge set is regularized by the Hausdorff measure (*i.e.* length). It is known that minimizers of models containing length regularizers have segments whose endpoints either terminate perpendicularly to the boundary of the domain, terminate at a triple junction where three segments connect, or terminate at a free endpoint where the segment does not connect to any other edges or the boundary of the domain. However, standard segmentation methods (those that are based on the level set method) are only able to capture edge structures which contain the first two types of segments. Part I

generalizes the level set based image segmentation methods to be able to detect free endpoint structures. This results in the ability to capture a larger class of edge structures using the length regularizer while recovering homogeneous regions.

Aside from edge recovery, cartoon-texture regularization applied to ill-posed imaging problems allows for the reconstruction of many small-scale (patterned) details. The cartoon component is typically modeled by functions of bounded variation and has been shown to be a successful descriptor of the large geometric structures in images. However, current texture models are not universal and may depend on the problem or the particular class of images. In general, texture is defined by its highly oscillatory nature and its well patterned structure. Exploiting each of these properties, two texture models are provided, one using weak functional spaces to promote oscillations and the other using matrix theory to define patterns.

The first texture regularization is measured by duality with a space of functions which approximates $W^{1,\infty}$, thereby encouraging oscillations. In order to provide a differentiable approximation to the L^∞ norm, a concentration of measures approach is taken. This model works well for reconstructing texture in highly degraded blurry images.

The second texture model is defined by the nuclear norm applied to patches in the image, interpreting the texture patches to be low-rank. This provides a mathematical description for highly patterned texture as well as an easy to implement numerical method. This particular texture model has the advantage of separating noise from texture and has been shown to better reconstruct texture for other applications such as denoising, known deblurring, sparse reconstruction, and pattern regularization.

The dissertation of Hayden Kyler Schaeffer is approved.

Paul Thompson

John Garnett

Inwon Kim

Luminita Vese, Committee Co-chair

Stanley Osher, Committee Co-chair

University of California, Los Angeles

2013

To my love.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank all those who made this possible. I am grateful for all of their support and help throughout the years.

I have had the pleasure of working with many outstanding mathematicians – especially my advisors Stanley Osher and Luminita Vese. I thank my advisors for their mathematical, professional, and personal guidance. I want to thank Stanley Osher for his encouragement, wisdom, and enthusiam- all of which have been invaluable to my research. I want to thank Luminita Vese for her support, advice, and insight. I am glad I was able to attend their classes and thank them for their inspiring lectures. I have benefited from their mentorship and am truly grateful.

I would like to express my gratitude for the support I received from the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program. With it, I was able to focus on learning and researching. The published work of "Active Contours with Free Endpoints," Journal of Mathematical Imaging and Vision, (2013), pages 1-17 contains much of the research presented in Chapter 3 and was completed with the help of Luminita Vese. Also, Chapters 6 and 7 contain research with Stanley Osher and can be found in "A Low Patch-Rank Interpretation of Texture," SIAM Journal on Imaging Sciences, 6-1 (2013), pp. 226-262 .

I would like to thank my committee members for their time and suggestions. I thank Inwon Kim for her excellent lectures in partial differential equations and helpful discussions. I thank John Garnett for his thorough questions, productive conversations, and his help in analysis. And I would like to thank Paul Thompson for his interest in my work and for his valuable questions.

I have had the opportunity to collaborate with many incredible individuals. I thank them for all that they have taught me. I would like to thank my friends, colleagues, and offices mates for their support and enjoyable discussions. I would also like to express my

appreciation for the wonderful UCLA mathematics department and IPAM staff for going beyond their jobs to help others.

And most importantly, I thank my love.

Vita

2010-2013 Research Fellow, University of California, Los Angeles.

Publications

H. Schaeffer and S. Osher. "A Low Patch-Rank Interpretation of Texture." SIAM Journal on Imaging Sciences, 6-1 (2013), pp. 226-262

H. Schaeffer and L. Vese. "Active Contours with Free Endpoints." Journal of Mathematical Imaging and Vision (2013), pages 1-17.

CHAPTER 1

Introduction

Modeling is fundamental to the science of imaging. Since images are essentially digital representations of our visual perception, they can be as complex as the scenes they depict. They are geometrically multi-structured, typically including a wide range of shapes, patterns, scales, and even randomness. The role of the model is to provide a mathematical description of the main features which constitute all images. By understanding these features, we can better analyze images and videos as well as create more accurate and efficient methods for various applications.

This thesis focuses on variational models, which are concerned with minimizing a functional associated with a given problem. The functional incorporates the preferred quantitative behavior, where their minimizers represent the ideal image. The difficult task is identifying the important features and relating those features to a mathematical quantity. Classical methods center around constructing functionals whose minimizers contain the appropriate boundaries between objects. Object boundaries are modeled by discontinuities between homogeneous regions, thereby providing a mathematical representation of large scale geometric features in images. The discontinuities in intensity values also create sharp contrast, which is visually desired.

Modern methods are concerned with the mathematical formulation of texture, since this is an important visual feature. Unlike edges, texture can be dense and highly oscillatory, related to its fine scale and non-local nature. The texture component provides details, making images appear less cartoon-like and more realistic.

This thesis builds on previous work related to edges and texture, to create general and

effective representations, models, and methods for recovering and analyzing fine-scale details.

1.1 Classical Models for Fine Structures

There are three main variational models which are essential to understanding the mathematics of fine details in modern imaging methods. Each of the models can be written in the following form:

$$\inf_{u \in X} E(u) = \mu ||u||_X + ||f - u||_{L^2}^2 \tag{1.1}$$

where u is an approximation to the true image with respect to a space X, f is the given image, and $\mu > 0$ is a tuning parameter. The choice of X defines the properties of the minimizers, and thus the structure of the recovered image. In this form, the main task is to construct a choice for X which best represents the important features in images.

The first is the Mumford and Shah (MS) model which attempts to find a piecewise smooth approximation, u, of a given image f that can contain jumps along an edge set Γ by minimizing the following [58]:

$$\inf_{u,\Gamma} E_{MS}(u, \Gamma) = \mu \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \gamma \int_{\Omega} |u - f|^2 dx + \lambda \mathcal{H}^1(\Gamma) \tag{1.2}$$

The first term requires u to be smooth outside of the jump set Γ, the second term ensures that u remains close to f in the L^2 sense, and the last term is the Hausdorff measure ("length") of the jump set which regularizes the edge set. The minimizing edge set is made up of $C^{1,1}$ segments whose endpoints either terminate perpendicularly to the boundary of the domain, terminate at a triple junction where three segments connect, or terminate at a free endpoint where the segment does not connect to any other segments. In the MS model, the object boundaries are explicitly given by Γ.

The second, is the Rudin, Osher, and Fatemi (ROF) model, which can be written as

follows [67]:

$$\inf_u E_{ROF}(u) = \mu||u||_{TV} + \frac{1}{2}||f-u||_{L^2}^2 \tag{1.3}$$

where $||\cdot||_{TV}$ is the total variation (TV) semi-norm for the space of functions of bounded variation (BV). In practice, minimizers of the ROF model are nearly piecewise constant and thus recover both sharp edges and homogeneous regions in the image.

Both the MS and ROF models are related to the space BV. Let Ω be an open, bounded and connected subset of $\mathbb{R}^n$, with Lipschitz boundary, then the space BV is defined below.

Definition 1.1. A function $u : \Omega \to \mathbb{R}$ is of bounded variation if and only if $u \in L^1(\Omega)$ and there exists a finite $\mathbb{R}^n$ valued Radon measure Du such that for all $\phi \in C_c^1(\Omega, \mathbb{R})$,

$$\int_\Omega \frac{\partial \phi}{\partial x_j}\ u\ dx = -\int_\Omega \phi\ (Du)_j, \quad \text{for } j = 1, ..., n. \tag{1.4}$$

and the TV semi-norm

$$||u||_{TV} = \int_\Omega |Du| := \sup\left\{ \int_\Omega u\ \mathrm{div}\,(\phi)\, dx \;\middle|\; \phi \in C_c^1(\Omega, \mathbb{R}^n), ||\phi||_{L^\infty} \leq 1 \right\} \tag{1.5}$$

is bounded.

If $u \in BV(\Omega)$, then the following decomposition holds (see Lebesgue's decomposition theorem):

$$\begin{aligned} Du &= \nabla u \cdot \mathcal{L}_n + D_s u \\ &= \nabla u \cdot \mathcal{L}_n + (u^+ - u^-)\nu \cdot \mathcal{H}_{S_u}^{n-1} + C_u \end{aligned}$$

where $\mathcal{L}_n$ is the Lebesgue measure, $\mathcal{H}_{S_u}^{n-1}$ is the Hausdorff measure of the jump set, and lastly C_u is the Cantor measure. The norm on the Banach space $BV(\Omega)$ is:

$$||u||_{BV(\Omega)} := ||u||_{L^1(\Omega)} + ||u||_{TV(\Omega)} \tag{1.6}$$

The space of special functions of bounded variation (SBV) is a subspace of BV, corresponding to the case when the Cantor component of the measure Du is identically zero (*i.e.* $C_u \equiv 0$). The ROF and MS models are equivalent (in some sense) to equation (1.1) when taking $X = BV$ and $X = SBV$, respectively.

For both models, the minimizers are very smooth away from the jump sets. This gives the appearance of a cartoon version of the input image f. In order to capture many of the details, Meyer [55] proposed using the G-norm, a norm weaker than the L^2 norm, to model the highly oscillatory nature of texture. The space G of distributions v which can be written as $v = \text{div}(\vec{g})$, where $g = (g_1, g_2) \in L^\infty$, and is equipt with the norm

$$||v||_G := \inf \left\{ ||\sqrt{g_1^2 + g_2^2}||_\infty \;\middle|\; v = \text{div}(\vec{g}) \right\}. \tag{1.7}$$

This space G closely resembles the dual of BV (in fact, it is the predual) and minimizers of the G norm are very oscillatory.

The $TV - G$ model represents the recovered image as $u + v$, where $u \in BV$ is the cartoon component containing edges and geometric features and $v \in G$ is the texture component containing small scale details. Altogether the $TV - G$ model is:

$$\inf_{u,v} E_{TV-G}(u, v) = \mu||u||_{TV} + \gamma||v||_G + \frac{1}{2}||f - u - v||_{L^2}^2 \tag{1.8}$$

In this way, the model decomposes the image into two main components and attempts to model the preferred behavior of each feature separately.

These three models provide the building blocks for many modern imaging methods and incorporate the basic features currently used in image recover and analysis.

1.2 Outline

The thesis is divided into two parts which both include mathematical formulations, numerical methods, and examples. The first part focuses on the problem of representing general edge structures that minimize length, both in theory and in practice. The second part contains two different perspectives on texture modeling, one from the point of view of weak spaces and the other from matrix representation.

Part I

Curve Evolution and Image Segmentation

CHAPTER 2

Introduction to Mathematical Segmentation

Image segmentation is one of the most important yet challenging problems in imaging science. The goal of segmentation is to locate important edges and boundaries in an image. The difficulty resides in the notion of the edge or boundary set, and thus the definition of segmentation itself. One way to define segmentation is as the process of partitioning up an image into different features or objects. In this way, the edge set must be made up of curves with no endpoints (loops) or with endpoints that terminate at the boundaries of the image domain. An alternative definition, one that is employed by many mathematical segmentation techniques, categorizes segmenting images as locating boundaries between important features and objects. The latter is more general, since the boundary set may contain curves with free endpoints which do not partition the domain in the classical sense. We propose an extension to the classical level set based segmentation techniques which allow for the more general class of boundaries, including curves with free endpoints. We will do so by extending the active contours models [18, 74], using a different formulation of the edge set which can capture a large class of edges.

Recall, that the Mumford and Shah (MS) problem is defined as follows: find a piecewise smooth approximation, u, of a given image f, which may have jumps along a set Γ by minimizing the following [58]:

$$\inf_{u,\Gamma} E_{MS}(u,\Gamma) = \mu \int_{\Omega\setminus\Gamma} |\nabla u|^2 dx + \gamma \int_{\Omega} |u-f|^2 dx + \lambda \mathcal{H}^1(\Gamma) \tag{2.1}$$

The first term requires u to be smooth outside of the jump set Γ, the second term ensures that

u remains close to f in the L^2 sense, and the last term is the Hausdorff measure ("length") of the jump set which regularizes the edge set. Theoretical results on the existence and regularity of minimizers can be found in the works of Morel and Solimini [57, 56], Dal Maso, Morel, Solimini [22], and De Giorgi, Carriero and Leaci [36]. The edge set has a simple structure, it is composed of $C^{1,1}$ curves whose endpoints either terminate perpendicularly to the boundary of the domain, terminate at a triple junction where three segments connect, or terminate at a free endpoint where the segment does not connect to any other edges.

Based on Γ-convergence [1], Ambrosio and Tortorelli proposed an elliptic approximation to the Mumford and Shah model. The set Γ is replaced by a function $v \in [0,1]$, which is 1 away from an edge and 0 on an edge, thus acting as a continuous approximation of an indicator function. The approximated functional is:

$$\begin{aligned}\inf_{u,v} E_{AT}(u,v) = \mu \int_\Omega (v^2 + \eta_\epsilon)|\nabla u|^2 dx + \gamma \int_\Omega |u-f|^2 dx \\ +\lambda \int_\Omega \left(\frac{\epsilon}{2}|\nabla v|^2 dx + \frac{1}{2\epsilon}|v-1|^2 \right) dx \end{aligned} \tag{2.2}$$

where $\epsilon > 0$ is a small parameter and $\eta_\epsilon = o(\epsilon)$. The last integral replaces the length term and enforces that the function v is smooth and close to 1 except for a small region around an edge. It can be shown that as $\epsilon \to 0$ the functional above Γ-converges to the weak formulation of the Mumford and Shah functional. One advantage of this formulation over the level set based methods is that the edge set can contain all types of curves theoretically possible as minimizers of the Mumford and Shah model. On the other hand, since the indicator function v does not sharply define edges, the reconstructed image may not have sharp jumps. The width of the edge is determined by the parameter ϵ, which may also cause thickening of the edge set. This was extended in [70] to include a texture regularizer.

Our model is formulated using the level set method, proposed by Osher and Sethian [63]. The level set method provides an implicit representation for curves by defining them as the zero level set of a Lipschitz continuous function $\phi : \Omega \to \mathbb{R}$. Using the level set framework allows the curve to undergo changes in topology and allows for the formation of cusps and

corners during the minimization process. The level set method does restrict the class of possible edge sets to curves made up of segments without endpoints or that terminate at the boundary of the domain.

In [18], the authors proposed a region based segmentation and restoration method formulated within the level set framework. The reconstructed image u is defined as a piecewise constant function, equal to c_1 inside the region enclosed by the curve and c_2 outside the region enclosed by the curve. The energy minimization is as follows:

$$\inf_{c_1,c_2,\Gamma} E_{CV}(c_1,c_2,\Gamma) = \gamma_1 \int_{inside(\Gamma)} |f-c_1|^2 dx + \gamma_2 \int_{outside(\Gamma)} |f-c_2|^2 dx + \lambda \text{Length}(\Gamma) \quad (2.3)$$

The first two terms enforce that the given image f must remain close to the constants c_1 and c_2 in each of their respective regions. When minimized, the last term regularizes the edge set by making it as small as possible while still separating the two regions. Equation (2.3) is a special case of the Mumford and Shah functional [58], in which the reconstructed image is restricted to the class of piecewise constant solutions.

In [74], the authors proposed an extension of the original active contours model without edges, providing a practical implementation of the full Mumford and Shah model. Two level set functions are used to define four regions and the reconstructed function u is defined piecewise by four auxiliary functions u_j, $1 \leq j \leq 4$, which are smooth in each of their respective regions. The model is as follows:

$$\begin{aligned}\inf_{u_j,\Gamma} E_{VC}(u_j,\Gamma) = \mu &\left\{ \int_{\Omega_1} |\nabla u_1|^2 dx + \int_{\Omega_2} |\nabla u_2|^2 dx + \int_{\Omega_3} |\nabla u_3|^2 dx + \int_{\Omega_4} |\nabla u_4|^2 dx \right\} \\ &+ \gamma \left\{ \int_{\Omega_1} |u_1 - f|^2 dx + \int_{\Omega_2} |u_2 - f|^2 dx + \int_{\Omega_3} |u_3 - f|^2 dx + \int_{\Omega_4} |u_4 - f|^2 dx \right\} \\ &+ \lambda \text{Length}(\Gamma) \end{aligned} \quad (2.4)$$

and the regions $\Omega_1, ..., \Omega_4$ are represented using two open sets. The minimizing function u is smooth outside of the set Γ, where it may have jumps. The minimizing edge set Γ

is composed of curves that terminate perpendicularly to the boundary of the domain and curves without endpoints, but cannot have curves with free endpoints. See also [73] for related work, [14] for an alternative model based on geodesics, and [17] for an extension to vector valued images.

An alternative level set formulation can be found in [19], where the authors proposed a multilayer extension to the piecewise constant active contours model. Their model uses multiple level lines of ϕ in order to segment many embedded objects. The resulting reconstructed image is piecewise constant and the resulting edge set is composed of curves that terminate perpendicularly to the boundary of the domain and curves without endpoints which may be enclosed in each other.

The level set based segmentation methods thus far are unable to locate edges with free endpoints. We propose an extension to the level set techniques for segmentation by defining a more general edge set. Using the method from the work of Smereka [71] to represent curves with endpoints, we propose a level set based segmentation method which can capture free curves, in addition to all previously possible curves.

The majority of previous work on free curves rely on fixed endpoints or explicit particles/markers. In the Snakes active contours model [40], the curve is represented by splines which make it possible to have either free curves or curves without free endpoints. In [42], curves with free endpoints are represented with a marker-points method, which must be explicitly re-adjusted after each iteration. In [43], an explicit curve representation using polylines are used. A particle based curve evolution approach for segmentation can be found in [54]. For curve matching, a similar framework is proposed in [48]. For more on related work on curves with free endpoints see also [20, 44].

CHAPTER 3

Active Contours with Free Endpoints Model

3.1 Description of the Model

Before discussing the model, the level set method with the addition of free curves will be reviewed. After that, the energy will be presented with the associated equations of motion. Lastly, we will review the theory of Sobolev gradients and its application to our evolution equations. Note that a version of this chapter appears in [69].

3.1.1 Representation of Curves with Free Endpoints and Domain Partitioning

Osher and Sethian proposed the level set method as an implicit representation of curves [63]. In the classical level set formulation, a curve Γ is represented as the zero level set of a Lipschitz continuous function, $\phi : \Omega \to \mathbb{R}$. The assumption is that Γ is the boundary of an open set, and thus Γ is composed of curves without endpoints and curves which terminate at the boundary of the domain. The interior region is defined as the set of points where $\phi > 0$ and the exterior region is defined as the set of points where $\phi < 0$. The standard example of a level set function is the signed distance function to the curve.

Using the Heaviside function, defined as $H(s) = 1$, if $s \geq 0$, and $H(s) = 0$, if $s < 0$, in conjunction with the level set function, one can reformulate geometric quantities into easier-to-handle equations. Instead of looking at quantities along the curve, ϕ allows the calculations to be extended to the entire domain, making calculations more practical. For example, the length of the curve and the area enclosed by the curve can be written as [27]:

$$L(\Gamma) = \int |\nabla H(\phi)|, \qquad A(\Gamma) = \int H(\phi) dx$$

The derivative of the Heaviside function is taken in the sense of measures. The problem can be regularized by taken a differentiable approximation, H_ϵ, which limits to the Heaviside function as $\epsilon \to 0$. This provides an approximation to the Dirac delta function, $\delta_\epsilon = H'_\epsilon$, and the quantities above can be approximated by:

$$L_\epsilon(\Gamma) = \int \delta_\epsilon(\phi)|\nabla\phi| dx, \qquad A_\epsilon(\Gamma) = \int H_\epsilon(\phi) dx$$

These equations can then be minimized by introducing an artificial time and descending using (the negative) of the Euler-Lagrange equations respectively:

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi) \text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) \qquad \frac{\partial \phi}{\partial t} = -\delta_\epsilon(\phi)$$

Since the delta function's approximation is assumed to be strictly positive, the equations can be rescaled to the following equations, which have the same steady state solutions as the equations above:

$$\frac{\partial \phi}{\partial t} = |\nabla\phi| \text{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right), \qquad \frac{\partial \phi}{\partial t} = -|\nabla\phi|$$

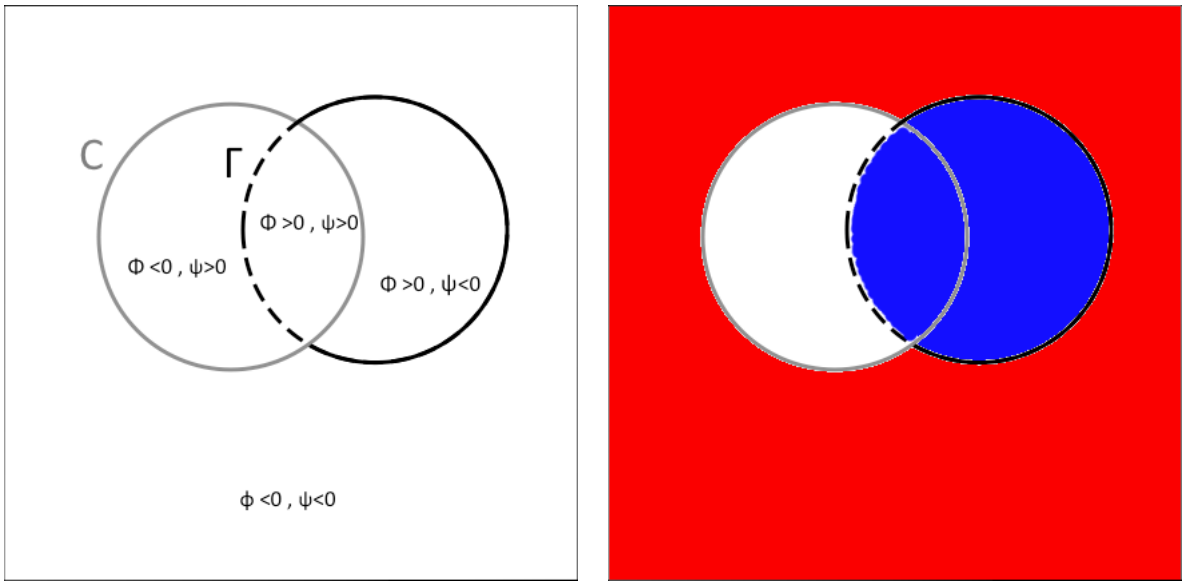
The length minimizing equation is the mean curvature flow, while the area minimizing equation is a Hamilton-Jacobi equation. Each of the two equations above play an important role in level set based segmentation models. For example, let us consider the model [18]:

$$E_{CV}(c_1, c_2, \phi) = \lambda \int \delta(\phi)|\nabla\phi| dx + \int \left((c_1 - f)^2 H(\phi) + (c_2 - f)^2 (1 - H(\phi))\right) dx$$

In terms of ϕ, the length term acts as the regularizer while the area terms are connected to the regional fidelity terms. Each region is clearly defined by the level set function, where the sign determines the regions. Using the same formulation, the two phase piecewise smooth

model is [74]:

$$E_{VC}(u_1, u_2, \phi) = \lambda \int \delta(\phi)|\nabla\phi| dx + \int \left((u_1 - f)^2 H(\phi) + (u_2 - f)^2 (1 - H(\phi))\right) dx$$
$$+\mu \int \left(|\nabla u_1|^2 H(\phi) + |\nabla u_2|^2 (1 - H(\phi))\right) dx.$$



(a) Level Set Representation of the Curve Γ, the dotted line above

(b) Color Coded Partition of Space

Figure 3.1: Level Set Representation of the curve Γ, with free endpoints

Again, the level set function partitions the domain into two regions. This partitioning allows the function u, which is an approximation of f, to have clearly defined jumps. We wish to further extend the level set based segmentation methods to allow for jumps on curves with free endpoints. This is done by introducing a second level set function ψ, which acts as an indicator function partitioning up the zero level set of ϕ into two segments (see Figure 3.1 a), based on the work of Smereka [71]. The curve Γ is defined as $\{x \in \Omega \mid \phi(x) = 0 \text{ and } \psi(x) > 0\}$, which now allows the curve to have loops, segments terminating at the boundary of the domain, and segments with free endpoints. Revisiting the length functional from before, the

new formulation is as follows:

$$L(\Gamma) = \int |\nabla H(\phi)| H(\psi), \qquad L_\epsilon(\Gamma) = \int \delta_\epsilon(\phi) H_\epsilon(\psi) |\nabla \phi| dx,$$

The main difference is the addition of the term $H(\psi)$ to the integral. The length can be minimized by descending using the first variation:

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \delta(\phi) \text{div}\left(H(\psi) \frac{\nabla \phi}{|\nabla \phi|} \right) \\ \frac{\partial \psi}{\partial t} &= -\delta(\psi)\delta(\phi)|\nabla \phi|, \end{aligned}$$

assuming the Heaviside and Dirac delta functions are replaced by their smooth counterparts. The standard rescaled versions become:

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= |\nabla \phi| \text{div}\left(H(\psi) \frac{\nabla \phi}{|\nabla \phi|} \right) \\ \frac{\partial \psi}{\partial t} &= -\delta(\phi)|\nabla \phi||\nabla \psi|. \end{aligned}$$

The equation for ϕ defines a mean curvature flow, while the equation for ψ defines an area minimizing Hamilton-Jacobi equation.

Although similar in structure, there is a key difference between the classical and the proposed formulations – in the proposed case, the domain does not have a natural partition, since there is no clear concept of the interior and exterior of Γ. A proper partition for the domain should enforce that the reconstructed function only has jumps along the curve, so one choice (and the choice we use here) is to divide the domain in three regions: region 0: $\phi < 0$ and $\psi < 0$ (red in Figure 3.1 b), region 1: $\phi > 0$ (blue in Figure 3.1 b), and region 2: $\phi < 0$ and $\psi > 0$ (white in Figure 3.1 b). In this way, the important boundary is the one separating the white and blue regions. As in the classical formulation, two auxiliary functions u_1 and u_2 are chosen so that each function is smooth in their respective regions, but unlike the other methods, they must be smooth over all zero level curves outside of $\phi = 0$

and $\psi > 0$. If we define the reconstructed image as a linear combination of the two auxiliary functions, then choosing the partition in this manner ensures that the only discontinuities in u lie on Γ. If we denote the regions using the following characteristic functions: $\chi_1 = H(\phi)$, $\chi_2 = H(\psi)(1 - H(\phi))$, and $\chi_0 = (1 - H(\psi))(1 - H(\phi))$, then u_1 exists in region 0 and region 1 and u_2 exists in region 0 and region 2. The reconstructed function becomes:

$$u = u_1\chi_1 + u_2\chi_2 + \left(\frac{u_1 + u_2}{2}\right)\chi_0 \tag{3.1}$$

where the two auxiliary functions are averaged in region 0, although any non-trivial linear combination is sufficient. Using the partition and function above, we can formulate the level set based MS energy for a general curve.

3.1.2 The Energy

Recall that the MS energy in terms of the reconstructed image u and jump set Γ, with given (possibly corrupted) image f is:

$$E_{MS}(u, \Gamma) = \mu \int_{\Omega\setminus\Gamma} |\nabla u|^2 dx + \lambda \mathcal{H}^1(\Gamma) + \int_\Omega |u - f|^2 dx.$$

Using equation (3.1) and our definition of Γ the energy becomes:

$$\begin{aligned} E(u_1, u_2, \phi, \psi) = \mu \int_\Omega & \left(|\nabla u_1|^2(\chi_1 + \frac{\chi_0}{4}) + |\nabla u_2|^2(\chi_2 + \frac{\chi_0}{4}) + \frac{1}{2}\nabla u_1 \cdot \nabla u_2 \chi_0 \right) dx \\ & + \int_\Omega \left(|u_1 - f|^2\chi_1 + |u_2 - f|^2\chi_2 + \left| \frac{u_1 + u_2}{2} - f \right|^2 \chi_0 \right) dx \\ & + \lambda \int_\Omega \chi_3 \, |\nabla \chi_1| \end{aligned} \tag{3.2}$$

where $\chi_3 = H(\psi)$ and all measure zero terms are ignored. The first three terms are the regional regularities of the reconstructed function, the next three terms are the regional fidelity terms, and the last term is the length regularity for the curve. The average of

the auxiliary function is necessary to ensure that both functions are smooth over the non-essential zero level sets. For the sake of argument, let us assume the reconstructed function u is exactly u_o (the original image), and, instead of an average, we took the sum of the auxiliary: $u = u_1\chi_1 + u_2\chi_2 + (u_1 + u_2)\,\chi_0$. Then we have the following: in region 1 $u_1 = u_o$, in region 2 $u_2 = u_o$, and in region 0 $u_1 + u_2 = u_o$. For this to be possible, one of the functions must take values that are at most half of u_o, so that function's values would sharply decrease over the boundary between region 1 (or 2) and region 0. By taking the average, neither function may have jumps over the non-essential zero level sets.

Next, assuming that the Heaviside functions are regularized, the Euler-Lagrange equations are as follows:

$$\begin{cases} \frac{\partial u_j}{\partial t} & = \mu \text{ div}\left(2\chi_j \nabla u_j + \frac{\chi_0}{2}\nabla(u_1+u_2)\right) - 2(u_j - f)\chi_j - \left(\frac{u_1+u_2}{2} - f\right)\chi_0 \\ \frac{\partial \phi}{\partial t} & = \delta(\phi)\Big\{\lambda \text{div}\left(\chi_3 \frac{\nabla\phi}{|\nabla\phi|}\right) - |u_1 - f|^2 + \chi_3|u_2 - f|^2 + (1-\chi_3)\left|\frac{u_1+u_2}{2} - f\right|^2 \\ & \qquad -\mu|\nabla u_1|^2\left(\frac{3}{4} + \frac{\chi_3}{4}\right) + \mu|\nabla u_2|^2\left(\frac{1}{4} + \frac{3\chi_3}{4}\right) + \frac{1}{2}\mu\nabla u_1 \cdot \nabla u_2(1-\chi_3)\Big\} \\ \frac{d\psi}{dt} & = -\delta(\psi)\Big\{\lambda|\nabla\chi_1| + (1-\chi_1)\Big(|u_2 - f|^2 - \left|\frac{u_1+u_2}{2} - f\right|^2 - \frac{\mu}{4}|\nabla u_1|^2 \\ & \qquad + \frac{3\mu}{4}|\nabla u_2|^2 - \frac{\mu}{2}\nabla u_1 \cdot \nabla u_2\Big)\Big\} \end{cases}$$

with the following boundary conditions:

$$\begin{cases} \chi_j \frac{\partial u_j}{\partial n} + \frac{1}{4}\chi_0 \frac{\partial}{\partial n}(u_1 + u_2) & = 0 \\ \frac{\chi_3}{|\nabla\phi|}\frac{\partial \phi}{\partial n} & = 0 \end{cases}$$

and the initial conditions: $u_j(0,x) = u_j^0(x)$, $\phi(0,x) = \phi^0(x)$, and $\psi(0,x) = \psi^0(x)$. For simplicity, the boundary condition can be reduced to the standard Neumann boundary condition, which we will show in Section 4. The system of PDEs includes diffusion equations for u_1 and u_2, mean curvature flow for ϕ and the area minimizing ODE for ψ. As before, the level set functions' PDE can be rescaled by the magnitudes of their gradients.

Remark 3.1. *Although not always necessary, in practice, extra regularization on the level*

set functions may ensure a smoother evolution. For example, the following regularized PDEs give a smoother flow for the system above:

$$\begin{cases} \frac{\partial \phi}{\partial t} & = \delta(\phi)\Big\{\lambda div\Big((\chi_3+\epsilon_1)\frac{\nabla\phi}{|\nabla\phi|}\Big) - |u_1-f|^2 + \chi_3|u_2-f|^2 + (1-\chi_3)\left|\frac{u_1+u_2}{2}-f\right|^2 \\ & \qquad -\mu|\nabla u_1|^2\left(\frac{3}{4}+\frac{\chi_3}{4}\right) + \mu|\nabla u_2|^2\left(\frac{1}{4}+\frac{3\chi_3}{4}\right) + \frac{1}{2}\mu\nabla u_1\cdot\nabla u_2(1-\chi_3)\Big\} \\ \frac{\partial \psi}{\partial t} & = \delta(\psi)\Big\{\epsilon_2\Delta_\infty\psi - \lambda|\nabla\chi_1| + (1-\chi_1)\Big(|u_2-f|^2 - \left|\frac{u_1+u_2}{2}-f\right|^2 - \frac{\mu}{4}|\nabla u_1|^2 \\ & \qquad +\frac{3\mu}{4}|\nabla u_2|^2 - \frac{\mu}{2}\nabla u_1\cdot\nabla u_2\Big)\Big\} \end{cases}$$

where ϵ_1 and ϵ_2 are (small) parameters and $\Delta_\infty\psi = \left\langle \frac{\nabla\psi}{|\nabla\psi|}, D^2\psi\frac{\nabla\psi}{|\nabla\psi|}\right\rangle$ is the renormalized infinity Laplacian. In terms of energy minimization, this is equivalent to adding a length regularizer on the zero level curve of ϕ and a (rescaled) infinity norm regularizer for $|\nabla\psi|$. For more details to the theory of the infinity Laplacian, see [15, 21, 28, 50] and for discretizations see [61].

3.1.3 Sobolev Gradient

In order to minimize the proposed energy, a gradient descent method is used. The first variation (or Euler-Lagrange equation) is embedded in a dynamic scheme as follows: given an energy $E(\phi)$ with Euler-Lagrange equations that are denoted as $\nabla_{L^2}E(\phi)$, the gradient descent is $\frac{\partial\phi}{\partial t} = -\nabla_{L^2}E(\phi)$. For a general energy, this PDE may not be well-posed and could lead to many issues. To better pose the equation, the Sobolev Gradient, denoted as $\nabla_{H^1}E(\phi)$ can be used. Here we provide a short derivation, following[59] and [65]. Assume that the energy can be written with a potential V as $E(\phi) = \int_\Omega V(D\phi)dx$ and it is to be minimized over $H^1(\Omega)$, where $D : H^1(\Omega) \to H^1(\Omega)\times L^2(\Omega)$ is the operator $D\phi = (\phi, \nabla\phi)^T$.

For all $\phi \in H^1(\Omega)$ and for any $h \in H_0^1(\Omega)$, the directional derivative is:

$$(E'(\phi), h) \;=\; \int_\Omega V'(D\phi)Dhdx = \langle\nabla V(D\phi), Dh\rangle_{L^2} = \langle D^*\nabla V(D\phi), h\rangle_{L^2}$$

where D^* is the adjoint of D. Assuming that all the terms above are in $L^2(\Omega)$, the L^2 gradient is defined as $\nabla_{L^2} E(\phi) := D^* \nabla V(D\phi)$ (which is the standard Euler-Lagrange equation). On the other hand, we can equate the L^2 inner product with the H^1 inner product in terms of the Sobolev gradient:

$$(E'(\phi), h) \quad = \quad \langle \nabla_{L^2} E(\phi), h \rangle_{L^2} = \langle \nabla_{H^1} E(\phi), h \rangle_{H^1}.$$

Using the operator D, the inner product becomes:

$$\langle \nabla_{H^1} E(\phi), h \rangle_{H^1} \quad = \quad \langle D(\nabla_{H^1} E(\phi)), Dh \rangle_{L^2} = \langle D^* D(\nabla_{H^1} E(\phi)), h \rangle_{L^2}$$

This yields the following expression for the Sobolev gradient: $\nabla_{H^1} E(\phi) = (D^* D)^{-1}(\nabla_{L^2} E(\phi)) = (I - \Delta)^{-1}(\nabla_{L^2} E(\phi))$. This is interpreted as a gradient descent with respect to a more appropriate inner product space. This can also be seen as a preconditioned descent, with the smoothing operator $(I - \Delta)^{-1}$.

This smoothing allows the Euler-Lagrange equations to reside in a large function class. Recall that the dual of $H^1(\Omega)$, denoted $H^{-1}(\Omega) := (H^1(\Omega))^*$ (assuming Neumann boundary conditions), is larger than L^2, since it contains weaker functions. The operator $(I - \Delta)^{-1}$ can be considered as a map from $H^{-1}(\Omega) \to H^1(\Omega)$ such that for $G \in H^{-1}(\Omega)$, there exists (by Lax-Milgram) a unique $v \in H^1(\Omega)$ which solves the weak problem:

$$v - \Delta v = G \tag{3.3}$$

with Neumann boundary conditions. Thus for all $\nabla_{L^2} E(\phi) \in H^{-1}(\Omega)$, we can find a $\nabla_{H^1} E(\phi) \in H^1(\Omega)$. Now lets examine this in terms of the evolution equations, which define an iterative process. In the semi-discrete case, we construct the sequence ϕ^n by $\phi^{n+1} = \phi^n - \Delta t \nabla_{L^2} E(\phi^n)$, with $\phi^0 \in H^1(\Omega)$ and $\Delta t > 0$, such that $E(\phi^{n+1}) < E(\phi^n)$. In order to have $\phi^{n+1} \in H^1(\Omega)$, we would require that $\nabla_{L^2} E(\phi^n) \in H^1(\Omega) \subset L^2(\Omega)$. In other words, the solution ϕ would be required to be more regular than necessary. This require-

ment is one of the reasons for the small time-steps used in L^2 gradient descent (in terms of stability). In terms of the energy minimization, the following theorem provides further benefits of the Sobolev gradient.

Theorem 3.2. *For all $t > 0$, let ϕ and ϕ_s be the solution of the L^2 gradient descent ($\frac{\partial \phi}{\partial t} = -\nabla_{L^2} E(\phi)$) and Sobolev gradient descent equation ($\frac{\partial \phi_s}{\partial t} = -\nabla_{H^1} E(\phi_s)$), respectively; then:*

$$\begin{aligned} \frac{dE(\phi)}{dt} &= -||\nabla_{L^2} E(\phi)||^2_{L^2} \\ \frac{dE(\phi_s)}{dt} &= -||\nabla_{H^1} E(\phi_s)||^2_{H^1} \end{aligned}$$

Proof. Assume that $E(\phi) = \int_\Omega V(D\phi)$ then we can formally take the time derivative as follows:

$$\begin{aligned} \frac{dE(\phi)}{dt} &= \int_\Omega \nabla_{L^2} E(\phi) \frac{\partial \phi}{\partial t} dx \\ &= -\int_\Omega |\nabla_{L^2} E(\phi)|^2 \, dx \\ &= -||\nabla_{L^2} E(\phi)||^2_{L^2} \end{aligned}$$

The equations above hold if $\nabla_{L^2} E(\phi) \in L^2(\Omega)$. Next, taking the Sobolev gradient descent:

$$\begin{aligned} \frac{dE(\phi_s)}{dt} &= \int_\Omega \nabla_{L^2} E(\phi_s) \frac{\partial \phi_s}{\partial t} dx \\ &= -\int_\Omega \left((I - \Delta) \frac{\partial \phi_s}{\partial t} \right) \frac{\partial \phi_s}{\partial t} dx \\ &= -\int_\Omega \left(\left| \frac{\partial \phi_s}{\partial t} \right|^2 + \left| \nabla \frac{\partial \phi_s}{\partial t} \right|^2 \right) dx \\ &= -\left|\left| \frac{\partial \phi_s}{\partial t} \right|\right|^2_{H^1} \\ &= -||\nabla_{H^1} E(\phi_s)||^2_{H^1} \end{aligned}$$

assuming that all terms satisfy Neumann boundary conditions. The equation above holds if $\nabla_{L^2} E(\phi_s) \in H^{-1}(\Omega)$. $\square$

Applying the Sobolev gradient to the system of equations from our model yields:

$$\begin{cases} \frac{\partial u_j}{\partial t} & = (I-\Delta)^{-1}\left\{\mu \text{ div}\left(2\chi_j \nabla u_j + \frac{\chi_0}{2}\nabla(u_1+u_2)\right) - 2(u_j - f)\chi_j - \left(\frac{u_1+u_2}{2} - f\right)\chi_0\right\} \\ \frac{\partial \phi}{\partial t} & = (I-\Delta)^{-1}\Big\{\delta(\phi)\Big(\lambda \text{div}\left(\chi_3 \frac{\nabla\phi}{|\nabla\phi|}\right) - |u_1 - f|^2 + \chi_3|u_2 - f|^2 + (1-\chi_3)\left|\frac{u_1+u_2}{2} - f\right|^2 \\ & \qquad -\mu|\nabla u_1|^2\left(\frac{3}{4} + \frac{\chi_3}{4}\right) + \mu|\nabla u_2|^2\left(\frac{1}{4} + \frac{3\chi_3}{4}\right) + \frac{1}{2}\mu\nabla u_1 \cdot \nabla u_2(1-\chi_3)\Big)\Big\} \\ \frac{\partial \psi}{\partial t} & = -(I-\Delta)^{-1}\Big\{\delta(\psi)\Big(\lambda|\nabla\chi_1| + (1-\chi_1)\Big(|u_2 - f|^2 - \left|\frac{u_1+u_2}{2} - f\right|^2 - \frac{\mu}{4}|\nabla u_1|^2 \\ & \qquad + \frac{3\mu}{4}|\nabla u_2|^2 - \frac{\mu}{2}\nabla u_1 \cdot \nabla u_2\Big)\Big)\Big\} \end{cases}$$

The versions above are used in practice. In terms of the equations in u_j, the Sobolev gradient's application is clear. For the other two equations, this may not be the case. In the case of ϕ, since the Dirac delta function and the characteristic functions are smooth and strictly positive and (for many applications) $|\nabla\phi| = 1$ a.e., the equation resembles a typical anisotropic diffusion equation, which exists in H^{-1}. Another interpretation is that, in terms of the level set functions, the delta function acts to concentrate the motion around the zero level sets, whose width is dependent on the smoothness of the approximations. The operator, $(I-\Delta)^{-1}$ in turn, continues to smooth the main area of influence of the delta function. With respect to Equation (3.3), it is easy to show that the equation can be re-written as an optimization problem:

$$\inf_v E(v) = \int |\nabla v|^2 dx + \int |v - G|^2 dx$$

For a simple example, take G to be a smooth and strictly positive version of the Dirac delta function. Then it can be shown that $v \geq 0$, $\int v dx = \int G dx$, and $||v||_\infty \leq ||G||_\infty$. In this way, v is a smoother and more "spread out" version than the original G. Like rescaling by the magnitude to the derivative, this operator can be viewed as a rescaling.

Remark 3.3. *Using the inner product* $\langle u, v\rangle_{H^1,A} = \langle Du, ADv\rangle_{L^2}$ *for positive definite matrices* A*, a more general* H^1 *gradient can be defined. For example, a simple rescaling* $\langle u, v\rangle_{H^1,\alpha,\beta} = \alpha \langle u, v\rangle_{L^2} + \beta \langle \nabla u, \nabla v\rangle_{L^2}$ *yields the following gradient:* $\nabla_{H^1,\alpha,\beta} E = (\alpha I - \beta\Delta)^{-1}(\nabla_{L^2} E)$*, for* $\alpha, \beta > 0$*. In our problem, it would be appropriate to choose* $\alpha = 1$ *and* $\beta = \mu$*, since that inner product naturally appears in the energy, but for consistency between results, we set both to 1.*

For further applications of Sobolev gradients to imaging problems, see [7, 39, 72, 66].

3.2 Analytical Remarks

In this section, we will analyze our approximation by showing that our model is consistent with the MS functional, via point-wise convergence, and discuss its relation to the other level set based segmentation models by looking at degenerate cases.

3.2.1 Consistency with the Mumford-Shah Functional

To derive our proposed energy, recall that we defined the subregions Ω_j $0 \leq j \leq 2$, where $\Omega \setminus \Gamma = \cup_j \Omega_j$, by the following characteristic functions: $\chi_1 = H(\phi)$, $\chi_2 = H(\psi)(1 - H(\phi))$, and $\chi_0 = (1 - H(\psi))(1 - H(\phi))$. Using these regions, the reconstructed function is defined by two auxiliary functions, u_1 and u_2, as follows: $u = u_1\chi_1 + u_2\chi_2 + \left(\frac{u_1+u_2}{2}\right)\chi_0$. Using these definitions, the L^2 norm in equation (2.1) becomes:

$$\begin{aligned}
\int_\Omega |u - f|^2 dx &= \int_{\Omega_1} |u - f|^2 dx + \int_{\Omega_2} |u - f|^2 dx \int_{\Omega_0} |u - f|^2 dx \\
&= \int_{\Omega_1} |u_1 - f|^2 dx + \int_{\Omega_2} |u_2 - f|^2 dx + \int_{\Omega_0} \left|\frac{u_1 + u_2}{2} - f\right|^2 dx \\
&= \int_\Omega \left(|u_1 - f|^2\chi_1 + |u_2 - f|^2\chi_2 + \left|\frac{u_1 + u_2}{2} - f\right|^2 \chi_0 \right) dx
\end{aligned}$$

and the H^1 semi-norm on $\Omega \setminus \Gamma$ in equation(2.1) becomes:

$$\begin{aligned}
\int_{\Omega\setminus\Gamma} |\nabla u|^2 dx &= \int_{\Omega_1} |\nabla u|^2 dx + \int_{\Omega_2} |\nabla u|^2 dx + \int_{\Omega_0} |\nabla u|^2 dx \\
&= \int_{\Omega_1} |\nabla u_1|^2 dx + \int_{\Omega_2} |\nabla u_2|^2 dx + \int_{\Omega_0} \left| \nabla \left(\frac{u_1 + u_2}{2} \right) \right|^2 dx \\
&= \int_{\Omega} |\nabla u_1|^2 \chi_1 dx + \int_{\Omega} |\nabla u_2|^2 \chi_2 dx + \int_{\Omega} \left| \nabla \left(\frac{u_1 + u_2}{2} \right) \right|^2 \chi_0 dx \\
&= \int_{\Omega} \left(|\nabla u_1|^2 (\chi_1 + \frac{\chi_0}{4}) + |\nabla u_2|^2 (\chi_2 + \frac{\chi_0}{4}) + \frac{1}{2} \nabla u_1 \cdot \nabla u_2 \chi_0 \right) dx
\end{aligned}$$

ignoring measure zero terms. Lastly, the length term in equation(2.1) becomes:

$$\text{Length}(\Gamma) = \int_{\Gamma} ds = \int_{\phi=0} H(\psi) ds = \int_{\Omega} |\nabla H(\phi)| H(\psi) = \int_{\Omega} \chi_3 \, |\nabla \chi_1|$$

where the equation above is in the sense of measures and $H' = \delta$, the Dirac delta function. All together, these three terms make up our proposed energy. In order to formally take the Euler-Lagrange equations, each Heaviside function is replaced with a smooth approximation, also yielding a continuous approximation to the delta function.

There is much freedom in the choice of approximations. In general, given any function $\delta_1 \in C^0$ such that $\int_{\mathbb{R}} \delta_1(x) dx = 1$, one can construct an approximation to the Dirac delta function by setting $\delta_\epsilon(x) := \frac{1}{\epsilon} \delta_1 \left(\frac{x}{\epsilon} \right)$ and an approximation to the Heaviside function by setting $H_\epsilon(x) := \int \delta_\epsilon(x) dx$. This yields the following properties:

1. $H_\epsilon(x) \to H(x)$ point-wise everywhere except at $x = 0$
2. $\delta_\epsilon = H'_\epsilon$
3. $H_\epsilon \in C^1$

These conditions are easily satisfied by our particular choice of approximations: $H_\epsilon(x) = \frac{1}{2} + \frac{1}{\pi} \text{arctan} \left(\frac{x}{\epsilon} \right)$ and $\delta_\epsilon(x) = \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + x^2}$.

As ϵ goes to zero, the approximations to the L^2 error term and H^1 regularity term converge to their un-regularized forms (by Lebesgue Dominated Convergence). For the length term, the following theorem is provided, see also [19].

Theorem 3.4. *Let $L_\epsilon(\phi,\psi) = \int_\Omega |\nabla H_\epsilon(\phi)| H_\epsilon(\psi)$ with H_ϵ satisfying the properties above and let ϕ and ψ be Lipschitz continuous; then*

$$\lim_{\epsilon\to 0} L_\epsilon(\phi,\psi) = \int_{\{\phi=0\}} H(\psi)ds = \text{Length}(\Gamma)$$

where $\Gamma := \{x \in \Omega \Big| \ \phi(x) = 0, \psi(x) > 0\}$.

Proof. Using the co-area formula (see [27]) and the fact that H_ϵ is smooth, the regularized length becomes:

$$\begin{aligned} L_\epsilon(\phi,\psi) &= \int_{\mathbb{R}} \left(\int_{\phi=\rho} \delta_\epsilon(\phi) H_\epsilon(\psi) ds \right) d\rho \\ &= \int_{\mathbb{R}} \delta_\epsilon(\rho) \left(\int_{\phi=\rho} H_\epsilon(\psi) ds \right) d\rho. \end{aligned}$$

Define $S_\epsilon(\rho) := \int_{\phi=\rho} H_\epsilon(\psi) ds$. By the scalability property of the delta function,

$$\begin{aligned} L_\epsilon(\phi,\psi) &= \int_{\mathbb{R}} \delta_\epsilon(\rho) S_\epsilon(\rho) d\rho \\ &= \int_{\mathbb{R}} \frac{1}{\epsilon} \delta_1\left(\frac{\rho}{\epsilon}\right) S_\epsilon(\rho) d\rho \end{aligned}$$

By the change of variable $p = \frac{\varrho}{\epsilon}$, we obtain

$$\begin{aligned} \lim_{\epsilon\to 0} L_\epsilon(\phi,\psi) &= \lim_{\epsilon\to 0} \int_{\mathbb{R}} \delta_1(p) S_\epsilon(\epsilon p) dp \\ &= S_0(0) \int_{\mathbb{R}} \delta_1(p) dp = S_0(0) \\ &= \int_{\phi=0} H(\psi) ds = \text{Length}(\Gamma), \end{aligned}$$

It is clear that $S_0(0) = \int_{\phi=0} H(\psi)ds$, since

$$\begin{aligned} S_\epsilon(\epsilon p) - S_0(0) &= \int_{\phi=\epsilon p} H_\epsilon(\psi)ds - \int_{\phi=0} H(\psi)ds \\ &= \left(\int_{\phi=\epsilon p} H_\epsilon(\psi)ds - \int_{\phi=\epsilon p} H(\psi)ds\right) - \left(\int_{\phi=\epsilon p} H(\psi)ds - \int_{\phi=0} H(\psi)ds\right) \\ &= \left(\int_{\phi=\epsilon p} \left(H_\epsilon(\psi) - H(\psi)\right) ds\right) - \left(\int_{\phi=\epsilon p} H(\psi)ds - \int_{\phi=0} H(\psi)ds\right) \end{aligned}$$

Since $H_\epsilon(x) \to H(x)$ in $\mathbb{R} \setminus \{0\}$ and since the length of the level sets are finite, the first term goes to zero, while the second term goes to zero by continuity of the integral. □

3.2.2 Relation to Other Models

In practice, the curve Γ can change its topology freely. Even if it is initialized with free endpoints, it may become an endpoint free curve, or vice versa. The standard splitting and merging behavior now includes breaking (or cracking) where, during its evolution, the curve can crack itself to develop endpoints. With this addition, our model can be viewed as a natural extension to other level set based methods for segmentation. Here we briefly examine the degeneration of the energy (and thus the curve evolution) into the classical models.

Firstly, the model can completely degenerate to the endpoint free structure when the indicator level set function has fixed sign, *i.e.* $\psi > 0$. The characteristic functions become:

$$\begin{aligned} \chi_1 &= H(\phi) \\ \chi_2 &= H(\psi)(1 - H(\phi)) = 1 - H(\phi) \\ \chi_3 &= H(\psi) = 1 \\ \chi_0 &= (1 - H(\psi))(1 - H(\phi)) = 0 \end{aligned}$$

and the reconstructed function takes the form:

$$\begin{aligned} u &= u_1\chi_1 + u_2\chi_2 + \left(\frac{u_1+u_2}{2}\right)\chi_0 \\ &= u_1 H(\phi) + u_2(1-H(\phi)) \end{aligned}$$

Thus the function is smooth in each of the two regions defined by the sign of ϕ. The energy becomes:

$$\begin{aligned} E(u_1,u_2,\phi,-) &= \mu\int_\Omega \left(|\nabla u_1|^2 H(\phi) + |\nabla u_2|^2(1-H(\phi))\right)dx + \lambda\int_\Omega |\nabla H(\phi)| \\ &\quad + \int_\Omega \left(|u_1-f|^2 H(\phi) + |u_2-f|^2(1-H(\phi))\right)dx \end{aligned}$$

which is the two-phase piecewise smooth model [74]. Furthermore, if the regularization parameter is set to a large value, $\mu >> 1$, or if u is restricted to the set of piecewise constant solutions, and where $u_1 = c_1$ and $u_2 = c_2$, and $c_1, c_2 \in \mathbb{R}$, then the reconstructed function becomes $u = c_1 H(\phi) + c_2(1-H(\phi))$ and the energy becomes:

$$E(c_1,c_2,\phi,-) = \lambda\int_\Omega |\nabla H(\phi)| + \int_\Omega \left(|c_1-f|^2 H(\phi) + |c_2-f|^2(1-H(\phi))\right)dx$$

which recovers the two-phase piecewise constant model [18].

3.3 Numerical Results for Image Segmentation

Since the model is non-convex and highly non-linear, an alternating minimization is used, where the energy is minimized with respect to each variable separately. During each minimization, three steps are done: calculate the Euler-Lagrange equation, find the Sobolev

gradient, and step forward in time.

To calculate the Euler-Lagrange equations, the PDEs are discretized using forward differences for the gradient and backwards differences for the divergence, in order to retain their adjoint relationship. The magnitudes of the gradient are replaced with a regularized version to avoid dividing by zero. With respect to the time discretization, the Euler-Lagrange equation is completely explicit and the inversion of the preconditioning operator (Sobolev gradient) is done using a semi-implicit method. Specifically, let $G = -\nabla_{L^2} E$ be the Euler-Lagrange equation and v equals the Sobolev gradient. Then we have the relationship from before: $v - \Delta v = G$, which we solve using a Gauss-Seidel sweep of the following discretization:

$$v_{i,j}^{n+1} - \left(v_{i+1,j}^{n} - 2v_{i,j}^{n+1} + v_{i-1,j}^{n}\right) - \left(v_{i,j+1}^{n} - 2v_{i,j}^{n+1} + v_{i,j-1}^{n}\right) = G_{i,j}$$

In practice, a few iterations are sufficient. Alternatively, v can be found using the Fourier transform: $v = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(G)}{1+|\xi|^2}\right)$. Once v is found, a forward Euler step is used to update the variable (recall that the time derivative of the variable is equal to the Sobolev gradient of the energy).

Next, with respect to the boundary conditions for the reconstructed image, if the auxiliary functions u_1 and u_2 are initialized to have Neumann boundary conditions at $t = 0$, then in semi-discrete terms the future boundary conditions are:

$$\begin{aligned} \chi_1 \frac{\partial}{\partial N} u_1^{n+1} + \frac{\chi_0}{4} \frac{\partial}{\partial N}(u_1^{n+1} + u_2^{n}) &= 0 \\ \chi_2 \frac{\partial}{\partial N} \nabla u_2^{n+1} + \frac{\chi_0}{4} \frac{\partial}{\partial N}(u_1^{n+1} + u_2^{n+1}) &= 0 \end{aligned}$$

Since the characteristic functions are assumed to be nonzero everywhere (based on our choice

of approximations):

$$\begin{aligned}\frac{\partial}{\partial N}u_1^{n+1} &= -\frac{\chi_0}{4\chi_1+\chi_0}\frac{\partial}{\partial N}u_2^n \\ \frac{\partial}{\partial N}u_2^{n+1} &= -\frac{\chi_2}{4\chi_2+\chi_0}\frac{\partial}{\partial N}u_1^{n+1}\end{aligned}$$

Since $\frac{\partial}{\partial N}u_1^0 = \frac{\partial}{\partial N}u_2^0 = 0$, they stay that way for all $n > 0$.

This process is iterated until convergence with respect to one variable is achieved, and then each of these alternating steps are repeated until total convergence. In terms of the level set functions, ϕ and ψ, only partial convergence is needed.

Lastly, once the updates for u_1 and u_2 are found in order to calculate the various differences across the curve Γ for the level set equations, each function needs to be extended. In general, any C^1 extension is appropriate; in particular, we solve $\Delta u_1 = 0$ in region 2 and $\Delta u_2 = 0$ in region 1 with prescribed boundary conditions (Dirchelet), and the extensions are labelled Eu_1 and Eu_2.

Algorithm: Free Curve Segmentation

Initialize $u_1^0, u_2^0, \phi^0, \psi^0$

while Not Converged **do**

Substep 1:	Compute $G_{u_1}(u_1^n, u_2^n, \phi^n, \psi^n)$, Solve for $v_{u_1}^n$, Iterate Forward to u_1^{n+1}
Substep 2:	Compute $G_{u_2}(u_1^{n+1}, u_2^n, \phi^n, \psi^n)$, Solve for $v_{u_2}^n$, Iterate Forward to u_2^{n+1}
Substep 3:	Extend u_1^{n+1} and u_2^{n+1} to Eu_1^{n+1} and Eu_2^{n+1}
Substep 4:	Compute $G_\phi(Eu_1^{n+1}, Eu_2^{n+1}, \phi^n, \psi^n)$, Solve for v_ϕ^n, Iterate Forward to ϕ^{n+1}
Substep 5:	Compute $G_\psi(Eu_1^{n+1}, Eu_2^{n+1}, \phi^{n+1}, \psi^n)$, Solve for v_ψ^n, Iterate Forward to ψ^{n+1}

end while

The convergence is typically measured by the difference in energy between the two iterations. Compared to the general level set based segmentation methods, this algorithm is more sensitive to initialization. In the standard methods, if the curve were to be initialized as a large circle that encloses all the objects, then it would shrink inward until it captured all the edges. In our case, if the curve was initialized as a large arc, it may shrink along the curve more quickly than inward (shrinking along the tangents rather than the normal vectors). In general, this can be controlled by the number of iterations in Substeps 4 and 5 or by initializing the curve to intersect the desired edges. This can be done in practice by over-segmenting the image using classical edge-detectors (for example, the Canny edge-detector) and then using the result to provide regions of interest for initializations. From this perspective, this segmentation algorithm can also be initialized using a result from another level set based method, for example, the methods from [18, 74, 14]. This may be useful to produce a better initial region of interest than using a classical edge-detector.

3.4 Further Remarks

In terms of the forward Euler step in our method, starting from the same data, the Sobolev gradient decreases the energy more than the L^2 descent.

Theorem 3.5. *Let ΔE_S and ΔE_{L^2} be the discrete changes in energy using Sobolev and L^2 gradient descent, respectively. Starting at the same value, if Δt_S and Δt_{L^2} are the time steps for the discretization of the Sobolev and L^2 gradient descent algorithm, respectively, then $\Delta E_S = \Delta E_{L^2} \frac{\Delta t_S}{\Delta t_{L^2}}$.*

Proof. Let $v = \nabla_{H^1} E(\phi)$ and $G := \nabla_{L^2} E(\phi)$ then we have $(I - \Delta)v = G$ in a weak sense with Neumann conditions. This equation is equivalent to the following weak formulation: for all $h \in H^1(\Omega)$

$$\int_\Omega [vh + \nabla v \cdot \nabla h]\, dx = \int_\Omega Gh\, dx$$

which is equivalent to:

$$\langle v, h \rangle_{H^1(\Omega)} = G(h) \tag{3.4}$$

where $G(h)$ is the linear form from the right hand side of the weak equation. Assuming that $G \in H^{-1}(\Omega)$, by the Riesz Representation theorem, there exists a unique $g \in H^1(\Omega)$ such that $G(h) = \langle g, h \rangle_{H^1(\Omega)}$ for all $h \in H^1(\Omega)$ (or equivalently by the Lax-Milgram theorem) and $||G||_{H^{-1}} = ||g||_{H^1}$. Combining this with the equation above yields $\langle v - G, h \rangle_{H^1} = 0$. Therefore we can see that $v = g$ *a.e.* and $||G||_{H^{-1}} = ||v||_{H^1}$

If we look at the ratio of changes in energy at a given iteration with Euler time steps then we have:

$$\frac{\Delta E_S}{\Delta E_{L^2}} = \frac{\Delta t_S ||v||^2_{H^1}}{\Delta t_{L^2} ||G||^2_{L^2}} = \frac{\Delta t_S}{\Delta t_{L^2}}$$

Note that since we assumed $G \in L^2$, we have that the H^{-1} norm is just the L^2 norm of G. □

In general, the preconditioned PDEs are more stable, which lets $\Delta t_S \geq \Delta t_{L^2}$, so that we will almost always get $\Delta E_S \geq \Delta E_{L^2}$. We see that not only is the Sobolev descent method better posed theoretically, it is also preferred numerically.

3.5 Experimental Results

We use time steps $\Delta t \in [0.01, 0.1]$, space steps $\Delta x = 1$, and $\epsilon = \Delta x$. Without the Sobolev gradient, Δt must be very small to guarantee stability, which in many cases incurs other numerical issues. The number of iterations in each minimization step for u_1 and u_2 is set to a maximum of about 150 (although they converge before reaching the maximal amount of iterations), while the level set minimization steps are set to a maximum of about 2-5 iterations. The algorithm converges between several seconds and a few minutes depending on the size of the image. In these examples, no re-initialization to the signed distance function is used for the level set function.

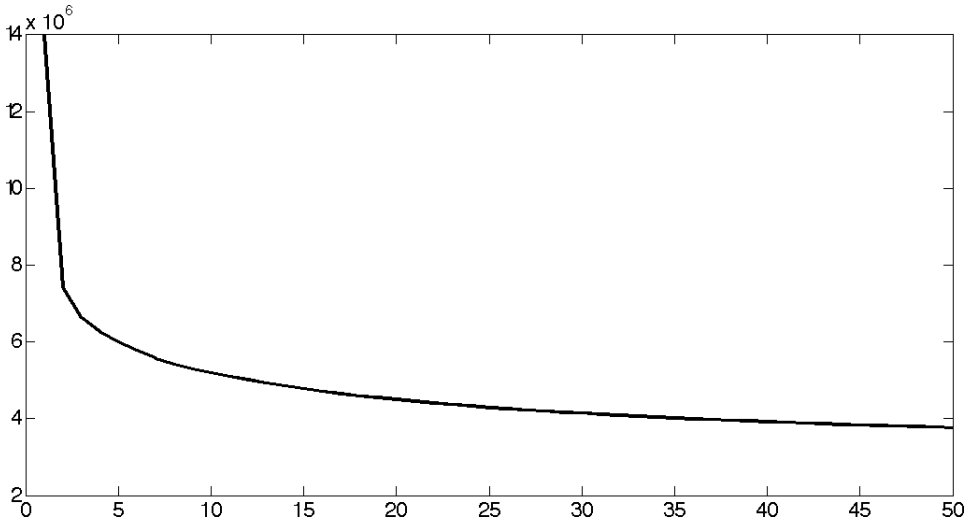


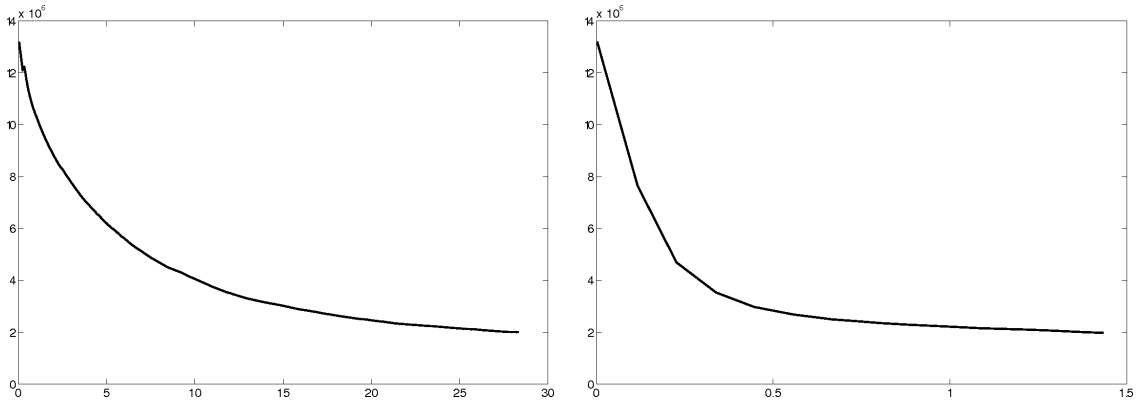
Figure 3.2: Plot of Energy versus Iteration using Sobolev Gradient descent

In Figure 3.2, we visualize the energy versus iterations corresponding to the experiment from Figure 3.2 and we can see that it is strictly decreasing. In Figure 3.3(a), we show the energy versus the CPU time corresponding to the experiment from Figure 8, as well as the energy versus CPU time for the equivalent L^2 gradient descent method. Both descent methods have the exact same initialization and are stopped when they have decreased the energy to below 2×10^6. The time steps are chosen so that the energy is strictly decreasing for all time, thus providing a measure of stability. The standard L^2 gradient descent needed 1084 iterations with a total time of 32.85 seconds to decrease the energy below the threshold, while the Sobolev descent needed 17 iterations at 1.75 seconds total. In general, the Sobolev descent trades longer time during each iterations for fewer iterations overall.

In Figure 3.4, the method is applied to a noisy synthetic image with an edge set composed of two free curves. The initial curve is made up of one segment, which first locates the edges and then separates into two segments (taking only a few iterations to change topology). In Figure 3.5, the method is applied to a very noisy synthetic image with an edge that has one endpoint which terminates at the boundary of the image and one endpoint that is free. The curve is initialized near the edge and the algorithm converges in seconds. In Figure 3.6, the method is applied to a noisy synthetic image composed of one segment with endpoints and

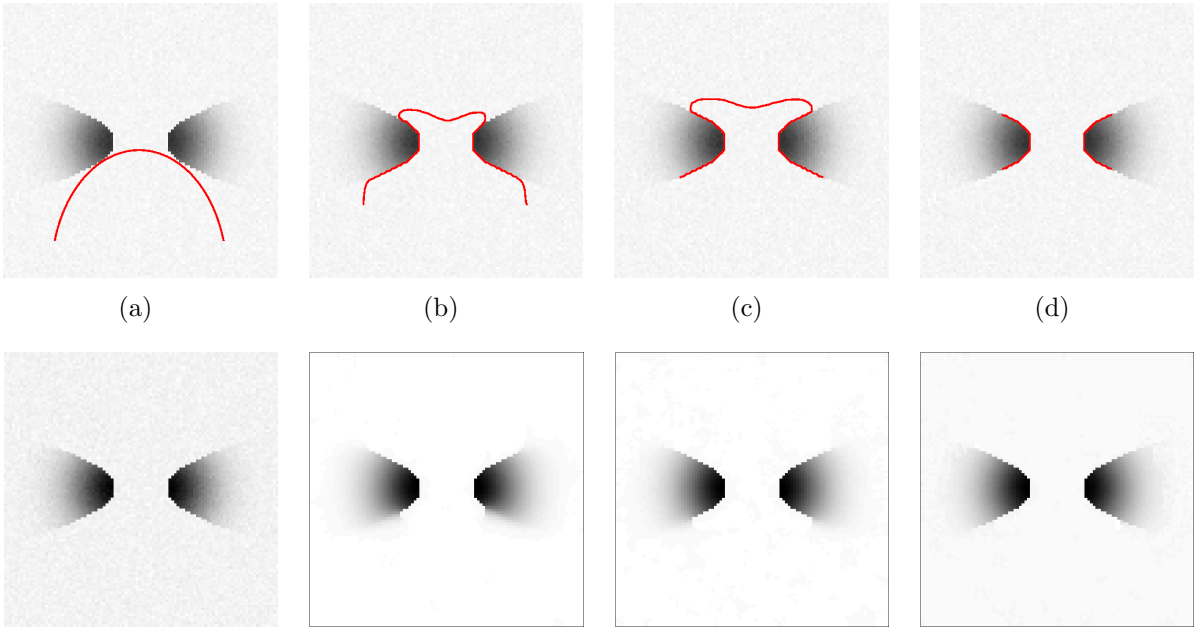
one without endpoints. The curve is initialized as two circles, but one changes its topology in order to capture the free endpoint structure. From these examples, the results depict the robustness of the algorithm to the various edge structures.

The method was applied to a photographic image (Figure 3.7), with close ups of two re-



(a) Standard Gradient Descent, $dt = 0.025$ (b) Sobolev Gradient Descent, $dt = 0.001$

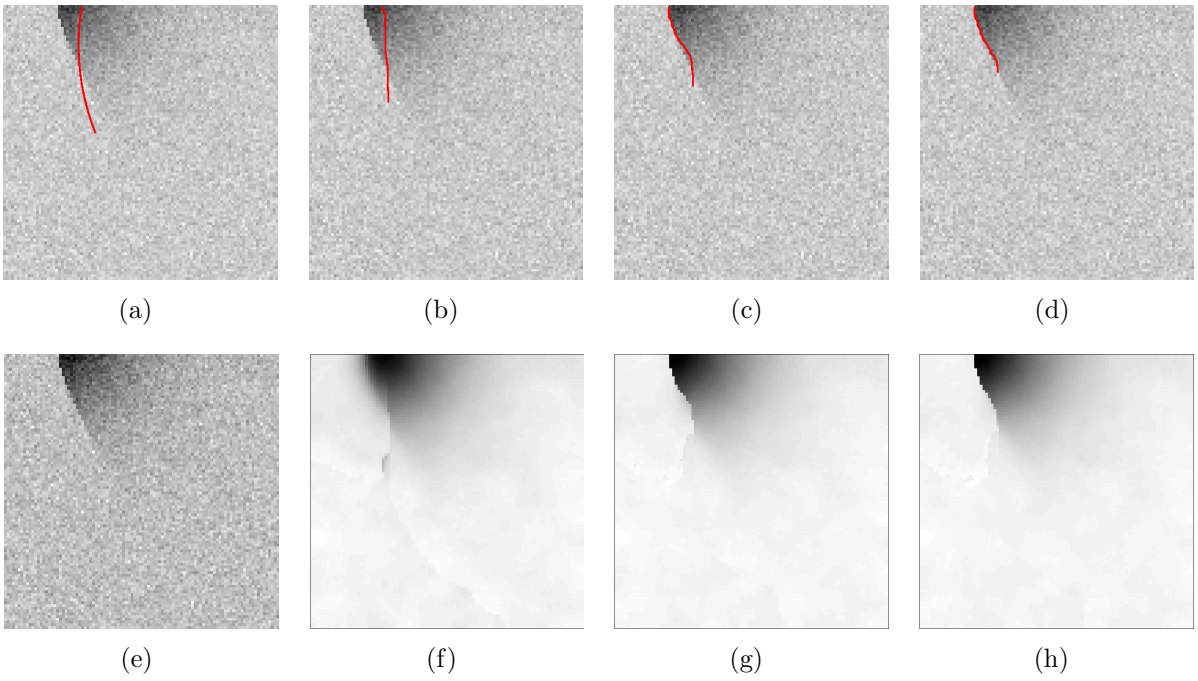
Figure 3.3: Plot of Energy versus CPU time using Sobolev gradient descent and the standard gradient descent



(a) (b) (c) (d)

(e) (f) (g) (h)

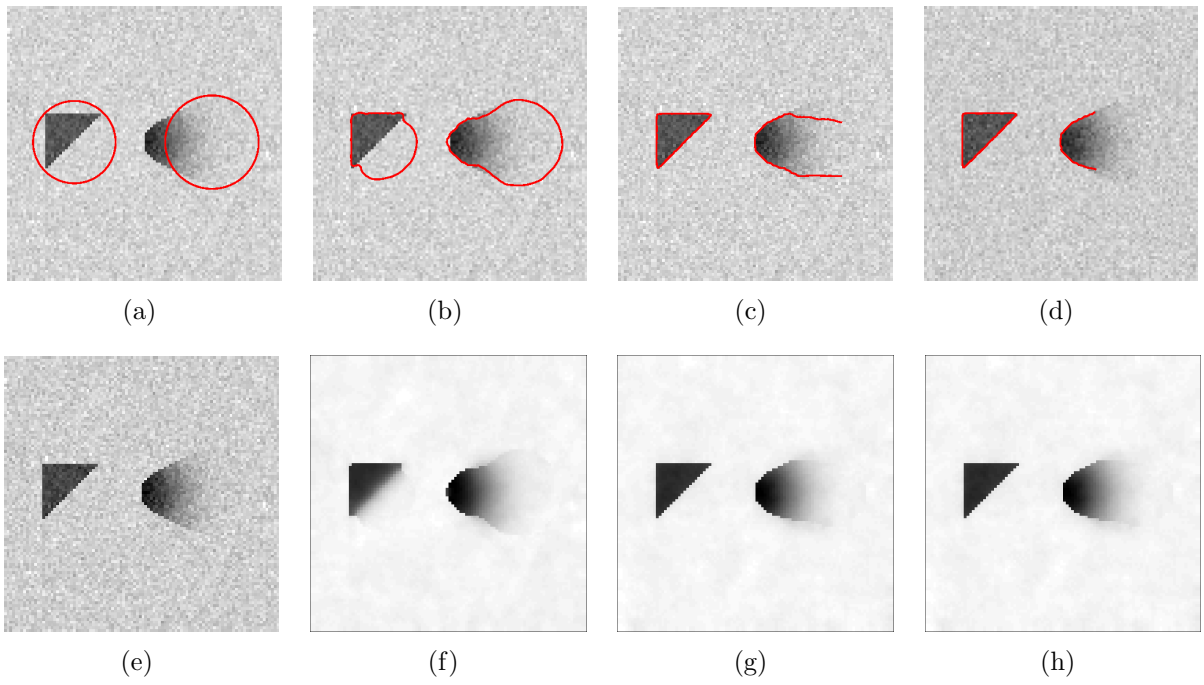
Figure 3.4: Segmentation and Restoration of a Synthetic Image with Two Free Edge Sets: the curve evolution (a-d) and the restoration (e-h) with $\mu = 4$ and $\lambda = 0.02 \times 255^2$.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.5: Segmentation and Restoration of a Synthetic Image with a Half Edge: $\mu = 17$ and $\lambda = 0.01 \times 255^2$.



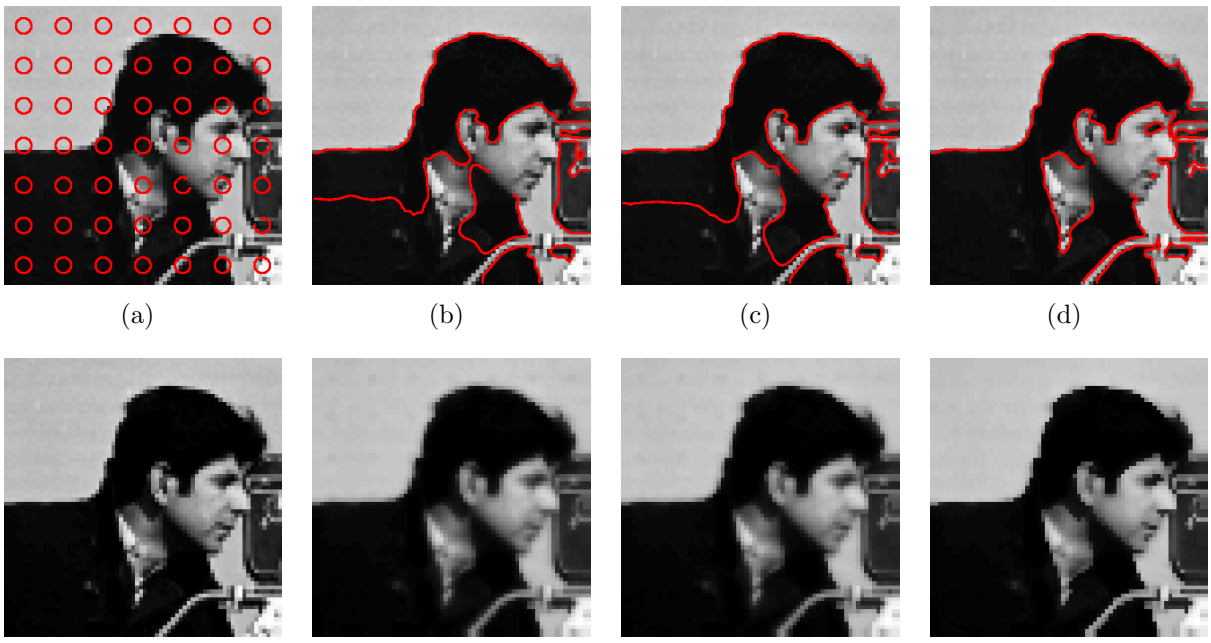
(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.6: Segmentation and Restoration of a Synthetic Image with Different Topologies: $\mu = 5$ and $\lambda = 0.05 \times 255^2$.

Figure 3.7: Original Cameraman image

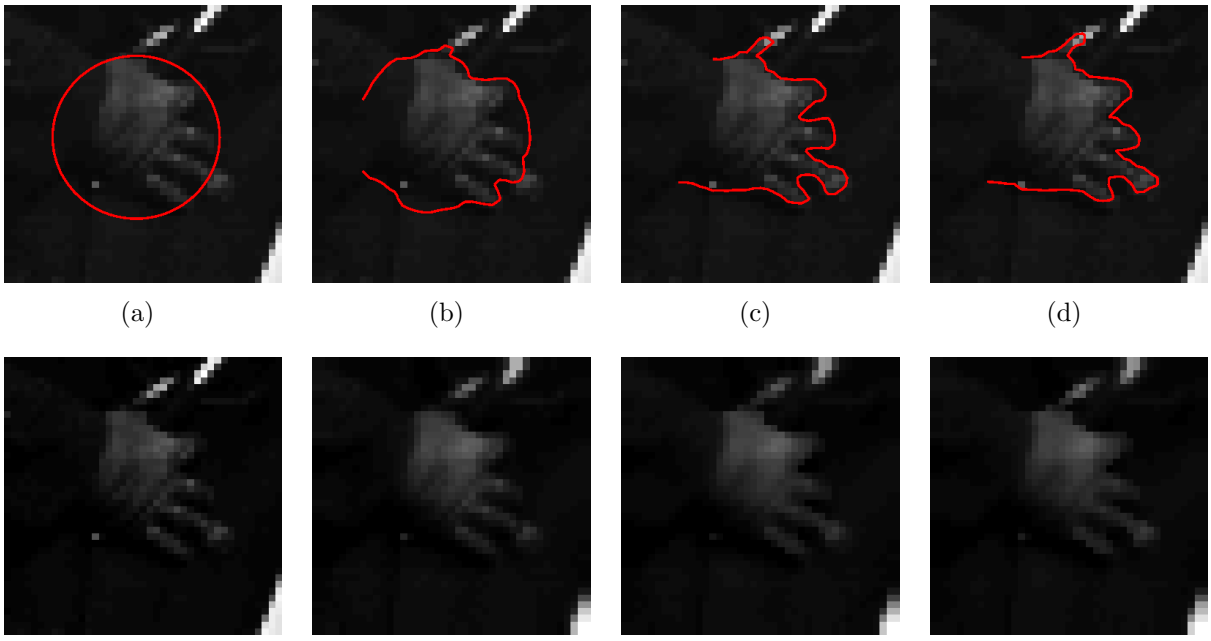
gions. In Figure 3.8, the curve is initialized as many small circles throughout the image. This particular initialization circumvents the need for manual location or an initial segmentation



(a) (b) (c) (d)

(e) (f) (g) (h)

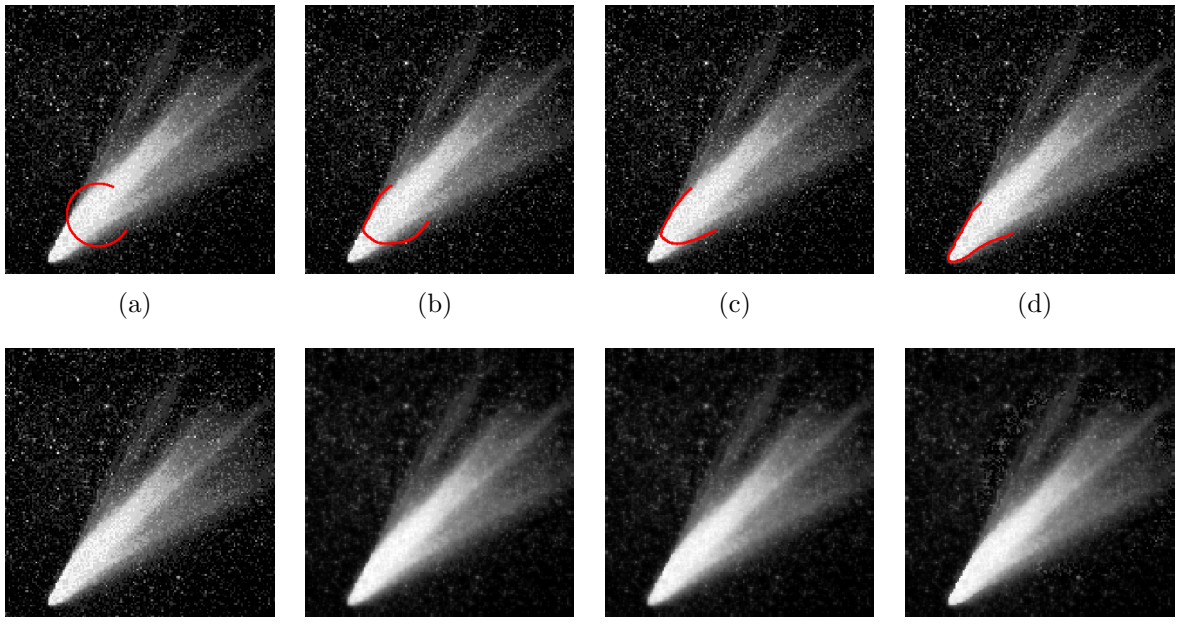
Figure 3.8: Segmentation of Cameraman, close up of the top region, $\mu = 0.5$ and $\lambda = 0.01 \times 255^2$



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3.9: Segmentation of Cameraman, close up of the middle region, $\mu = 0.5$ and $\lambda = 0.01 \times 255^2$

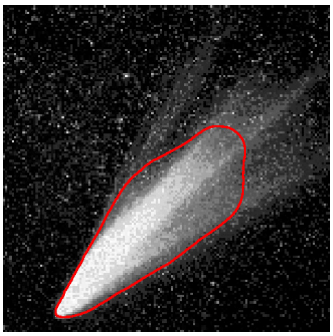


(a) (b) (c) (d)

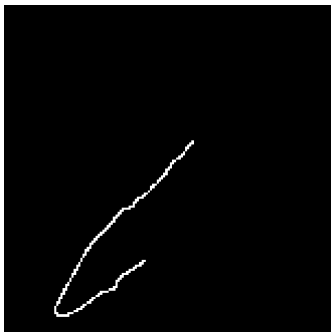
(e) (f) (g) (h)

Figure 3.10: Segmentation and Restoration of a Comet with Noise: $\mu = 0.5$ and $\lambda = 0.01 \times 255^2$

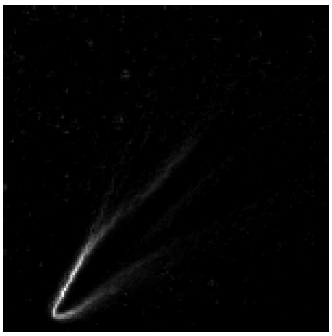
(a) Our Method



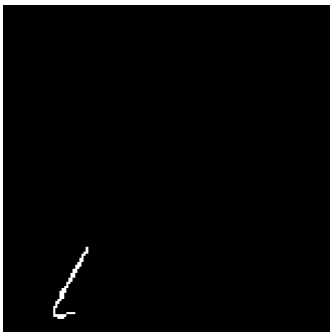
(b) Chan-Vese



(c) Canny Edge Detector Edge Image



(d) Ambrosio-Tortorelli

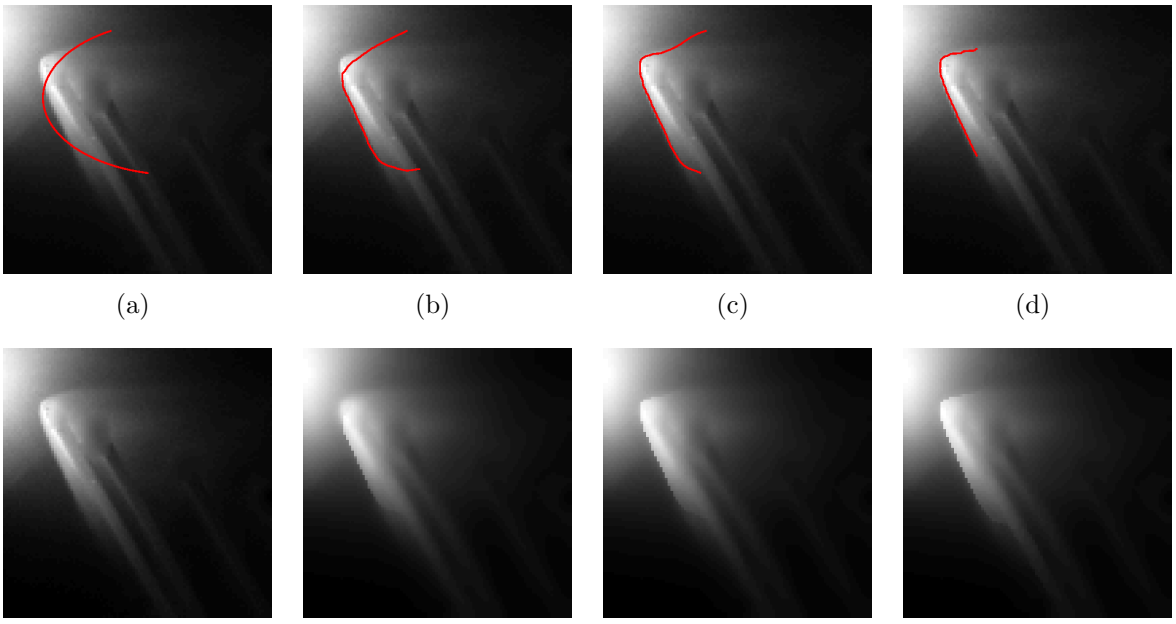


(e) Ambrosio-Tortorelli Edge Image after thresholding

Figure 3.11: Comparison of Segmentation Results: Comet Image

result in order to locate a region of interest. The final curve locates both free edges and contours without endpoints. The resulting reconstruction resembles a smooth cartoon form of the original with enhanced edges. The total computation time was about 9.77 seconds. In Figure 3.9, the curve is initialized as an endpoint free structure in the center of the image and quickly breaks topology to locate the hand. The resulting reconstruction sharpens edges and denoises. The total computation time was about 2.29 seconds.

The method is also applied to two astronomical images: one of a comet and the other of a plasma. In Figure 3.10, the curve locates the front of the comet, and the restored image sharpens the contrast between the comet and the background and removes noise from the comet, while preserving large stars (point structures) in the background. The total computation time was about 62.24 seconds. The final segmentation is compared to other techniques in Figure 3.11. The method from [18] locates the correct front, but over



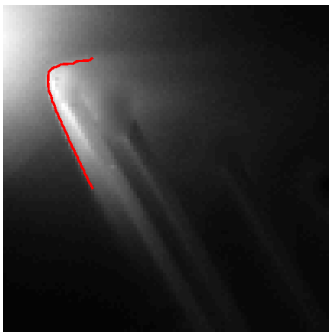
(a) (b) (c) (d)

(e) (f) (g) (h)

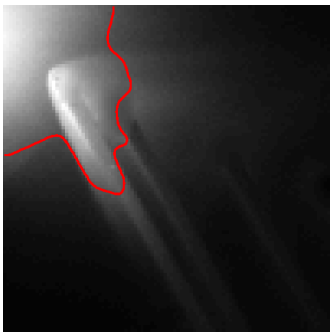
Figure 3.12: Segmentation and Restoration of a Plasma at UCLA [60]: $\mu = 25$ and $\lambda = 0.8 \times 255^2$

segments the comet, since it must be a loop. The Canny edge detector over-segments the white interior region of the comet, missing the faint boundary which defines the comet front. Similarly, the Ambrosio-Tortorelli method mainly locates the white region, where the gradient is sharpest. Based on these results, one possible application of this method could be to enhance and extract geometric information for astronomical classifications. Large scale geometric objects, such as galaxies or comets (depending on perspective) can be sharpened while removing noise and other forms of degradation.

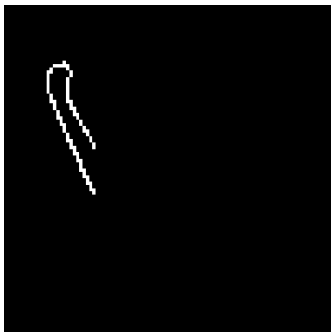
Lastly, we test our algorithm on a real plasma image. In Figure 3.12, the curve locates the plasma front and the restored image sharpens the contrast between the plasma front and the background, while removing the small amount of noise present in the image. The total computation time was about 47.01 seconds. This particular segmentation is made difficult by the light region in the top left quadrant near the plasma front. Region based methods would try to group the lighter intensities together, avoiding the actual edge. The final segmentation is compared to other techniques in Figure 3.13. The method from [18]



(a) Our Method

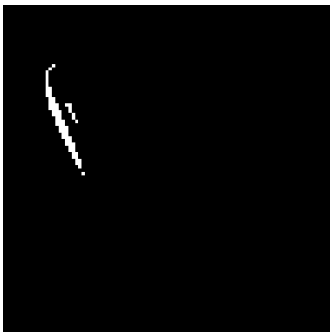


(b) Chan-Vese



(c) Canny Edge Detector Edge Image

(d) Ambrosio-Tortorelli Edge Image



(e) Ambrosio-Tortorelli Edge Image after thresholding

Figure 3.13: Comparison of Segmentation Results: Plasma Image

does not properly locate the edge, since the edge set does not enclose a region. The Canny edge detector does not locate the correct edge – locating places of high gradient inside of the plasma. Similarly, the Ambrosio-Tortorelli method does locate the front correctly but also includes excess edges, which cannot be removed by thresholding.

Part II

Texture Models and Analysis

CHAPTER 4

Introduction to Texture Modeling

One of the most important problems in image processing is the recovery of a corrupted image, f, which may be degraded by noise, blur, missing data, etc. The goal is to reconstruct important structural features of the original image, such as large scale objects (smooth regions), edges (discontinuities), textures (patterned small scale details), and noise (random and of mean zero). This problem is typically written as an inverse problem: given f, find a u which is a smooth approximation of f in some sense [2]. Specifically, given an f, decompose $f = u + v$, where u is the recovered image and v is the residual assumed to be noise. For such two component decompositions, the general formulation is:

$$\inf_{(u,v)\in X_1\times X_2} \mathcal{E}(u,v) = \{\mu||u||_{X_1} + ||v||_{X_2}\}$$
$$\text{s.t.} \quad f = u + v$$

where $\mu > 0$ is a tuning parameter. The recovered image is assumed to reside in the space X_1, which contains functions of positive differentiability (i.e. the "derivative" is well-behaved in some sense), and the noise is taken to be in $X_2 = L^2$. In this class of image recovery methods, notable models include Rudin, Osher, and Fatemi (ROF) [67] and Chambolle and Lions [16]. In particular, recall the ROF model below:

$$\inf_{(u,v)\in BV\times L^2} \mathcal{E}_{ROF}(u,v) = \left\{ \mu||u||_{TV} + \frac{1}{2}||v||^2_{L^2} \right\} \quad \text{s.t.} \quad f = u + v \tag{4.1}$$

which is known to reconstruct piecewise constant solutions very well, thereby recovering both large scaled features and edges. With the introduction of the Bregman technique [62] and the split Bregman method [37], total variation regularized problems can be solved quickly and efficiently while also remedying defects such as loss of contrast.

On the other hand, the texture is considered to be highly oscillatory, which is not well captured by the X_1 norms. Thus in many of these reconstruction models, v also contains some textures and edges. In order to further decompose the texture and noise, texture regularized models have appeared in the literature, beginning with the work of Meyer [55], who first proposed recovering texture using spaces that are weaker than L^2. These weak spaces encourage oscillatory behavior, since their norms decrease as the amount of oscillations increase. In this framework, the image is decomposed into the cartoon component and the texture component, while also removing any noise or residual. This becomes a three component decomposition: given f, decompose $f = u + v + \rho$ by solving:

$$\inf_{(u,v)\in BV\times T} \mathcal{E}_{C-T}(u,v) = \mu||u||_{TV} + \gamma||v||_T + \frac{1}{2}||u + v - f||^2_{L^2} \tag{4.2}$$

The cartoon component u is appropriately modeled by the space BV, using the TV semi-norm in the energy. The recovered image is now $u + v$, the sum of the cartoon and texture components. The residual, ρ, is assumed to reside in L^2, although other L^p norms can be used. Meyer proposed several texture norms in his work; most notably, the G-norm (the predual of BV, *i.e.* $G^* = BV$) has seen much success in applications, including: color image restoration [6], road detection [35], denoising [34], inpainting [9], image classification

[5], and anomaly removal and pattern regularization [33]. The space is defined below.

Definition 4.1. The space G consists of all distributions v which can be written as $v = \text{div}(\vec{g})$, where $\vec{g} = (g_1, g_2)$ and $g_1, g_2 \in L^\infty$. The norm on this space is defined as:

$$||v||_G := \inf \left\{ ||\sqrt{g_1^2 + g_2^2}||_\infty \;\middle|\; v = \text{div}(\vec{g}) \right\} \tag{4.3}$$

This texture space is particularly elegant because of its symmetry (specifically duality) with the cartoon space BV. However, this space is difficult to handle numerically. In [75], Vese and Osher proposed a method to approximate the G-norm by using the Sobolev spaces of negative differentiability, defined as $W^{-1,p} := \left\{ v = \text{div}\vec{g} \;\middle|\; g_1, g_2 \in L^p \right\}$, and sending $p \to \infty$. Rather than working with the texture v directly, this method works with the vector-field $\vec{g}$:

$$\inf_{(u,g) \in BV \times (L^p)^2} \mathcal{E}_{OV}(u, g) = \mu ||u||_{TV} + ||\vec{g}||_{L^p} + \frac{\lambda}{2} ||u + \text{div}\vec{g} - f||_{L^2}^2 \tag{4.4}$$

This formulation yields appropriate Euler-Lagrange Equations for all $p \geq 1$, which is easy to compute numerically. The results give satisfactory decompositions, particularly in the case of $p = 1$. Osher, Sole, and Vese [64] proposed an alternative norm for the case of $p = 2$, which coincides with the space H^{-1} (the dual of the Hilbert space H^1). This dual space has an explicit norm defined as: $||v||_{H^{-1}} = ||\nabla \Delta^{-1} v||_{L^2}$, and the resulting Euler-Lagrange equation of the model reduces to a forth order non-linear partial differential equation. This texture norm was also used in cartoon-texture-edge set separation in [70]. Later, Aujol, Aubert, Blanc-Feraud, and Chambolle [4] used projections to solve the original $BV - G$ decomposition model and applied it to denoising of cartoon images (where the G norm captures the noise) and to cartoon-texture separation. These weak spaces inspired many other texture norms, mainly in the form of dual spaces. In Lieu and Vese [49], the negative Hilbert spaces H^{-s}, for $s > 0$, which are dual to the Hilbert-Sobolev spaces $(H^s)^*$, were used for texture extraction and noise removal. Later, in Kim and Vese [41], the negative

exponential Sobolev spaces $W^{\alpha,p}$, for $-2 \leq \alpha < 0$, which are dual to the Sobolev spaces with pseudo-derivatives, were used for texture reconstruction in the presence of blur. We provide a table of spaces below to summarize some of these models.

Model	Texture	Description
[55]	G	$G^* = BV$, divL^∞, $W^{-1,\infty}$
[31]	F	div BMO
[55]	E	div$B_\infty^{-1,\infty}$
[76]	G_p	divL^p
[64]	H^{-1}	$H^{-1} = (H^1)^*$
[49]	$H^{-s}, s > 0$	$H^{-s} = (H^s)^*$
[41], [31]	$\Delta W^{\alpha+2,p}$, $-2 \leq \alpha < 0$,	$W^{\alpha,p}$

Note: $p \in (1, \infty)$

CHAPTER 5

A texture model based on a concentration of measure

As in [23, 41], a general degradation model using both the cartoon and texture component takes the form of:

$$f = K(u_0 + v_0) + \eta$$

and the corresponding optimization becomes:

$$\inf_{(u,v)\in X_1\times X_2} \mathcal{E}(u,v) = \left\{\mu||u||_{X_1} + \gamma||v||_{X_2} + ||f - K(u+v)||^2_{L^2}\right\}.$$

As mentioned before, the cartoon space X_1 is typically assumed to be a subset of BV with the TV semi-norm, but there is no standard choice for the texture space X_2. Most of the choices for X_2 in the literature are Banach space duals of spaces defined by some positive degree of differentiability, and as the degree of differentiability is increased, the dual norms of highly oscillatory functionals are decreased, which makes $||\cdot||_{X_2}$ favor high oscillations more.

Based on the literature and numerical experiments, the weaker the texture space, the finer the scale of details recovered. In this section, we use a texture norm that is related to the dual of the space of Lipschitz functions. There is evidence in the literature that supports this choice. In particular, the space G_1 from [76] is contained in our texture space. However, our norm is more isotropic and thus yields different minimizers numerically. We also make use of a substitution for the space of Lipschitz functions and use its dual to define the texture space. Numerically, this substitute norm is identical to the Lipschitz norm and in practice results in equations that are easier to handle than the standard substitute of taking $W^{1,p}$

for p large (curvature or diffusion equations rather than the highly non-linear p-Laplacians). The texture regularization is used to deblur images with unknown blur kernels by means of a semi-known variational approach. The original blurring kernel may be noisy, and thus one can only approximate the original kernel.

Since this texture norm is very weak (in the sense that the underlying function space is very large), the reconstructed images contain more fine-scale details. This is typically difficult to recover from blurred images. The addition of the texture regularization to the semi-known variational deblurring model helps support the approximation of the kernel. For semi-known deblurring, the method settles to the correct kernel quickly, even when the kernel contains noise. But before describing the deblurring model, we will first recall several definitions and preliminary details.

5.1 Terminology

Let Ω be an open, bounded and connected subset of $\mathbb{R}^n$, with Lipschitz boundary. For our cartoon space, we will use the space of bounded variation.

Definition 5.1. A function $u : \Omega \to \mathbb{R}$ is of bounded variation if and only if $u \in L^1(\Omega)$ and there exists a finite $\mathbb{R}^n$ valued Radon measure Du such that for all $\phi \in C_c^1(\Omega, \mathbb{R})$,

$$\int_\Omega \frac{\partial \phi}{\partial x_j}\, u \, dx = -\int_\Omega \phi \, (Du)_j, \quad \text{for } j = 1, ..., n. \tag{5.1}$$

and the TV semi-norm

$$||u||_{TV} = \int_\Omega |Du| := \sup \left\{ \int_\Omega u \, \operatorname{div}(\phi)\, dx \;\middle|\; \phi \in C_c^1(\Omega, \mathbb{R}^n), ||\phi||_{L^\infty} \leq 1 \right\} \tag{5.2}$$

is bounded.

The norm on the Banach space $BV(\Omega)$ is:

$$||u||_{BV(\Omega)} := ||u||_{L^1(\Omega)} + ||u||_{TV(\Omega)} \tag{5.3}$$

The Sobolev space $W^{1,1}(\Omega)$ is contained in $BV(\Omega)$, with the standard distributional derivative, *i.e.* $Du = \nabla u$ dx.

For our texture space, we would like to capture the highly oscillatory nature of texture, while leaving the cartoon component without any fine-scale detail. This is accomplished by choosing a strong auxiliary space and then constructing a very weak dual space. The resulting dual space will capture the behavior of texture, since it is well known that these types of dual norms decrease when the frequencies increase. One particular choice (supported by [41, 76]) would be the dual of $W^{1,\infty}(\Omega)$ (with semi-norm $||\nabla u||_{L^\infty}$) for texture.

Definition 5.2. A function $v : \Omega \to \mathbb{R}$ belongs to the dual of $W^{1,\infty}(\Omega)$ if v is of zero mean and the following quantity is finite:

$$||v||_{(W^{1,\infty})^*} := \sup \left\{ \langle v, w \rangle \;\middle|\; \|\nabla w\|_{L^\infty} \leq 1 \right\}. \tag{5.4}$$

where $\langle v, w \rangle = \int vw \; dx$.

Formally, given a function v, one way to calculate $||v||_{(W^{1,\infty})^*}$ is to find w which achieves the maximum (assuming this is possible) of equation (5.4). Unfortunately, the resulting equations may not be well-posed, due to the appearance of the infinity Laplacian (for more on the infinity Laplacian see [50, 21, 28]). Alternatively, one can approximate the infinity norm as the following limit:

$$||\nabla w||_{L^\infty} = \lim_{p \to \infty} ||\nabla w||_{L^p}.$$

In terms of the dual norms, this process results in approximating the space $(W^{1,\infty}(\Omega))^*$ by the spaces $W^{-1,q}(\Omega)$. The resulting maximizing equation in terms of the auxiliary variable w is the q-Laplacian, which is both non-linear and numerically stiff.

Instead we propose approximating the $L^\infty(\Omega)$ norm using the $L^1(\Omega)$ norm with respect to a measure which concentrates near the maximum:

$$||\nabla w||_{L^\infty} = \max_x |\nabla w(x)| = \int |\nabla w(x)| d\delta_{\tilde{x}},$$

where $\tilde{x} = \text{argmax } |\nabla w(x)|$ (assuming this is attained). There are many choices for approximations that converge to the Dirac delta function; however, we propose the following:

Definition 5.3. For $k > 0$ we define the measure $d\sigma_k := \sigma_k dx$ where

$$\sigma_k := \frac{e^{k|\nabla w|}}{\int e^{k|\nabla w|} dx} .$$

In [28], a similar formulation is given when the L^1 norms above are replaced by the L^2 norms with the corresponding measure $\frac{e^{k|\nabla w|^2}}{\int e^{k|\nabla w|^2} dx}$. In practice, the L^2 norm does not provide satisfactory results, because (for finite k) it yields overly smooth textures. With the measure in Definition 2.3, we define our approximate norms as follows.

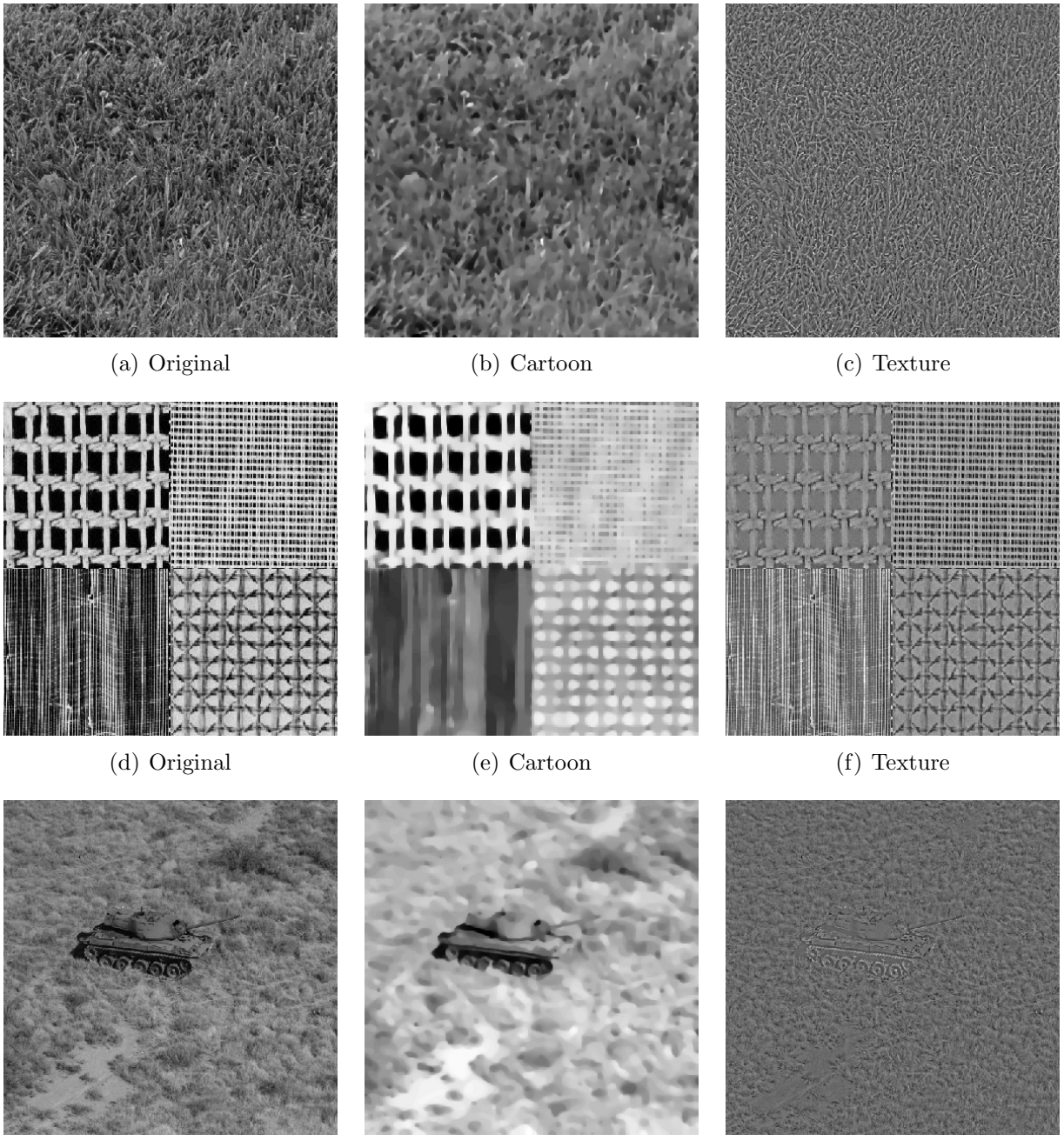
Definition 5.4. A function $v : \Omega \to \mathbb{R}$ belongs to T_k if and only if v is of zero mean and the following quantity is finite

$$||v||_{T_k} := \sup_{w \in W^{1,\infty}} \left\{ \langle v, w \rangle \;\middle|\; \int_\Omega |\nabla w| d\sigma_k(w) \leq 1 \right\}. \tag{5.5}$$

where k is fixed. For notational purposes, $T := T_\infty = (W^{1,\infty})^*$ and the "k-norms" are defined as $||\nabla w||_k = \int_\Omega |\nabla w| d\sigma_k(w)$.

5.2 The Model

We propose a semi-known deblurring model which is regularized using the BV space for the cartoon and the T_k space (for all k) for the texture. Similar work using cartoon-texture regularized deblurring has appeared in the literature, but only for the case when the kernel is fixed (see [23, 41]). For an example of a semi-blind deblurring model related to our model



(a) Original (b) Cartoon (c) Texture

(d) Original (e) Cartoon (f) Texture

(g) Original (h) Cartoon (i) Texture

Figure 5.1: Cartoon-Texture Separation of Grass, Brodatz, and Tank Images. In each case, the Cartoon component (b), (e), and (h) has sharp contrast with homogenous regions.

below, see [8].

The general model is as follows:

$$\inf_{u,v,K}\left\{E_{general}(u,v,K)=\mu||u||_{TV}+\gamma||v||_{T}+\frac{1}{2}||K*(u+v)-f||_2^2+\beta||\nabla K||_p^p\right\}, \quad (5.6)$$

where K is a blur kernel. We assume that the kernel belongs to a parametrized family, say $K(\alpha)$, thus the problem becomes:

$$\inf_{u,v,\alpha}\left\{E(u,v,\alpha)=\mu||u||_{TV}+\gamma||v||_{T}+\frac{1}{2}||K(\alpha)*(u+v)-f||_2^2+V(\alpha)\right\}, \quad (5.7)$$

where $V(\alpha):=\beta||\nabla K(\alpha)||_p^p$ is a regularizer only dependent on the parameter α. Replacing the texture norm by a ratio results in the following saddle point problem:

$$\inf_{u,v,\alpha}\sup_{w}\left\{\mu||u||_{TV}+\gamma\frac{|\langle v,w\rangle|}{\int_\Omega|\nabla w|d\sigma}+\frac{1}{2}||K(\alpha)*(u+v)-f||_2^2+V(\alpha)\right\}.$$

In this form, we regularize the gradient $|\nabla w|:=\sqrt{w_x^2+w_y^2+1}$, so that we may take derivatives. Also, this breaks the 1-homogeneity in k of the ratio between the L^2 inner product and the functional measure on ∇w.

5.3 Theoretical and Analytical Remarks

The texture space $T=(W^{1,\infty})^*$ is larger than many of the spaces appearing in the literature. In practice, this means the texture component is able to capture a wider range of fine scale details and oscillatory patterns. The following lemma provides some insight to the relationship between T and other spaces.

Lemma 5.5. *The following inclusions hold:* $\Delta W^{1,1}\subsetneq G_1\subset(W^{1,\infty})^*$.

Proof. First let $v = \text{div } g$ where $g \in L^1 \times L^1$, then

$$\begin{aligned} ||v||_{(W^{1,\infty})^*} &= \sup_{||\nabla w||_{L^\infty} \leq 1} \left| \int_\Omega vw \ dx \right| \\ &= \sup_{||\nabla w||_{L^\infty} \leq 1} \left| \int_\Omega \text{div} gw \ dx \right| \\ &= \sup_{||\nabla w||_{L^\infty} \leq 1} \left| \int_\Omega g \cdot \nabla w \ dx \right| \\ &\leq \sup_{||\nabla w||_{L^\infty} \leq 1} ||g||_{L^1} ||\nabla w||_{L^\infty} \\ &\leq ||g||_{L^1} \end{aligned}$$

which is bounded by assumption. So $G_1 \subset (W^{1,\infty})^*$.

Next, to see that $\Delta W^{1,1} \subset G_1$, take $v \in \Delta W^{1,1}$. Then there exists an $h \in W^{1,1}$ such that $v = \Delta h$. Therefore, $v = \text{div } \nabla h$ with $\nabla h \in L^1 \times L^1$ so $v \in G_1$. However, the converse is not true. To show this, let $v \in G_1$. There exists a $g \in L^1 \times L^1$ such that $v = \text{div} g$, but it is not true that g must be a gradient of a function in L^1, thereby establishing $\Delta W^{1,1} \subsetneq G_1$. □

Of course, the space T itself is quite large; however, it will be clear in the following sections that the minimizers are well-behaved and our approximation T_k for sufficiently large k, yields appropriate results.

5.3.1 Behavior with respect to k

In this section, some results provide information on the model's dependence on the parameter k. Both the k-norms and the T_k norms have a proper ordering.

Proposition 1. *The functionals $||-||_k$ are increasing functions of k, strictly increasing over the class of non-constant functions.*

Proof. For any function f, the derivative of $||f||_k$ with respect to k is the following:

$$\begin{aligned}
\frac{d}{dk}||f||_k &= \frac{d}{dk}\left(\frac{\int |f|e^{k|f|}dx}{\int e^{k|f|}dx}\right) \\
&= \frac{\left(\int |f|^2 e^{k|f|}dx\right)\left(\int e^{k|f|}dx\right) - \left(\int |f|e^{k|f|}dx\right)^2}{\left(\int e^{k|f|}dx\right)^2} \\
&\geq 0
\end{aligned}$$

by the Cauchy-Schwarz inequality. Equality holds if and only if $|f|$ is constant. Note that the arguments above hold for the regularized absolute value as well. □

The following corollaries are a result of the increasing behavior of the k norms.

Corollary 5.6. *The norms $||-||_{T_k}$ are strictly decreasing with respect to k.*

Corollary 5.7. *The functionals $||-||_k$ converge from below to the L^∞ norm as $k \to \infty$.*

The proof is a direct consequence of Corollary 5.6 and the fact that the L^∞ norm is a sharp upper bound. These corollaries allow one to examine the convergence of the minimal energy values as functions of k.

Theorem 5.8. *Let $P_k = \inf_{u,v}\left\{\mu||u||_{TV} + \gamma||v||_{T_k} + \frac{1}{2}||u+v-f||_2^2\right\}$ for fixed $\mu, \gamma > 0$. Then the following holds.*

1. *$P_{k+1} < P_k$ for $k > 0$.*
2. *$P_{k+1} \to P_\infty$ as $k \to \infty$, where $P_\infty = \inf_{u,v}\left\{\mu||u||_{TV} + \gamma||v||_T + \frac{1}{2}||u+v-f||_2^2\right\}$*

Proof. For the inequality, assume (u_k, v_k) are minimizers of the functional for a given k, then

$$\begin{aligned} P_k &= \inf_{u,v}\left\{\mu||u||_{TV} + \gamma||v||_{T_k} + \frac{1}{2}||u+v-f||_2^2\right\} \\ &= \mu||u_k||_{TV} + \gamma||v_k||_{T_k} + \frac{1}{2}||u_k+v_k-f||_2^2 \\ &> \mu||u_k||_{TV} + \gamma||v_k||_{T_{k+1}} + \frac{1}{2}||u_k+v_k-f||_2^2 \\ &\geq \mu||u_{k+1}||_{TV} + \gamma||v_{k+1}||_{T_{k+1}} + \frac{1}{2}||u_{k+1}+v_{k+1}-f||_2^2 \\ &= P_{k+1} \end{aligned}$$

by Corollary 5.6.

For the second consequence, note that the sequence is strictly decreasing with the sharp lower bound

$$P_k \geq \inf_{u,v}\left\{\mu||u||_{TV} + \gamma||v||_{T} + \frac{1}{2}||u+v-f||_2^2\right\}.$$

□

5.3.2 Existence, Uniqueness, and Characterization of Minimizers

For regularized models of the form presented in this work, the minimizers always enjoy the property that $v \in L^2(\Omega)$. This provides sufficient regularity to the texture component to avoid unwanted elements from either the T or T_k space. Although both methods share this regularity, the minimizers of their respective energies will not be the same in general. In this section, we focus on

$$\inf_{u,v,u+v=f}\left\{E_1(u,v) = \mu||u||_{TV} + \gamma||v||_{T_k}\right\}, \tag{5.8}$$

and

$$\inf_{u,v}\left\{E_2(u,v) = \mu||u||_{TV} + \gamma||v||_{T_k} + \frac{\lambda}{2}||u+v-f||_2^2\right\}. \tag{5.9}$$

Both functionals above are cases of our general model. In the constrained cartoon-texture decomposition, the theorem below guarantees existence and uniqueness.

Theorem 5.9. *For $f \in L^2(\Omega)$, equation (5.8) with constraint $f = u + v$ and $\int_\Omega f\ dx = \int_\Omega u\ dx$ has a solution $(u, v) \in BV \times T_k \cap L^2$.*

Proof. Let $\{(u_n, v_n)\}$ be a minimizing sequence, then there exists a constant C such that:

$$\begin{aligned} |u|_{BV} &\leq C \\ ||v||_{T_k} &\leq C \\ f &= u_n + v_n \\ \int_\Omega f\ dx &= \int_\Omega u_n\ dx \end{aligned}$$

Then by the estimate above and by the Poincare-Wirtinger inequality (for some constant $c > 0$),

$$\begin{aligned} ||u_n||_{L^2} &\leq \left\| u_n - \int_\Omega u_n\ dx \right\|_{L^2} + \left| \int_\Omega u_n\ dx \right| \\ &\leq c|u_n|_{BV} + \left| \int_\Omega f\ dx \right| \\ &\leq C \end{aligned}$$

with redefinition of C. This provides a uniform bounded for u_n with respect to the L^2 norm. Therefore, $||u_n||_{BV} \leq C$ uniformly, thus there exists $u \in BV$ and a subsequence u_n (with possible reindexing) which converges to u in the BV-weak * topology and strongly in the L^1 topology.

For v_n, since u_n and f are uniformly bounded in the L^2 norm, then $||v_n||_{L^2} \leq C$. Also if $||w||_k \leq C$ then $||w||_{L^2} \leq C$, then

$$\int_\Omega w v_n\ dx \to \int_\Omega w v\ dx$$

so v_n converges weakly in L^2 and in the weak * T_k topology. In terms of the texture norm, the following holds:

$$\begin{aligned}
||v||_{T_k} &= \sup_{w \in W^{1,\infty}} \left\{ |\langle v, w \rangle| \;\middle|\; \int_\Omega |\nabla w| d\sigma_k(w) \leq 1 \right\} \\
&\leq \lim_{n\to\infty} \sup_{w \in W^{1,\infty}} \left\{ |\langle v_n, w \rangle| \;\middle|\; \int_\Omega |\nabla w| d\sigma_k(w) \leq 1 \right\} \\
&= \lim_{n\to\infty} ||v_n||_{T_k}.
\end{aligned}$$

Similarly, by lower semi-continuity $|u|_{BV} \leq \liminf_{n\to\infty} |u_n|_{BV}$. Lastly, note that the constraints are also held in the limit and $f = u + v$ a.e.. Thus $E_1(u, v) \leq \liminf_{n\to\infty} E_1(u_n, v_n)$ and (u, v) are minimizers. □

Theorem 5.10. *For $f \in L^2(\Omega)$, equation (5.9) with the constraint $\int_\Omega f \; dx = \int_\Omega u \; dx$ has a solution $(u, v) \in BV \times T_k \cap L^2$, with uniqueness for $\int_\Omega f \; dx \neq 0$.*

Proof. Once again, let $\{(u_n, v_n)\}$ be a minimizing sequence, then there exists a constant C such that the following hold:

$$\begin{aligned}
|u|_{BV} &\leq C \\
||v||_{T_k} &\leq C \\
||u_n + v_n - f||_{L^2} &\leq C \\
\int_\Omega u_n \; dx &= \int_\Omega f \; dx
\end{aligned}$$

Using the estimate above and the Poincare-Wirtinger inequality, as before $||u_n||_{L^2} \leq C$. Therefore, $||u_n||_{BV} \leq C$ uniformly and so there exists $u \in BV$ and a subsequence u_n (with possible reindexing) which converges to u in the BV-weak * topology and strongly in the L^1 topology.

Since u_n is uniformly bounded with respect to the L^2 norm, so is v_n, *i.*e. $||v_n||_{L^2} \leq C$. Next, we wish to show that v_n limits to $v \in (T_k)^* \cap L^2$. For any w with $||w||_k \leq C$, $||w||_{L^2} \leq C$

holds, so

$$\int_\Omega wv_n \, dx \to \int_\Omega wv \, dx$$

therefore v_n converges weakly in L^2 and in the weak * T_k topology. By the lower semicontinuity of the respective norms, we have

$$\begin{aligned} |u|_{BV} &\leq \liminf_{n\to\infty} |u_n|_{BV} \\ ||v||_{T_k} &\leq \lim_{n\to\infty} ||v_n||_{T_k} \\ ||u+v-f||_{L^2} &\leq \lim_{n\to\infty} ||u_n+v_n-f||_{L^2} \end{aligned}$$

Thus $E_2(u,v) \leq \liminf_{n\to\infty} E_2(u_n,v_n)$ and (u,v) are minimizers. For uniqueness, see [47]. □

Theorem 5.11. *If $f \in L^2(\Omega)$, $\int_\Omega f \, dx \neq 0$, and (u_λ, v_λ) is the the solution of equation (5.9), then as $\lambda \to \infty$, $u_\lambda + v_\lambda \to f$ strongly in L^2 and the pair (u_λ, v_λ) converges to the (u,v) pair which is the minimizer of (5.8).*

Next, the minimizers can be characterized by the dual norm of the entire regularization. Let $||(u,v)||_* = \inf\{\mu||u||_{TV} + \gamma||v||_{T_k}\}$ be the norm induced on the pair (u,v) over the Banach space $BV \times T_k \cap L^2$, then the dual norm is given by $||h||' = \sup\left\{\frac{1}{\mu}||h||_G, \frac{1}{\gamma}||h||_{T_k^*}\right\}$ for $h \in L^2(\Omega)$. Using this full dual norm, the following theorem provides conditions for non-trivial solutions.

Theorem 5.12. *Let $(u,v) \in BV \times T_k \cap L^2$ be the solution of equation (5.9), then the following hold:*

1. $||f||' \leq \lambda \Leftrightarrow (u,v) = (0,0)$
2. *If $||f||' > \lambda$, then the optimal pair (u,v) must satisfy*

$$\begin{aligned} ||u+v-f||' &= \frac{1}{\lambda} \\ \langle f-u-v, u+v\rangle &= \frac{1}{\lambda}\{\mu||u||_{TV} + \gamma||v||_{T_k}\} \end{aligned}$$

The proof can be generalized from [55, 47]. Lastly, when the auxiliary variable w is considered, Uzawa's method provides conditions for convergence of the saddle point problem, as previously done in [46].

5.4 Numerical Results for Weak Textures

Since the model is a saddle point problem, the numerical method is essentially an Uzawa-type method (see [25]). The algorithm first finds a w which maximizes the texture norm for the current approximation of the texture component. Then the calculated w is used to minimize the energy with respect to the other values. This is alternatively iterated until convergence. The algorithm is summarized below.

Algorithm

Initialize u^0, v^0, α^0, w^0

while Not Converged **do**

$$
\begin{aligned}
w^{n+1} &= \underset{w}{\operatorname{argmax}} \left\{ \frac{|\langle v^n, w \rangle|}{\int_\Omega |\nabla w| d\sigma(w)} \right\} \\
\left(u^{n+1}, v^{n+1}, \alpha^{n+1}\right) &= \underset{u,v,\alpha}{\operatorname{argmin}} \Big\{ \mu ||u||_{TV} + \gamma \frac{|\langle v, w^{n+1} \rangle|}{\int_\Omega |\nabla w^{n+1}| d\sigma(w^{n+1})} + \\
&\qquad \frac{1}{2} ||K(\alpha) * (u+v) - f||_2^2 + V(\alpha) \Big\}
\end{aligned}
$$

end while

In terms of each substep, the calculations are done using a semi-implicit finite difference gradient descent scheme.

5.4.1 Maximization of auxiliary variable

Given the approximated texture component v^n in order to compute w, we must maximize the following equation,

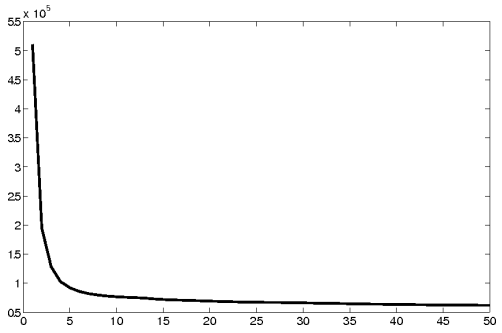


Figure 5.2: Energy versus number of iterations for Figure 5.10.

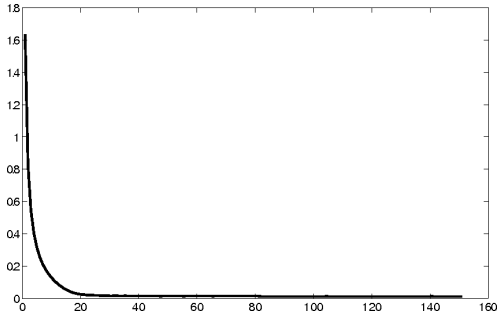


Figure 5.3: Inverse Value of the Maximization Step verse number of iteration for Figure 5.10.

$$\max_{w} \left\{ \frac{|\langle v^n, w \rangle|}{\int_\Omega |\nabla w| d\sigma(w)} \right\}.$$

To find the corresponding Euler-Lagrange equations, the dependence on the variable w in the measure σ is partially lagged. Assuming $w \in W^{1,\infty}$, the first variation (embedded in a gradient descent scheme) can be taken formally as:

$$\frac{\partial w}{\partial t} = \left(\frac{|\langle v^n, w \rangle|}{\int_\Omega |\nabla w| d\sigma} \right) \operatorname{div} \left(\sigma \frac{\nabla w}{|\nabla w|} \right) + \operatorname{sign} \left(\langle v^n, w \rangle \right) v^n$$

after some re-arrangement of terms and with a regularized absolute value function to avoid division by zero. Note that the expression in front of the curvature term is the energy that

is being maximized. For now we will drop the superscript on v^n, so as not to confuse it with the maximization iterations.

Next, fix $\Delta x = \Delta y = 1$ and let Δt be the time step. Let the n-th energy be discretely defined as: $E^n = \left(\frac{|\langle v, w^n \rangle|}{\int \sqrt{\left(w^n_{i+1,j} - w^n_{i,j}\right)^2 + \left(w^n_{i,j} - w^n_{i,j+1}\right)^2 + 1} d\sigma} \right)$. We define the following quantities related to the gradient of w^n:

$$\begin{aligned}
c^n_{1,i,j} &= \frac{1}{\sqrt{\left(w^n_{i+1,j} - w^n_{i,j}\right)^2 + .25\left(w^n_{i,j+1} - w^n_{i,j-1}\right)^2 + 1}} \\
c^n_{2,i,j} &= \frac{1}{\sqrt{\left(w^n_{i,j} - w^n_{i-1,j}\right)^2 + .25\left(w^n_{i-1,j+1} - w^n_{i-1,j-1}\right)^2 + 1}} \\
c^n_{3,i,j} &= \frac{1}{\sqrt{.25\left(w^n_{i+1,j} - w^n_{i-1,j}\right)^2 + \left(w^n_{i,j+1} - w^n_{i,j}\right)^2 + 1}} \\
c^n_{4,i,j} &= \frac{1}{\sqrt{.25\left(w^n_{i+1,j-1} - w^n_{i-1,j-1}\right)^2 + \left(w^n_{i,j} - w^n_{i,j-1}\right)^2 + 1}} \\
C^n_{i,j} &= \frac{1}{1 + E^n_{i,j}\Delta t \sigma_{i,j}\left(c^n_{1,i,j} + c^n_{2,i,j} + c^n_{3,i,j} + c^n_{4,i,j}\right)}
\end{aligned}$$

Next define the discrete inner product of the gradients of σ and w as:

$$d^n_{i,j} = \frac{\left(w^n_{i+1,j} - w^n_{i,j}\right)\left(\sigma_{i,j} - \sigma_{i-1,j}\right) + \left(w^n_{i,j+1} - w^n_{i,j}\right)\left(\sigma_{i,j} - \sigma_{i,j-1}\right)}{\sqrt{\left(w^n_{i+1,j} - w^n_{i,j}\right)^2 + \left(w^n_{i,j} - w^n_{i,j+1}\right)^2 + 1}}.$$

The forward gradient of w and backward gradient of σ is chosen for consistency with the energy. Altogether, the iteration step is defined below.

$$\begin{aligned}
w^{n+1}_{i,j} = C^n_{i,j}\Big\{ & w^n_{i,j} + \Delta t \text{sign}\left(\langle v, w^n \rangle\right) v + E^n_{i,j}\Delta t\Big(d^n_{i,j} + c^n_{1,i,j} w^n_{i+1,j} + \\
& c^n_{2,i,j} w^n_{i-1,j} + c^n_{3,i,j} w^n_{i,j+1} + c^n_{4,i,j} w^n_{i,j-1} \Big) \Big\}.
\end{aligned}$$

The semi-implicit discretization provides some stablity. This algorithm can be further stablized with the use of Sobolev gradient preconditioners, see [65, 72, 66].

5.4.2 Minimization of variables

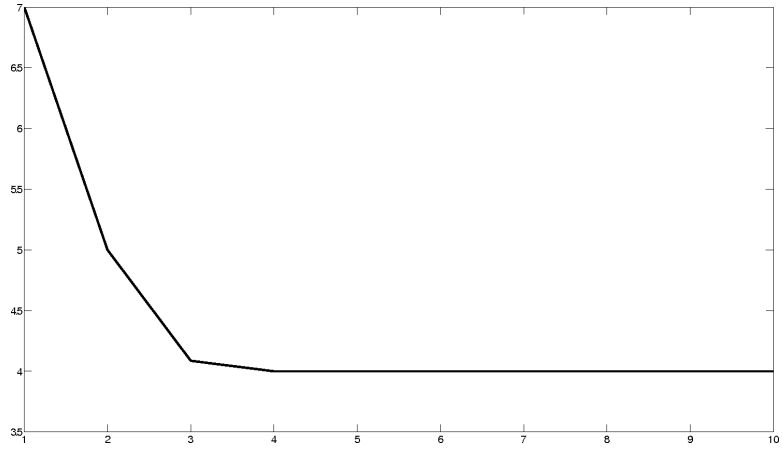


Figure 5.4: Kernel Parameter versus iteration for Figure 5.10.

For the minimization step in the algorithm, we assume that w is given. To compute the Euler-Lagrange equations in u and v, we assume that $(u, v) \in W^{1,1}(\Omega) \times L^2(\Omega)$, with $|u|_{TV} = \int_\Omega |\nabla u| dx$ the $W^{1,1}(\Omega)$ semi-norm. This is not a restriction, since functions in $BV(\Omega)$ can be sequentially approximated by functions in $W^{1,1}(\Omega)$ with respect to the strong topology $L^1(\Omega)$ (also u and v can only be calculated numerically in these stronger spaces). Now, taking the Euler-Lagrange equations for (u, v) and embedding them in a time-dependent scheme yields the following system.

$$\frac{\partial u}{\partial t} = \mu \text{div}\left(\frac{\nabla u}{|\nabla u|}\right) + K_\alpha^* \left(f - K_\alpha(u + v)\right) \tag{5.10}$$

$$\frac{dv}{dt} = -\frac{w \ \text{sign}\left(\langle v, w\rangle\right)}{\int_\Omega |\nabla w| d\sigma} + K_\alpha^* \left(f - K_\alpha * (u + v)\right). \tag{5.11}$$

In terms of the blur kernel parameters, if $\alpha \in \mathcal{K}$ (where $\mathcal{K}$ is a compact subset of $\mathbb{R}^d$, bounded away from zero) is a minimizer of the energy in equation (5.6), then α must satisfy the equation below:

$$0 \;=\; \int \left\{ (f - K_\alpha * (u+v)) \frac{\partial K_\alpha}{\partial \alpha} * (u+v) + p\beta \frac{\partial \left|\nabla K_\alpha\right|^p}{\partial \alpha} \right\} dx. \tag{5.12}$$

Define $F(\alpha)$ as the right hand side of equation (5.12). Then the problem reduces to finding the zeros of $F(\alpha)$. The roots of equation (5.12) are found by the bisection method.

To solve the system for (u, v), equation (6.10) is solved semi-implicitly, while equation (6.11) is explicit. Similar to the previous section, define the following quantities:

$$\begin{aligned}
k^n_{1,i,j} &= \frac{1}{\sqrt{\left(u^n_{i+1,j} - u^n_{i,j}\right)^2 + .25\left(u^n_{i,j+1} - u^n_{i,j-1}\right)^2 + \epsilon}} \\
k^n_{2,i,j} &= \frac{1}{\sqrt{\left(u^n_{i,j} - u^n_{i-1,j}\right)^2 + .25\left(u^n_{i-1,j+1} - u^n_{i-1,j-1}\right)^2 + \epsilon}} \\
k^n_{3,i,j} &= \frac{1}{\sqrt{.25\left(u^n_{i+1,j} - u^n_{i-1,j}\right)^2 + \left(u^n_{i,j+1} - u^n_{i,j}\right)^2 + \epsilon}} \\
k^n_{4,i,j} &= \frac{1}{\sqrt{.25\left(u^n_{i+1,j-1} - u^n_{i-1,j-1}\right)^2 + \left(u^n_{i,j} - u^n_{i,j-1}\right)^2 + \epsilon}} \\
A^n_{i,j} &= \frac{1}{1 + \mu\Delta t\left(k^n_{1,i,j} + k^n_{2,i,j} + k^n_{3,i,j} + k^n_{4,i,j}\right)}.
\end{aligned}$$

Then the semi-implicit iteration step for calculating the cartoon component is:

$$\begin{aligned}
u^{n+1}_{i,j} \;=\; A^n_{i,j}\Big\{ & u^n_{i,j} + \mu\Delta t\Big(k^n_{1,i,j}u^n_{i+1,j} + k^n_{2,i,j}u^n_{i-1,j} + k^n_{3,i,j}u^n_{i,j+1} + k^n_{4,i,j}u^n_{i,j-1}\Big) \\
& + \Delta t K^*_\alpha\left(f - K_\alpha(u^n + v)\right)\Big\}.
\end{aligned}$$

For the texture component, equation (6.11) is discretized using Euler's method, resulting in the following explicit iteration step,

$$v^{n+1} \;=\; v^n - \Delta t\left(\frac{w\ \mathrm{sign}\left(\langle v^n, w\rangle\right)}{\int_\Omega |\nabla w| d\sigma} - K^*_\alpha\left(f - K_\alpha * (u + v^n)\right)\right).$$

The time step Δt must be sufficiently small as to satisfy a CFL type condition on the system of equations. In practice an alternating minimization is used to solve the second substep in the overall scheme. The cartoon and texture are calculated first, then the parameters for the blur kernel. In all of the experiments in the next section, Δt is taken around 0.0025 to 0.025. Although the spaces T_k for various k's yield different texture components, for all the experiments here we fixed $k = 1$ for consistency.

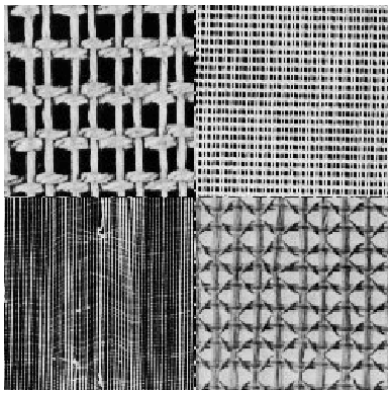
5.5 Experimental Results

In Figure 5.1, some examples of cartoon-texture decompositions are provided. The Grass and Tank images display very oscillatory behaviors, while the Brodatz image contains more patterned textures, all of which are captured well by the proposed texture norm.

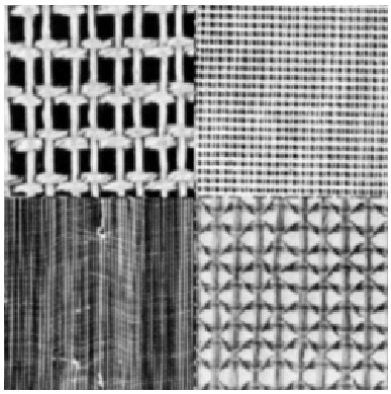
To verify the stability of our numerical scheme, Figure 5.2 shows that the total energy decreases over each iteration. Figure 5.3 shows that the inverse of the energy associated with the maximization substep is decreasing (which is expected since minimizers of the inverse energy are maximizers of the original energy). The extra regularity, in the form of the texture norm seems to also provide fast convergence of the parameters in the blur kernel. In Figure 5.4, notice that only 4 alternating minimization iterations are needed before the blur kernel converges (in fact, it converged to the exact parameter within algorithmic tolerance).

In the first set of examples, the blur kernel is known. In Figure 5.5, the Brodatz image is blurred with a Gaussian kernel with standard deviation equal to 0.75. The reconstructed image, the cartoon component, and the texture component are shown. Our reconstruction more than doubles the SNR, going from 12.33 to 35.15. Similarly, in Figure 5.6 the fingerprint image is blurred with a Gaussian kernel of standard deviation 0.74. The reconstructed image contains much of the lost texture, effectively sharpening the small scale detail.

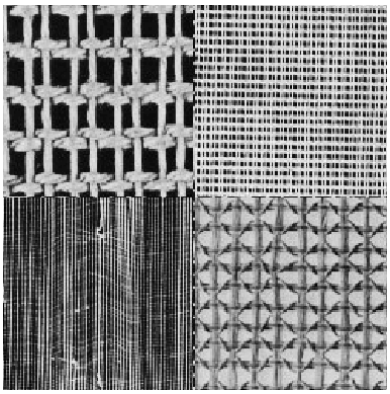
Unlike the previous examples, the semi-known case is more difficult, since small errors in the blur kernel propagate throughout the image. To make the problem more challenging, we have perturbed the blur kernel with random noise. Therefore, at best we will approximate



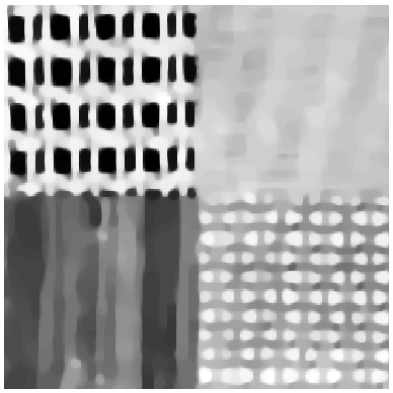
(a) Origina imagel



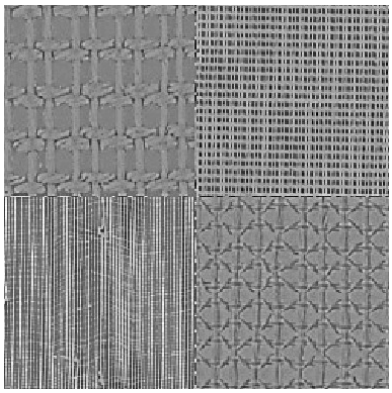
(b) Blurry image f, SNR=12.33



(c) Reconstructed image $u + v$, SNR=35.15



(d) Cartoon component u



(e) Texture component v

Figure 5.5: Deblurring results with known Gaussian Kernel with a standard deviation of 0.75 .

the blurring operation, since the model fits the true blur kernel to a class of parameterized kernels. This is preferred over fully unknown deblurring, since in practice, the class of kernels can be assumed from the type of degradation or from the acquisition process.

In Figure 5.7, the fingerprint image is blurred with a Gaussian kernel of standard deviation 0.9 and is perturbed by 5% relative noise. The reconstructed image recovers much of the texture and in terms of SNR, improves the quality of the image. In Figure 5.8, the Barbara image is blurred with an Out-of-Focus kernel of radius 2.5 and is perturbed by 15.2% relative noise. The blurred image loses many of the fine scale details of the scarf and the chair. Both of these patterns are recovered in the reconstructed image even with the highly noise kernel.

(a) Original image

(b) Blurry image f, SNR=13.32

(c) Reconstructed image $u + v$, SNR=32.77

(d) Cartoon component u

(e) Texture component v

Figure 5.6: Deblurring with Known Gaussian Kernel with a standard deviation of 0.74 .

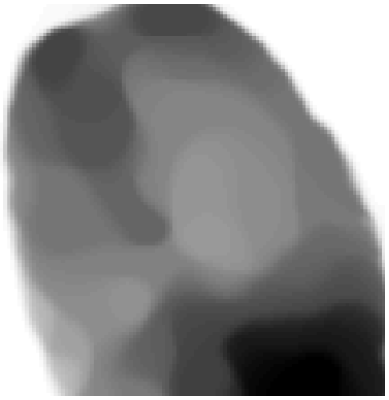
In Figure 5.9, the Brodatz image is blurred using the Out-of-Focus kernel with a radius of 1.5 and is perturbed by 5.0% relative noise. The finer textures (for example, in the bottom left quadrant of the image or along the lighter region in the top left quadrant) are removed. The reconstructed image has improved the SNR and restored the lost texture. Lastly, in Figure 5.10, the Chemical Plant image is blurred using the Out-of-Focus kernel with a radius of 4.0 and is perturbed by 8.0% relative noise. Similar to the previous examples, we see a gain in both edges and textures, with numerical support from the SNR.

Figure 5.11 provides comparisons with a standard deblurring regularizer. All images and SNRs for Figure 5.11 are found using the *known* blur. Even with the known kernels, the SNRs for images recovered by the ROF model are lower then those of our model with unknown

(a) Original image

(b) Blurry image f, SNR=12.02

(c) Reconstructed image $u + v$, SNR=17.67



(d) Cartoon component u

(e) Texture component v

Figure 5.7: Semi-Known Deblurring with Gaussian Kernel with a standard deviation of 0.9 and relative noise of 5%.

(a) Original image

(b) Blurry image f, SNR=15.77

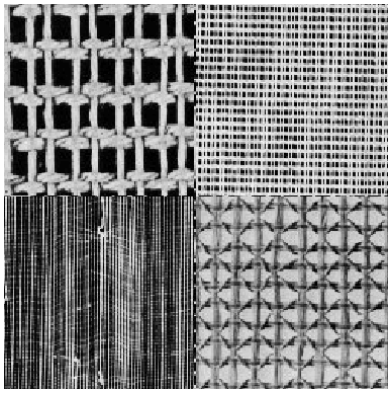
(c) Reconstructed image $u + v$, SNR=19.59

(d) Cartoon component u

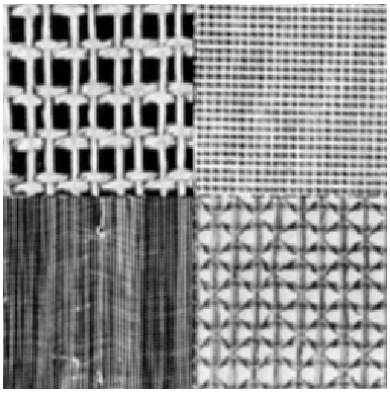
(e) Texture component v

Figure 5.8: Semi-Known Deblurring with Out-of-Focus Kernel with radius 2.5 and relative noise of 15.2%.

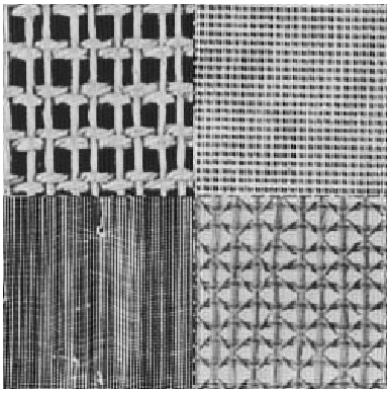
kernel parameters. This supports the use of a weak texture norm to further regularize image deblurring models, especially for unknown blur.



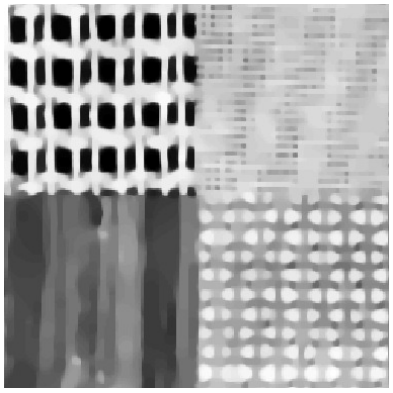
(a) Original image



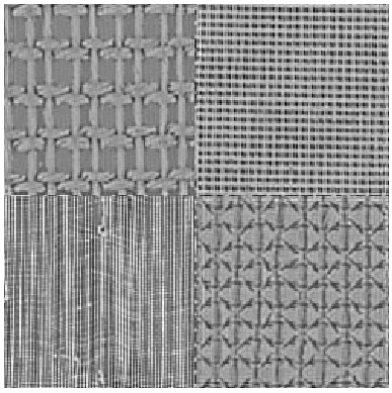
(b) Blurry image f, SNR=16.11



(c) Reconstructed image $u + v$, SNR=21.70



(d) Cartoon component u



(e) Texture component v

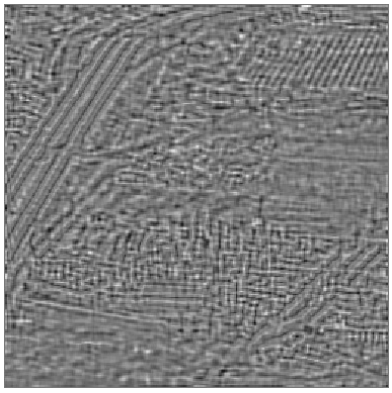
Figure 5.9: Semi-Known Deblurring with Out-of-Focus Kernel with radius 1.5 and relative noise of 5%.

(a) Original image

(b) Blurry image f, SNR=16.19

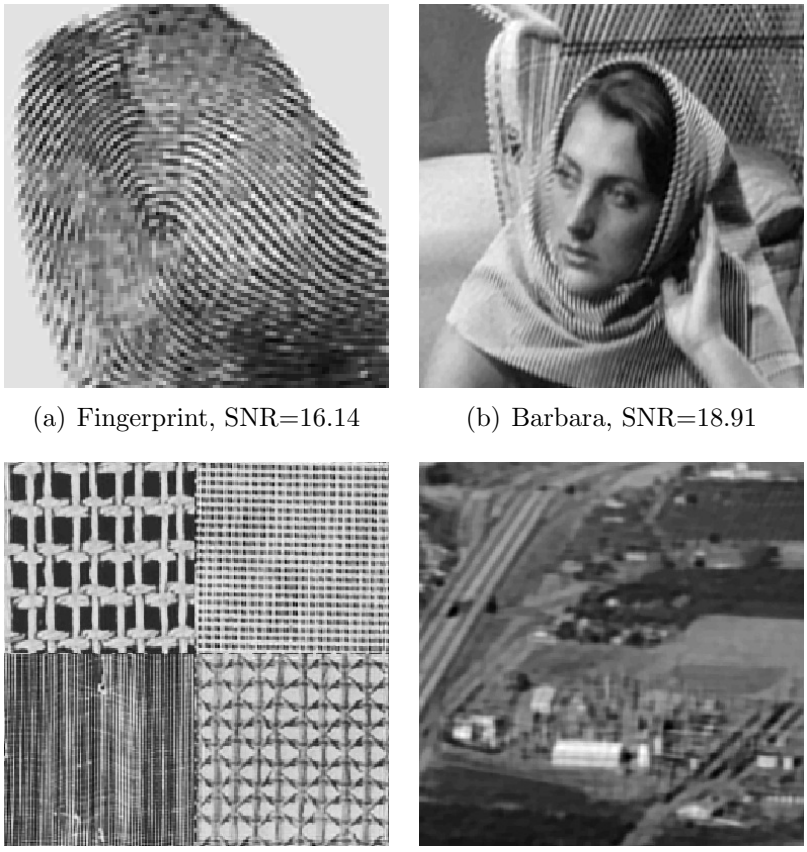
(c) Reconstructed image $u + v$, SNR=21.63

(d) Cartoon component u



(e) Texture component v

Figure 5.10: Semi-Known Deblurring of the Chemical Plant image with Out-of-Focus Kernel with radius 4 and relative noise of 8%.



(a) Fingerprint, SNR=16.14 (b) Barbara, SNR=18.91

(c) Brodatz, SNR=15.32 (d) Chemical Plant, SNR=20.16

Figure 5.11: Comparisons with TV-L^2 (the ROF model). In this case, the kernel is known.

CHAPTER 6

Low Patch-Rank Texture

6.1 Introduction to Patterned Texture

As seen in Chapter 5, the weak texture norms provide extra regularization on the recovered image, better recovering many of the features from the original image. However, they are unable to separate texture from noise. The reason for this is in the nature of the model: the underlying assumption is that texture is highly oscillatory. This property does not distinguish between texture and noise.

In this chapter, texture is modeled using the assumption that it is well patterned and appears throughout the image. The model presented here is related to the works of non-local imaging models and principal component analysis.

The non-local methods were first proposed by Buades, Coll, and Morel [10, 11] as a non-local filter and were later formulated in a variational framework by Gilboa and Osher [32, 33]. The general framework involves replacing local derivatives by their non-local counterpart:

$$(\nabla_w u)(x,y) := (u(y) - u(x))\sqrt{w(x,y)}$$

for all $x, y \in \Omega$ where $w(x,y) = e^{-d(u(x),u(y))}$, $d(u(x),u(y)) = \int_\Omega G(t)|u(x+t) - u(y+t)|^2 dt$, and $G(t)$ is a Gaussian with appropriate parameters. In the discrete version, each pixel is associated with a patch, and local differences are replaced by differences between pixels that have similar patches. Using these patch-based differences in the energy encourages repetitive behavior in the reconstructed image, thereby recovering the cartoon and texture while removing random noise. In particular, the non-local extension of the classical ROF

model:

$$\inf_u \mathcal{E}(u) = ||\nabla_w u||_{L^1} + \lambda ||u - f||_{L^2}^2 \tag{6.1}$$

has been shown to recover the texture well. Although these models are useful, due to the global nature of images, they tend to be very slow in practice because of the frequent re-computation of the weight function $w(x, y)$, especially for tasks such as deblurring and sparse reconstruction.

Robust Principal Component Analysis (PCA) was proposed by Candes, Li, Ma, and Wright [13] in order to recover the sparse and low rank parts of a given matrix f. This is also formulated as a decomposition problem: decompose $f = u + v$, where u is sparse and v is low-rank, by minimizing:

$$\inf_{u,v} \mathcal{E}_{PCA}(u, v) = ||u||_{L^1} + \lambda ||v||_*$$
$$\text{s.t.} \quad f = u + v$$

where $|| \cdot ||_*$ is the nuclear (or trace) norm, which is the sum of the singular values. Because this method almost exactly recovers the original sparse and low rank components, it is becoming increasingly popular in practice. In the past few years, Robust PCA has been applied to video surveillance [13], face recognition [13], and video denoising [38]. This variational model was later extended by Gao, Cai, Shen, and Zhao, who replaced the L^1 norm with a tight frame regularization [29] and a total variation regularization [30]. The PCA models also benefit from the split Bregman method, making their implementation both efficient and fast.

Both non-local methods and robust PCA exploit the patterned nature of data. For completeness, we mention that sparse decomposition methods exist in the class of dictionary learning techniques. The main idea of this technique is to find a sparse representation for an image with respect to a (learned) redundant dictionary. Elad and Aharon [26] and Zhou,

Chen, Paisley, Ren, Sapiro, and Carin [78] both applied (different) Bayesian approaches in order to construct dictionaries comprised of subparts of the given image. In [53], Mairal, Elad, and Sapiro extended the idea of sparse representation over a dictionary to color image restoration- in particular, denoising, inpainting and demosaicing. The idea of sparse coding was later combined with the nonlocal methods in the work of Mairal, Bach, Ponce, Sapiro, and Zisserman [51]. Similarly, in [24], the idea of structural clustering and dictionary learning was proposed and used for highly textured image restoration. Works such as [52] have also provided more efficient and less costly ways to implement these dictionary learning methods.

Both PCA methods and cartoon-texture models decompose a given f into two main components, where one is "sparse" in some sense and the other is "patterned". By combining the patch-based methodology of non-local methods with the pattern nature of PCA methods, the low patch-rank model shows a new way of interpreting texture.

Note that a version of this chapter appears in [68].

6.2 Description of the Model

The proposed model will be of a similar form to the classical cartoon-texture decomposition models. Given an f, we decompose $f = A(u+v) + \rho$ by minimizing the following energy:

$$\inf_{u,v} \mathcal{E}(u,v) = \mu||u||_{TV} + \gamma||v||_{Texture} + \frac{\lambda}{2}||A(u+v) - f||_2^2$$

where A is some degradation operator, for example a reduction operator for missing data. In practice, we solve this minimization using a double Bregman splitting, which can enforce $f = A(u+v)$ exactly or can be relaxed in the presence of noise. The cartoon norm we use is the (discrete) total variation defined as:

$$||u||_{TV} := ||Du||_1 = \sum_{i,j} |Du_{i,j}|$$

where $D = [D_x, D_y]$ is a differencing operator and $|Du_{i,j}| = \sqrt{(D_x u_{i,j})^2 + (D_y u_{i,j})^2}$. For the rest of this work, we take D to be the forward differences. Before going into more detail, we first define our texture norm.

6.2.1 The Texture Norm

From an intuitive perspective, the texture component is a global and well-patterned structure within a given image. The distinct patterns that make up the entire texture are called *base textures*. We expect the number of base textures to be low, since an image may exhibit only a few individual patterns. In particular, each patch (i.e. sub-block) should be comprised of a combination of these base textures. Therefore, the overall collection of patches can be spanned by a small set of base patches. If the patches are written as vectors, then the collection of *patch-vectors* are (highly) linearly dependent, and thus have low rank. This is the key to our definition of texture and its norm. Using this idea we have the following definition for the collection of patches and the texture norm:

Definition 6.1. The *patch map*, $\mathcal{P} : \mathbb{M}^{n,m} \to \mathbb{M}^{r^2, \frac{nm}{r^2}}$ is defined by the following: For $v \in \mathbb{M}^{n,m}$, partition up v into r-by-r (non-overlapping) submatrices, labeled $\{B_i\}_{i=1}^{\frac{nm}{r^2}}$. Next, transform each of the B_i's into a column vector of length r^2, called w_i. Then augment the vectors together to form the new matrix:

$$\mathcal{P}v := \left[w_1 .... w_{\frac{nm}{r^2}} \right]$$

The specific ordering which maps the patch to a patch-vector is not important, as long as it is consistent (see lemma 6.12). For the results here, a raster scan (row by row) was used to arrange the patches and a second raster scan was used to re-write each patch as a

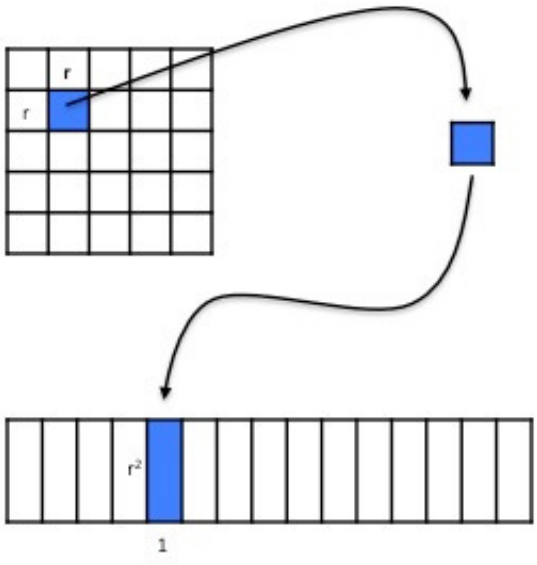


Figure 6.1: Patch Map

vector.

Note that the patch map does not increase the number of terms since the patches are non-overlapping. Also, the collection of texture patches is found by applying the patch map to the texture. Figure 6.1 depicts the patch map applied to a matrix. Using definition 6.1 and the behavior we expect on the texture, the natural norm on $\mathcal{P}v$ would be:

$$||v||_T \stackrel{?}{:=} rank\left(\mathcal{P}v\right)$$

However, this energy is non-convex and difficult to use in practice. Furthermore, *rank* is not a norm in the mathematical sense. Using the ideas from Robust PCA [13], we can replace the rank by the nuclear/trace norm, which is the convex envelope to the *rank* and in fact a norm. Thus, the suggested texture norm can be relaxed to:

Definition 6.2. A function $v \in T$ if it is mean zero and if the following quantity is finite:

$$||v||_T := ||\mathcal{P}v||_*$$

where $||\cdot||_*$ is the nuclear/trace norm, i.e. the sum of the singular values.

For a discrete function v, the norm $||v||_T$ is always finite, but we use this definition for a more general v which will be addressed in future work.

Unlike the non-local methods, this norm does not explicitly calculate the weights between patches, but rather compares the patches implicitly. In this way, our non-local measure is computationally more efficient, while still being easy to compute and simple to minimize. To better understand definition 6.2, here is an example of a texture and its norm.

Example 6.3. *Let the texture be the zero mean vertical stripe pattern defined as:*

$$v_{i,j} = \begin{cases} 1 & : \ j \ is \ even \\ -1 & : otherwise \end{cases}$$

of size N by N. Then after applying the patch map with 2 by 2 patches:

$$\mathcal{P}_{2\times 2}v = \begin{bmatrix} -1 & -1 & \\ 1 & 1 & \\ -1 & -1 & .... \\ 1 & 1 & \end{bmatrix}$$

Similarly, after applying the patch map with 3 by 3 patches:

$$\mathcal{P}_{3\times 3}v = \begin{bmatrix} -1 & 1 & \\ 1 & -1 & \\ -1 & 1 & \\ -1 & 1 & \\ 1 & -1 & \\ -1 & 1 & \\ -1 & 1 & .... \\ 1 & -1 & \\ -1 & 1 & \end{bmatrix}$$

In both cases the resulting matrices are of rank 1. It is easy to show that there is only one non-zero singular value equal to rN. Consequently, $||\mathcal{P}_{r\times r}v||_* = rN$. Of course, v has the same value in all L^p norms: $||v||_{L^p} = N^2$ and its total variation is $||v||_{TV} = 2N^2$. Since N is much larger than r, the norms are well ordered: $||\mathcal{P}_{r\times r}v||_* \ll ||v||_{TV}$ and $||\mathcal{P}_{r\times r}v||_* \ll ||v||_{L^p}$. This type of behavior is desired in cartoon-texture models.

6.2.2 Proposed Model

With definition 6.2, the model is as follows: Given an f, we decompose $f = A(u+v)+\rho$ by minimizing:

$$\mathcal{E}(u,v) = \mu||Du||_1 + \gamma||\mathcal{P}v||_* + \frac{\lambda}{2}||A(u+v)-f||_2^2 \tag{6.2}$$

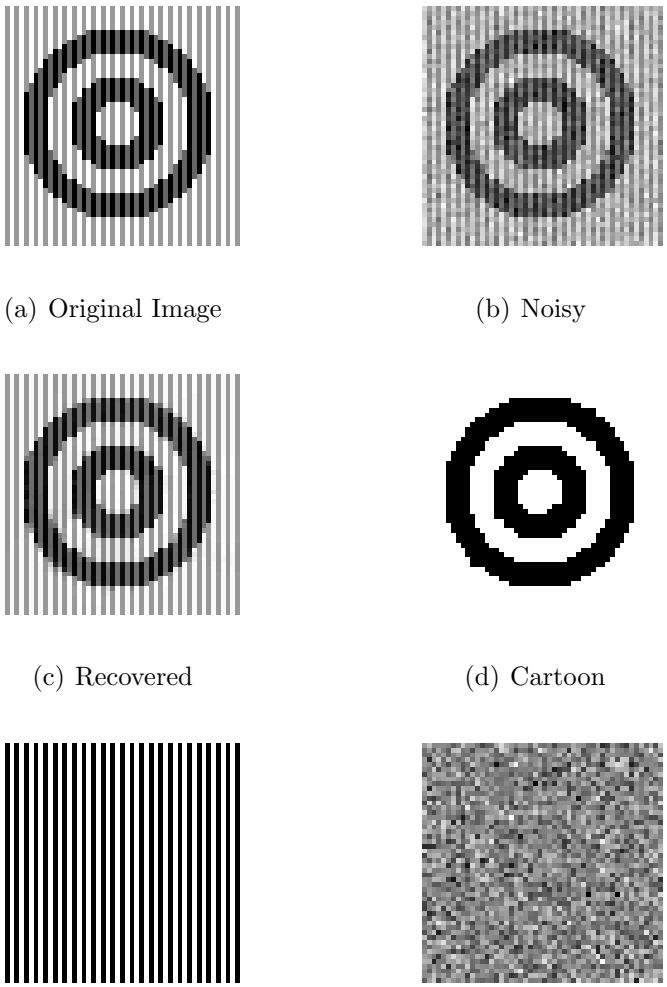
for all $(u,v) \in \mathcal{A}$, where the admissible set is $\mathcal{A} := \big\{(u,v)\,\big|\,\sum u = \sum f,\ \sum v = 0\big\}$. In practice, the minimizers remain in this admissible set without any formal constraints. The operator A is assumed to be linear. For denoising, A is the identity; for deblurring it is a

convolution with a blur kernel; and for inpainting or sparse reconstruction, it is a reduction operator.

Recall that minimizing this energy ensures that the cartoon will be in discrete BV, thereby being piecewise smooth with sharp edges. The texture norm ensures a low patch-rank collection of textures and thus a small amount of repetitive textures. Lastly, the residual (or noise) term $\rho := A(u+v) - f$ remains in L^2 since it is assumed to have no particular structure. In our examples, the noise is Gaussian; however, the residual term $\frac{\lambda}{2}||\rho||_2^2$ can be replaced with other norms depending on the type of corruption. For example, impulse noise and blind inpainting (of small regions) is better captured by using the L^1 norm, i.e. $\lambda||\rho||_1$.

In Figure 6.2, we provide a simple decomposition example. The cartoon consists of concentric annuli, the texture is a repetitive stripe pattern, and the noise is random. In this example, the rank of the texture is 1, while the cartoon and noise components have full rank. It is worth noting that this decomposition is not equivalent to thresholding the input images' singular values. Figure 6.3 plots the singular values in descending order of each of the images from Figure 6.2 (after applying the patch map). The texture has only one singular value since it is comprised of only one pattern. The cartoon has many patch patterns based on the various alignments of edges between the homogeneous regions, all with varying degrees of importance, with the most important being the constant patch. The noise has almost no coherent structures, which can be seen in its singular values. From Figure 6.3, it is clear that the texture is the only component to have many zero singular values. Even if the "smaller" singular values were thresholded, the texture component would still have the smallest singular value support.

Remark 6.4. *Like the cartoon-texture symmetry from the $BV - G$ model, our cartoon and texture norms are related in a curious way. For the cartoon u, the operator D is applied and the result is measured in L^1. For the texture v, the patch-texture is decomposed by the singular value decomposition; consequently, there exists two unitary matrices U and V such that: $\mathcal{P}v = U\Sigma V^*$, where Σ is a diagonal matrix that contains the singular values of $\mathcal{P}v$. This is equivalent to the L^1 norm on Σ, i.e. $||U^*\mathcal{P}vV||_1$. While the operator D looks at local*



(a) Original Image (b) Noisy

(c) Recovered (d) Cartoon

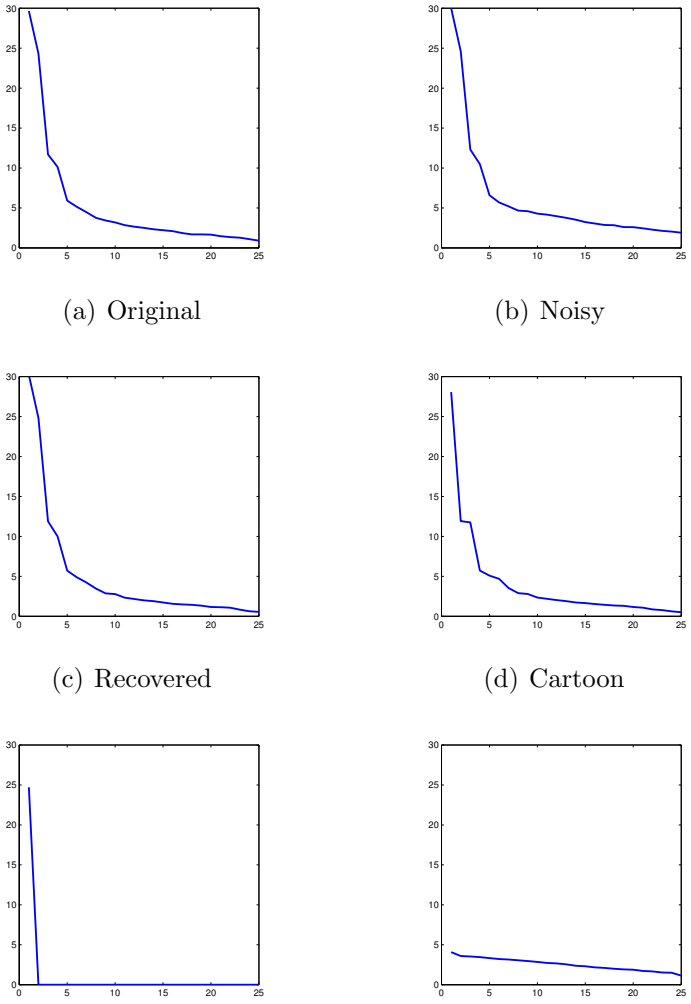
(e) Texture (rescaled) (f) Noise (residual)

Figure 6.2: Decomposition of a Synthetic Example Image

comparisons in pixel values, the operation $v \longmapsto U^*\mathcal{P}vV$ *compares patches non-locally. The* L^1 *norm is used to measure sparsity in jumps for the case of the cartoon and sparsity in pattern in the case of the texture.*

Remark 6.5. *There are many ways to view our texture norm.*

1. *From the perspective of non-local methods, our norm measures the similarity between patches based on linear dependence rather than element-wise differences. In these methods, the typical measure of similarity is the exponential of the* L^2 *distance between*



(a) Original (b) Noisy

(c) Recovered (d) Cartoon

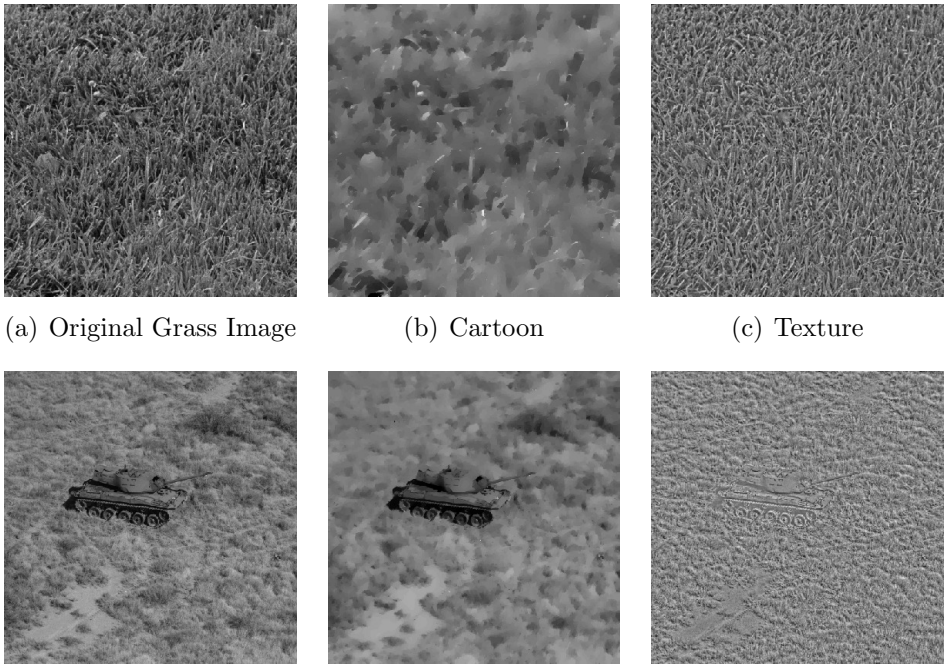
(e) Texture (f) Noise

Figure 6.3: Singular Values in descending order

patches. This can cause problems when two patches agree exactly in texture but have different means. Specifically, if the difference in means is large then the patches are considered to be dissimilar, which can lead to improper comparisons between patches. In practice, this leads to problems in contrast (see Figure 6.19).

2. *From a dictionary approach, our texture norm creates a basis* B_i *for* $i = 1, ..., rank\,(\mathcal{P}v)$ *comprised of the singular vectors of* $\mathcal{P}v$*. All patches are thus comprised of linear combinations from this basis, creating a implicit dictionary based on the image itself.*

3. *From a cartoon-texture point of view, our texture norm induces a space* $(T, ||\mathcal{P} \cdot ||_*)$ *which is used in the energy. As with the functional spaces, this texture norm decreases as the pattern becomes more repetitive.*



(a) Original Grass Image (b) Cartoon (c) Texture

(d) Original Tank Image (e) Cartoon (f) Texture

Figure 6.4: Examples of Image Decomposition. The patch-rank of the texture components for the Tank and Grass images are 51.0% and 19.1% respectively. Even seemingly random textures have a small patch-rank. Notice that spiking (or point structures) in the original images (a) and (d) are found in the cartoon components (b) and (e) and not the texture components (c) and (f).

6.3 Theoretical and Analytical Remarks

In this section we will examine the behavior of this cartoon-texture separation model, with the constraint that the input data f is of mean zero. This is not a restriction in practice since it is simply a rescaling of the data.

6.3.1 Characterization of Minimizers by Duality

In the continuous framework, the pre-dual of BV is G, the space of generalized functions. As in definition 4.1, the space G is equivalent to the space of functions which are the divergence of L^∞ vector fields. The space and norm first appeared in the characterization of minimizers for the ROF model and then in cartoon-texture models [55]. For our discrete model, we would like an analogous space which has a similar duality to our discrete BV space. In [3], a discrete G is given as follows:

Definition 6.6. The dual of (discrete) BV is the (discrete) space G which has the following norm:

$$||v||_G := \inf_{v=\text{div}\vec{g}} ||\sqrt{(g^1_{i,j})^2 + (g^2_{i,j})^2}||_{L^\infty}$$

where $\vec{g}_{i,j} = (g^1_{i,j}, g^2_{i,j})$.

We will consider the discrete divergence operator to be $\text{div} = \left(D_x^-, D_y^-\right)\cdot$ where D_i^- are the backward differences. It can be shown that this definition of divergence yields $\text{div} = -D^*$, where D^* is the adjoint of the operator from the definition of discrete TV. This is necessary to insure a discrete duality principle, or in other words, for any u and v $|\langle u, v\rangle| \leq ||Du||_1||v||_G$. A similar duality is needed with respect to the texture norm, which follows from the following properties of the patch map:

Lemma 6.7. *If $\mathcal{P}$ is a patch map then*

1. *It is a bijective linear operator.*
2. *$\mathcal{P}$ is an isometry with respect to all element-wise norms.*
3. *$||\mathcal{P}v||_* \leq \sqrt{n}||v||_2$, where n^2 is the size of the matrix v.*
4. *The dual norm of $||\mathcal{P}\cdot||_*$ is $||\mathcal{P}\cdot||_s$, where $||\cdot||_s$ is the spectral norm.*

There are sharper bounds for condition 3 above; however, they depend on the patch size and the texture component. The proofs for each of the above are easy to show. From these properties, the texture duality can be shown to be: $|\langle u, v\rangle| \leq ||\mathcal{P}u||_s ||\mathcal{P}v||_*$. Based on these dualities, there are four characterization theorems listed below. The characterization is centered around the pure decomposition version of the model (when $A = I$):

$$\mathcal{E}(u, v) = \mu||Du||_1 + \gamma||\mathcal{P}v||_* + \frac{\lambda}{2}||u + v - f||_2^2 \tag{6.3}$$

The theorems give a relationship between the parameters (μ, γ, λ) and the quantities $||f||_G$ and $||\mathcal{P}f||_s = \sigma_{max}(\mathcal{P}f)$ (which is the maximum singular value of the patch-form of the image). The first two theorems address the two trivial decompositions where no cartoon appears and the optimal solutions vary between the texture and residual terms. The proofs of these theorems can be found in Section 7.2.

Theorem 6.8. *If $0 < \gamma < \frac{2\mu}{n}$, then the minimizer (u, v, ρ) must have $u = 0$. If in addition $||\mathcal{P}f||_s \leq \frac{\gamma}{\lambda}$, then the minimizer (u, v, ρ) yields $u = \rho = 0$ and $v = f$.*

Theorem 6.9. *If $||f||_G \leq \frac{\mu}{\lambda}$ and $||\mathcal{P}f||_s \leq \frac{\gamma}{\lambda}$ then the minimizer (u, v, ρ) must yield $u = v = 0$ and $\rho = f$.*

The last two theorems are the more interesting cases. By choosing the parameters accordingly, the resulting optimal solutions will provide non-trivial decompositions.

Theorem 6.10. *If $||f||_G > \frac{\mu}{\lambda}$ and $||\mathcal{P}f||_s > \frac{\gamma}{\lambda}$, then the minimizer yields $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, $\langle \rho, u\rangle = \frac{\mu}{\lambda}||Du||_1$, and $\langle \rho, v\rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$.*

Theorem 6.11. *If $||f||_G \leq \frac{\mu}{\lambda}$ and $||\mathcal{P}f||_s > \frac{\gamma}{\lambda}$, then three optimal cases hold:*

(1) $u = 0$, $||\rho||_G < \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s \leq \frac{\gamma}{\lambda}$, and $\langle \rho, v\rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$

(2) $v = 0$, $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s < \frac{\gamma}{\lambda}$, and $\langle \rho, u\rangle = \frac{\mu}{\lambda}||Du||_1$

(3) $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, $\langle \rho, v\rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$, and $\langle \rho, u\rangle = \frac{\mu}{\lambda}||Du||_1$

These theorems provide some insight into choosing coefficients to obtain particular behaviors in the minimizers.

6.3.2 Characterization of Texture based on $\mathcal{P}$

We can further characterize solutions based on our patch map. The operator $\mathcal{P}$ is implicitly dependent on both the way in which the elements are re-assigned and the patch size. First, to address the "reshaping" we have the following theorem:

Lemma 6.12. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two patch maps which are identical except for the order in which they map the sub-blocks of v into the columns of $\mathcal{P}_i v$ and the elements within the sub-block of v into the rows of $\mathcal{P}_i v$. Then we have that for all v : $||\mathcal{P}_1 v||_* = ||\mathcal{P}_2 v||_*$.*

By this lemma we see that our method is independent of the ordering used to reshape the texture matrix into the patch-form.

Next, let us address the dependence of the texture component on the patch size r (for now assume the patches are square). Like the tuning parameters in the energy, this parameter also determines certain characteristics of the minimizers. For example, the patch size has a subtle relationship with the various "texture frequencies." Consider the following example:

Example 6.13. *Let $f(x, y)$ be a mean zero function that has an oscillation of period T only in the x-direction (and constant along the y-direction). Let r be the length of the square patches.*

1. *If $T < r$, then the patch-rank must be larger or equal to 1. If there are small scale symmetries during one-period, then the patch-rank can be equal to 1. For example, take $r = \frac{T}{2}$. If $f(x, y) = -f(x + \frac{T}{2}, y)$, then the patch on $[1, \frac{T}{2}] \times [1, \frac{T}{2}]$ is equal to the negative of the patch on $[\frac{T}{2} + 1, T] \times [\frac{T}{2} + 1, T]$. It follows that there is only one underlying base patch, so the rank is 1.*

2. *If $T > r$, then the patch-rank must be larger or equal to 1. This can be shown by using the previous argument with $r = \frac{3T}{2}$.*

3. *If $T = r$, then the patch-rank will be 1.*

From this example, it is clear that to minimize the amount of patches needed to describe a given texture, the parameter r should be as close to the the pattern period as possible.

Lastly, the following lemma address the partitioning of the texture component.

Lemma 6.14. *Let $\mathcal{P}$ be a patch map and let $\mathcal{S}$ be a index shifting operator (with periodic boundary conditions), then for any matrix v, $||\mathcal{P}\mathcal{S}v||_* = ||\mathcal{P}v||_*$ holds.*

The lemma shows that the patch map is invariant under uniform translations of the indices, therefore our proposed method is invariant of the partitioning grid. This is easy to show, since shifting the grid is equivalent to reordering the indices but does not change the relationship between indices, as in Lemma 6.12.

6.4 Numerical Method

The cartoon and texture are both defined with respect to fine (lower-dimensional) structures: co-dimensional 1 edges in the cartoon and co-dimensional 1 and 2 patterns in the texture. They are also both measured by L^1 type norms, which are efficiently solved by splitting methods — in our case the split Bregman Method [37]. The added advantage in these splitting methods is that lower-dimensional structures get enhanced. This is normally referred to as *contrast enhancement*, *i.e.* the sharpening of edges. An analogous effect seems to sharpen the "texture contrast," specifically linear features and isolated point structures. It is interesting to note that in practice, the Bregman iteration removes any block effects from the texture component, which occurs from the non-overlapping patch structure.

Below is an outline of the split Bregman technique applied to our model. We use a double Bregman, since both terms u and v are split. First introduce the auxiliary variables $d_1 = Du$ and $d_2 = \mathcal{P}v$:

$$\min_{d_1=Du, d_2=\mathcal{P}v} \mu||d_1||_1 + \gamma||d_2||_* + \frac{\lambda}{2}||A(u+v) - f||_2^2$$

Next add back the constraints $d_1 = Du$ and $d_2 = \mathcal{P}v$, enforcing them with the Bregman

variables b_1, b_2:

$$\min_{d_1,d_2,u,v} \mu||d_1||_1 + \gamma||d_2||_* + \frac{\lambda}{2}||A(u+v) - f||_2^2 + \frac{\lambda_1}{2}||d_1 - Du + b_1||_2^2 + \frac{\lambda_2}{2}||d_2 - \mathcal{P}v + b_2||_2^2$$

This splitting decouples the original equation into the following system:

$$\begin{aligned}
(u^{n+1}, v^{n+1}) &= \underset{u^n,v^n}{\operatorname{argmin}} \ \frac{\lambda}{2}||A(u^n + v^n) - f + f^n||_2^2 + \frac{\lambda_1}{2}||d_1^n - Du^n - b_1^n||_2^2 \qquad (6.4)\\
&\quad + \frac{\lambda_2}{2}||d_2^n - \mathcal{P}v^n - b_2^n||_2^2\\
d_1^{n+1} &= \underset{d_1^n}{\operatorname{argmin}} \ \mu||d_1^n||_1 + \frac{\lambda_1}{2}||d_1^n - Du^{n+1} - b_1^n||_2^2 \qquad (6.5)\\
d_2^{n+1} &= \underset{d_2^n}{\operatorname{argmin}} \ \gamma||d_2^n||_* + \frac{\lambda_2}{2}||d_2^n - \mathcal{P}v^{n+1} - b_2^n||_2^2 \qquad (6.6)\\
b_1^{n+1} &= b_1^n + Du^{n+1} - d_1^{n+1} \qquad (6.7)\\
b_2^{n+1} &= b_2^n + \mathcal{P}v^{n+1} - d_2^{n+1} \qquad (6.8)
\end{aligned}$$

In equation (6.4), we have also included the Bregman variable f^n, which is used when one wants to enforce the constraint $A(u+v) = f$. Each of the subproblems above can be easily solved, as follows. For the first subproblem, from Lemma 6.12 and the tensor discussion in the appendix section 7.1, we can re-write equation (6.4) as:

$$\begin{aligned}
(u^{n+1}, v^{n+1}) &= \underset{u^n,v^n}{\operatorname{argmin}} \ \frac{\lambda}{2}||A(u^n + v^n) - f + f^n||_2^2 + \frac{\lambda_1}{2}||d_1^n - Du^n - b_1^n||_2^2\\
&\quad + \frac{\lambda_2}{2}||\mathcal{P}^{-1}d_2^n - v^n - \mathcal{P}^{-1}b_2^n||_2^2
\end{aligned}$$

Since this problem is differentiable, taking the first variation yields the following linear system:

$$\begin{cases}
\left(\lambda A^*A - \lambda_1\Delta\right)u^{n+1} + \lambda A^*Av^{n+1} = F_1\\
\lambda A^*Au^{n+1} + \left(\lambda A^*A + \lambda_2\right)v^{n+1} = F_2
\end{cases} \qquad (6.9)$$

where $F_1 = \lambda A^* (f - f^n) - \lambda_1 D^* (d_1^n - b_1^n)$ and $F_2 = \lambda A^* (f - f^n) + \lambda_2 \mathcal{P}^{-1} (d_2^n - b_2^n)$. This equation can be solved completely (in the Fourier domain) or approximated with a few iterations of a Gauss-Seidel sweep. Next, equation (6.5) can be written out explicitly as a simple shrink:

$$d_1^{n+1} = shrink\left(Du^{n+1} + b_1^n, \frac{\mu}{\lambda_1}\right)$$

where the *shrink* function above is defined pointwise for 2-d vectors x as $shrink(x, \tau) := max\left(|x|_2 - \tau, 0\right) \frac{x}{|x|_2}$, and where $|\cdot|_2$ is the vector 2-norm and $\tau \in \mathbb{R}$. Lastly, equation (6.6) can be written explicitly as:

$$d_2^{n+1} = SVT\left(\mathcal{P}v^{n+1} + b_2^n, \frac{\gamma}{\lambda_2}\right)$$

where SVT is singular value thresholding, which is defined as follows: for a matrix M whose singular value decomposition (SVD) is given by $M = U\Sigma V^*$, the singular value thresholding function is defined as $SVT(M, \tau) := U\ max(\Sigma - \tau I, 0)\ V^*$, where the max is taken element-wise. There are methods to compute the SVT without using the SVD (see [12]), which can speed up the computations for large matrices. Using these formulas, the algorithm is presented in the next section.

6.4.1 The Algorithm

The splitting from the previous section reduces the problem from a difficult non-linear one to a sequence of simple linear (or explicit) subproblems. The algorithm involves two main loops: the inner loop, which solves each minimization, and the outer loop, which adds back the error and re-solves the minimization. The outer loop's termination is dependent on the problem we are solving, specifically on the amount of noise. Given a tolerance tol, for pure decomposition with no noise or for inpainting with no noise, the outer loop is iterated until $||f - A(u^{n+1} + v^{n+1})||_2 \leq tol$. In the presence of noise with standard deviation σ_{noise}, the stopping criteria becomes $||f - A(u^{n+1} - v^{n+1})||_2 \approx \sigma_{noise}$. The number of outer loops also determine the amount of texture-noise separation that occurs.

Algorithm

Initialize $u^0 = f, v^0 = 0$ $d_{1,x}^0 = d_{1,y}^0 = b_{1,x}^0 = b_{1,y}^0 = 0_{n,n}$, and $d_2^0 = b_2 = 0_{r^2, \frac{n^2}{r^2}}$

while Outer Iteration **do**

while Inner Iteration **do**

$$\begin{aligned}
(u^{n+1}, v^{n+1}) &= GS^n \\
d_1^{n+1} &= shrink\left(Du^{n+1} + b_1^n, \frac{\mu}{\lambda_1}\right) \\
d_2^{n+1} &= SVT\left(\mathcal{P}v^{n+1} + b_2^n, \frac{\gamma}{\lambda_2}\right) \\
b_1^{n+1} &= b_1^n + Du^{n+1} - d_1^{n+1} \\
b_2^{n+1} &= b_2^n + \mathcal{P}v^{n+1} - d_2^{n+1}
\end{aligned}$$

end while

$f^{n+1} = f^n + f - (u^{n+1} + v^{n+1})$

end while

where GS^n is the application of Gauss-Seidel sweeping to equation (6.9). Typically, only a few sweeps are necessary, since only partial convergence is needed.

6.5 Results

In the previous section, methods for choosing appropriate parameters were given in the examples and theorems. Before discussing the numerical results, a summary with one more parameter bound will be provided. Informally, an upper bound on the number of expected textures can be predicted, thereby providing a bound on $rank(\mathcal{P}v)$. As previously mentioned, the texture is comprised of highly oscillatory functions. If we model them as sums of sines and cosines, then the Fourier transform yields sums of Dirac delta functions at those frequencies. In practice, the textures appear as pairs of spikes (of various amplitudes) in the Fourier domain (see Figure 6.5). By counting the number of spikes in the Fourier domain, one can estimate an upper bound for the patch-rank (it is only an upper bound since jump discontinuities can contribute to the spikes).

Let $r \times r$ be the patch size of $\mathcal{P}v$, then we have the following:

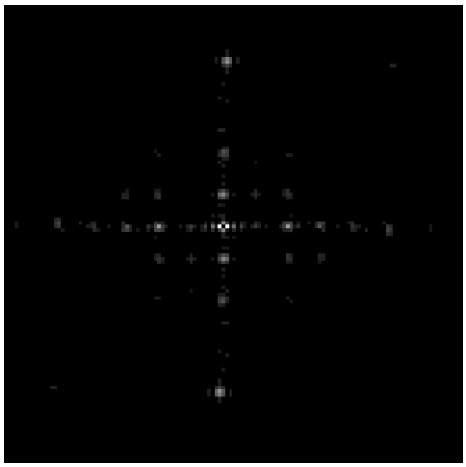


Figure 6.5: Fourier Transform of Brodatz

1. From the argument above, take r^2 to be close to half the number of "large" spikes in $|\mathcal{F}(v)|$.

2. From Example 3.8, if the largest texture period is T, then take r to be as close to T as possible.

3. Choose (μ, γ, λ) based on the theorems 6.8, 6.9, 6.10, and 6.11. Also, the ratio $\frac{\mu}{\lambda_1}$ determines the amount of cartoon, while the ratio $\frac{\gamma}{\lambda_2}$ determines the amount of texture.

4. Normalizing f such that $\sum |f_{i,j}| = 1$ (or $\max |f_{i,j}| = 1$) and $\sum f_{i,j} = 0$ helps when choosing (μ, γ, λ).

5. Setting $\lambda = \lambda_1 = \lambda_2$ gives appropriate results and removes two parameters from the model.

6. In regards to the algorithm, the number of outer loops determines the "amount of edges and texture" that is added back. If the original image is noisy, this parameter must be tuned in order to avoid adding back the noise. As mentioned in [37], the number of GS sweeps can be fixed to a value between 2 to 5 (only partially converging) and the method will give satisfactory results.

For the results that follow we take 5 by 5 patches (unless otherwise stated), $\lambda = 1$,

$\mu \in [.75, 2]$, and $\gamma \approx \sigma_{max}(\mathcal{P}v)$.

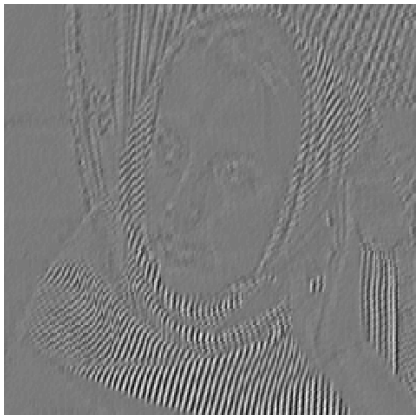
6.5.1 Decomposition

Figure 6.6: Barbara

For the pure cartoon-texture decomposition problem, we would like to remove the texture component without removing other key features, such as edges, shading, etc. Take, for example, the Barbara image (Figure 6.6). We estimated approximately 25 spikes in the Fourier domain, so a patch size of 5×5 is more than sufficient. In Figure 6.7, our method is compared to the standard $TV - G$ model [4] and the $TV - L^2$ model (i.e. ROF model). For $TV - L^2$, we use the Split Bregman approach from [37]. For all models, the parameters are chosen in order to have the same L^2 norm on the texture component. As seen in Figure 6.7, our method provides similar results to the classical $TV - G$ separation while removing fewer edges, for example, Barbara's hair and the background. Unlike our method, the $TV - L^2$ model does not remove the texture evenly.

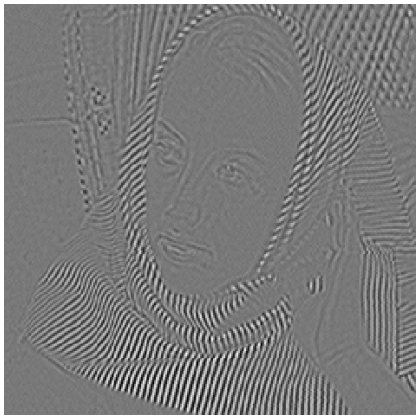
In Figure 6.8, the texture component's dependence on the patch size is examined. The similarity between the textures suggest some flexibility in determining the patch size, although the smaller patch size yields less texture while the large patch size removes non-texture features. In Figure 6.9, the texture component's dependence on γ (the texture norm's coefficient in the energy) is investigated by fixing μ and λ. Increasing γ gives smaller patch-ranks and removes less non-texture features, while decreasing γ removes more details from the cartoon (see theorem 6.8).

(a) Our Cartoon



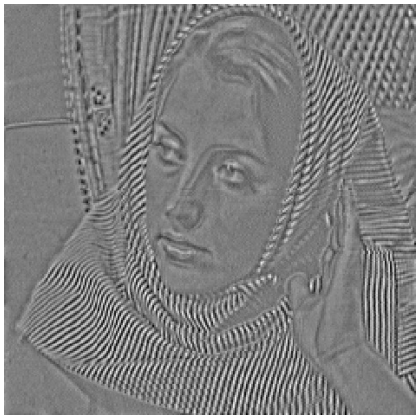
(b) Our Texture (patch-rank 6)

(c) $TV - G$ Cartoon



(d) $TV - G$ Texture

(e) $TV - L^2$ Cartoon



(f) $TV - L^2$ Texture

Figure 6.7: Cartoon-texture decomposition of Barbara

Fixing the texture from Figure 6.8d and decomposing the patch-form by the singular value decomposition yields: $\mathcal{P}v = U\Sigma V^*$, *i.e.* $v = \mathcal{P}^{-1}(U\Sigma V^*)$, where the diagonal matrix

(a) Cartoon p=4 (b) Texture p=4

(c) Cartoon p=5 (d) Texture p=5

(e) Cartoon p=7 (f) Texture p=7

Figure 6.8: Cartoon-texture decomposition of Barbara

$\Sigma = diag(\sigma_1, ..., \sigma_r, 0...0)$ contains the r singular values in descending order. In Figure 6.9, this texture is then reconstructed using only some singular vectors (thresholding Σ). Notice

(a) Cartoon (Patch-Rank 4) (b) Texture (Patch-Rank 4)

(c) Cartoon (Patch-Rank 6) (d) Texture (Patch-Rank 6)

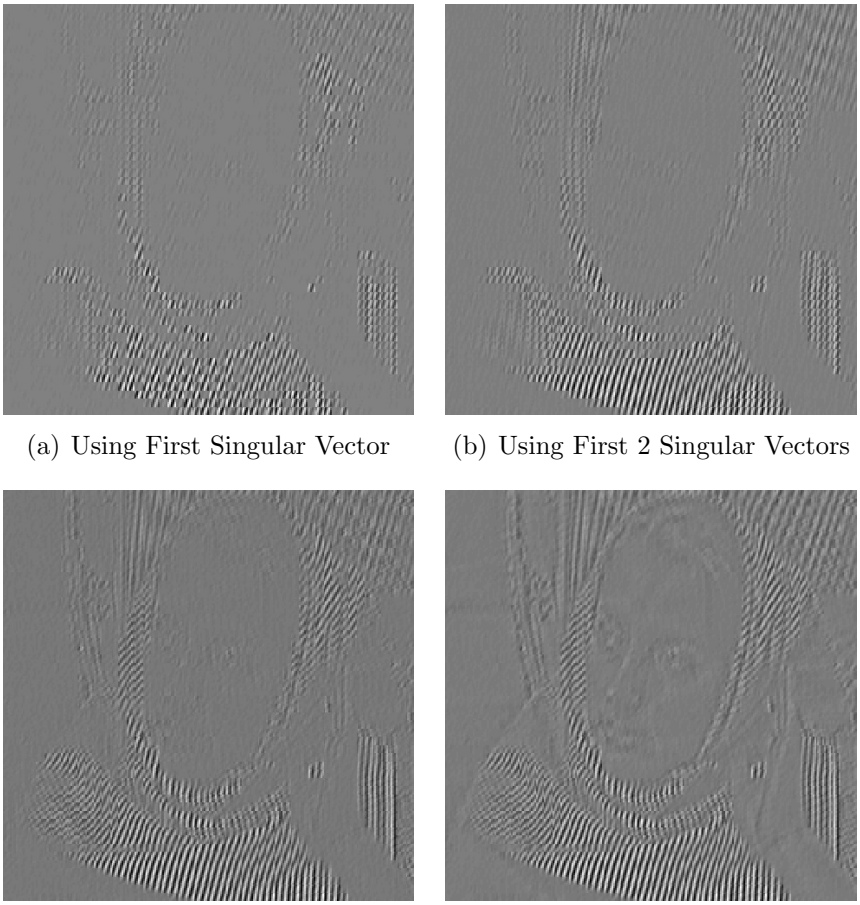
(e) Cartoon (Patch-Rank 7) (f) Texture (Patch-Rank 7)

Figure 6.9: Decompositions with different Patch-Ranks (fixed μ and λ, γ variable)

that the singular values capture the relative significances of the base textures.

Lastly, Figure 6.4 shows the wide range of textures that can be capture by our model. The

Tank and Grass images have visually random patterns yet have small patch-ranks (51.0% and 19.1% respectively).



(a) Using First Singular Vector (b) Using First 2 Singular Vectors

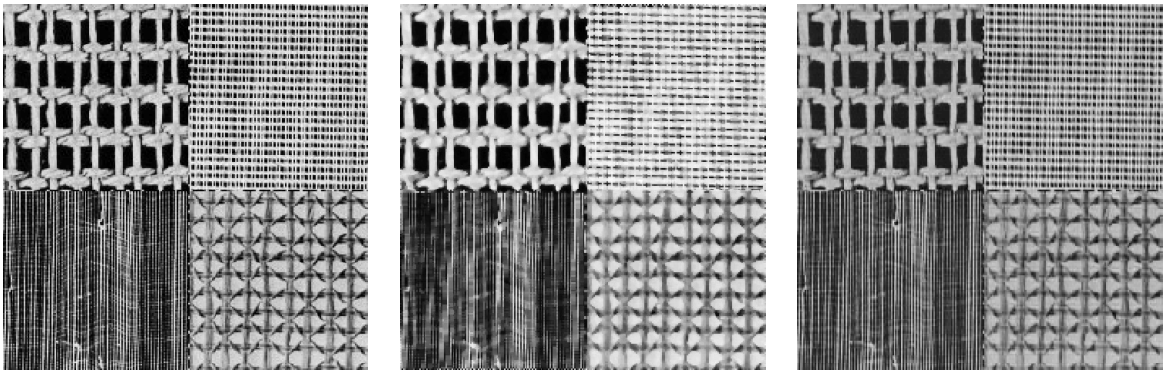
(c) Using First 4 Singular Vectors (d) Full Texture

Figure 6.10: Various Singular Vectors for a Fixed Texture

6.5.2 Pattern Regularization

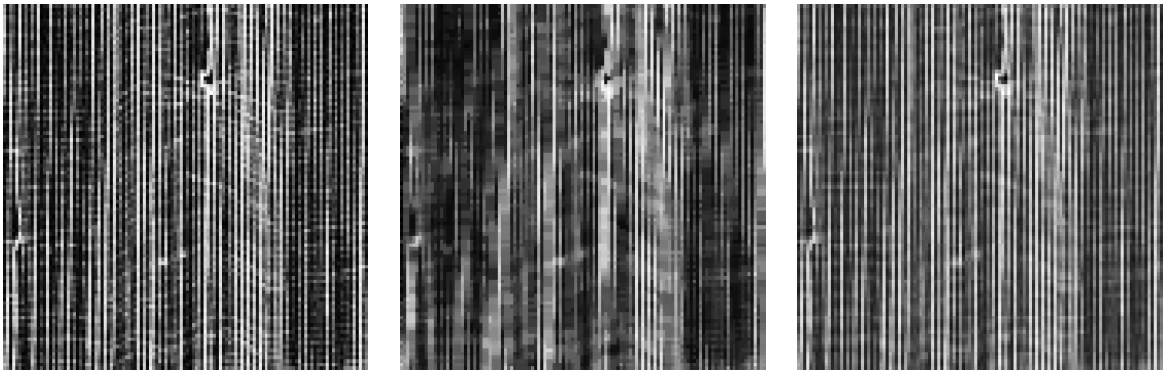
Since our texture norm penalizes non-uniform behavior between patches, our decomposition can be applied to image regularization. By regularizing the patterns, irregularities in the image are removed. In Figures 6.11 and 6.12, we regularize the highly textured Brodatz image using equation (6.3). The image f is decomposed into $f = u + v + \rho$, where $u + v$ is the regularized image and ρ contains the irregularities. In Figure 6.12, it can be seen that, in our method, the essential vertical pattern is preserved, while the unwanted point structures

are removed. Furthermore, our method also retains more texture than the standard median filter.

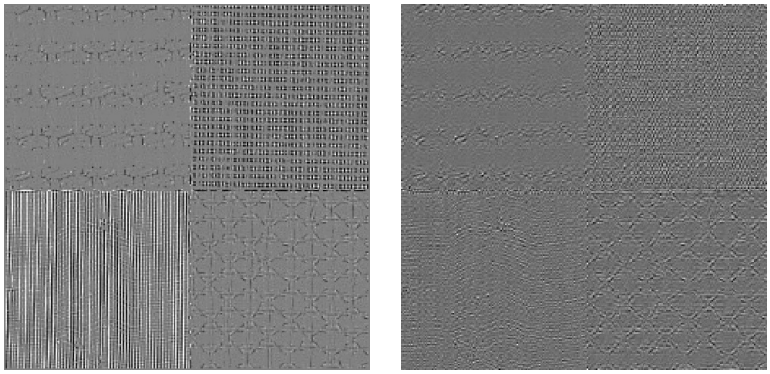


(a) Original Image (b) Median Filter (c) Our Method

Figure 6.11: Image Regularization



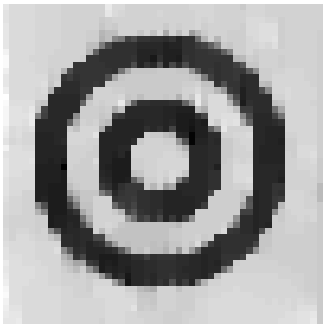
(a) Original-Zoomed In (b) Median Filter- Zoomed In (c) Our Method- Zoomed In



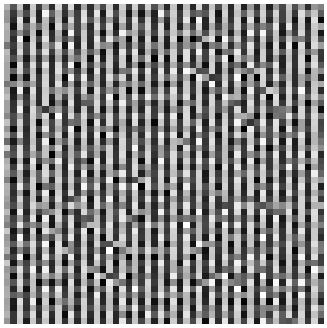
(d) Median Filter-Residual (e) Our Method-Residual

Figure 6.12: Image Regularization

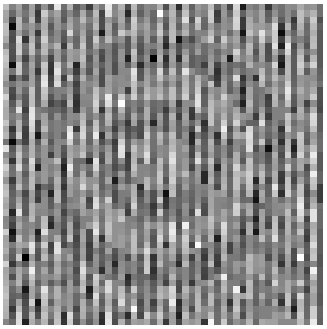
6.5.3 Denoising



(a) Cartoon



(b) Texture

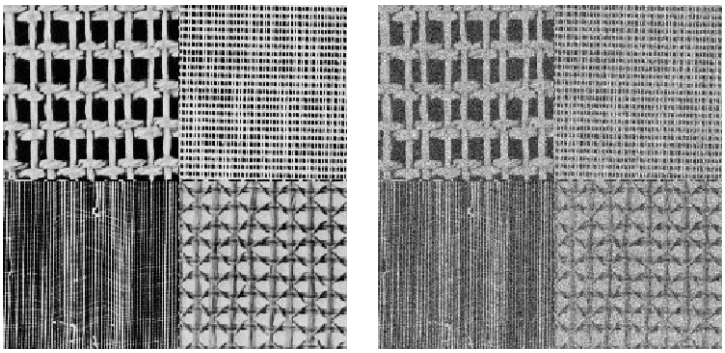


(c) Noise (residual)

Figure 6.13: Decomposition of a Synthetic Example Image using $TV - G$

In Figure 6.2, we saw the nearly-perfect recovery of a noisy synthetic image. Texture-noise separation is difficult for most cartoon-texture decomposition methods, see Figure 6.13. In Figures 6.14 and 6.15, the Brodatz image is degraded by Gaussian noise of zero mean and recovered by $TV - L^2$, $NLTV - L^2$, and our method. For $NLTV - L^2$, to make the comparison as similar as possible, we use a semi-local version with the same patch size as in our method [32, 33]. In both the $TV - L^2$ and the $NLTV - L^2$ models, the noise is removed from the top left quadrant. However, in more textured quadrants, more texture loss is seen (see Figure 6.15). For $TV - L^2$ (PSNR=30.2), although some texture is recovered, only the cartoon-like regions are recovered well. The $NLTV - L^2$ (PSNR=31.8) recovers both

the cartoon and texture well, but loses some texture in order to obtain a particular level of smoothness. Our method recovers both parts well, with the highest PSNR of 32.8, while being faster than the semi-local method (restricting the window size to 11 by 11). The slight oscillatory pattern that appears in the top left quadrant is an artifact of the non-locality of SVD.



(a) Original Image (b) Noisy Image

Figure 6.14: Decomposition of a Synthetic Example Image

6.5.3.1 Denoising: Quantitative Comparisons

Method	Recovered SNR	Cartoon SNR	Texture SNR	Noise Entropy	Error Entropy
Ours	**19.3**	**19.0**	**17.3**	**4.43**	**3.34**
TV-L^2	8.6	N/A	N/A	3.48	4.29
NLTV-L^2	19.1	N/A	N/A	4.42	3.50
TV-G-L^2	11.2	13.4	5.9	4.36	4.31
TV-H^{-1}-L^2	10.8	11.6	6.6	4.35	4.37

Table 6.1: Denoising of the Synthetic image. Each method has the same noise variance (L^2 norm). The entropy of the true noise is 4.53.

In general, there are no exact metrics in which to compare textures; however, there are some features which one prefers. In particular, a "good texture" component is noise-free and contains many sharp small-scale details. Since noise is dense and has many false spikes, we can measure the noise level in the texture component by computing the sparsity of the texture component (as a percent of total pixels) and the percent of pixels which are considered edges (after applying a gradient based edge detector). The smaller the value of both the edge and

Method	SNR	Edges	Sparsity	Patch-Rank	Noise Entropy	Error Entropy
Ours	**15.14**	**10.9%**	**61.8%**	**15.1%**	**3.81**	**3.72**
TV-L^2	14.37	N/A	N/A	N/A	3.75	3.78
NLTV-L^2	**15.14**	N/A	N/A	N/A	3.80	3.79
TV-G-L^2	11.13	16.4%	63.3%	83.9%	4.00	4.15
TV-H^{-1}-L^2	11.50	11.6%	64.4%	82.6%	4.05	4.14

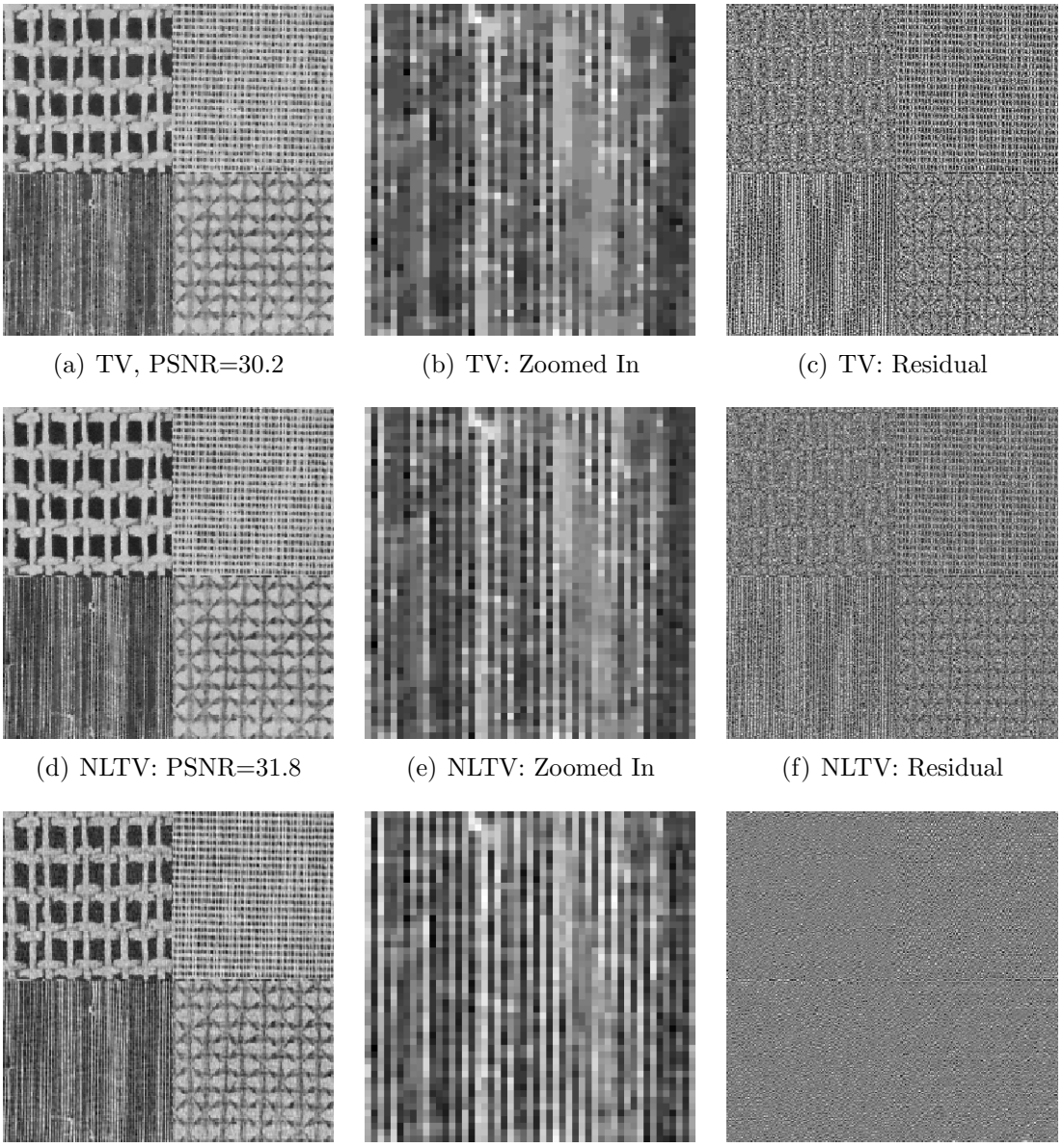
Table 6.2: Denoising of a noisy Brodatz-Wood image. Each method has the same noise variance (L^2 norm) and same L^2 norm on the texture component. The entropy of the true noise is 3.81. The patch-ranks for TV-G-L^2 and TV-H^{-1}-L^2 are computed after thresholding the smaller singular values.

Method	SNR	Edges	Sparsity	Patch-Rank	Noise Entropy	Error Entropy
Ours	**7.83**	**8.8%**	**66.5%**	**15.6%**	**3.79**	**3.28**
TV-G-L^2	4.26	10.6%	66.6%	98.1%	3.78	3.94
TV-H^{-1}-L^2	5.84	13.0%	67.3%	100%	3.78	3.78

Table 6.3: Denoising of the noisy Grass image. Each method has the same noise variance (L^2 norm) and same L^2 norm on the texture component. The entropy of the true noise is 3.80. The patch-ranks for TV-G-L^2 and TV-H^{-1}-L^2 are computed after thresholding the smaller singular values.

Method	SNR	Edges	Sparsity	Patch-Rank	Noise Entropy	Error Entropy
Ours	**11.86**	**7.2%**	**65.6%**	**40%**	**3.29**	**3.26**
TV-G-L^2	9.53	30.2%	67.6%	100%	3.35	3.49
TV-H^{-1}-L^2	9.57	0.3% *	67.8%	100%	3.34	3.49

Table 6.4: Denoising of the noisy Tank image. Each method has the same noise variance (L^2 norm) and the *exact* same cartoon component. The entropy of the true noise is 3.28. The patch-ranks for TV-G-L^2 and TV-H^{-1}-L^2 are computed after thresholding the smaller singular values. *Note: The TV-H^{-1}-L^2 decomposition has a texture component that is mostly low amplitude noise, so the gradient detector outputs almost no edges.



(a) TV, PSNR=30.2 (b) TV: Zoomed In (c) TV: Residual

(d) NLTV: PSNR=31.8 (e) NLTV: Zoomed In (f) NLTV: Residual

(g) Our Method: PSNR=32.8 (h) Our Method: Zoomed In (i) Our Method: Residual

Figure 6.15: Denoising of Brodatz Image

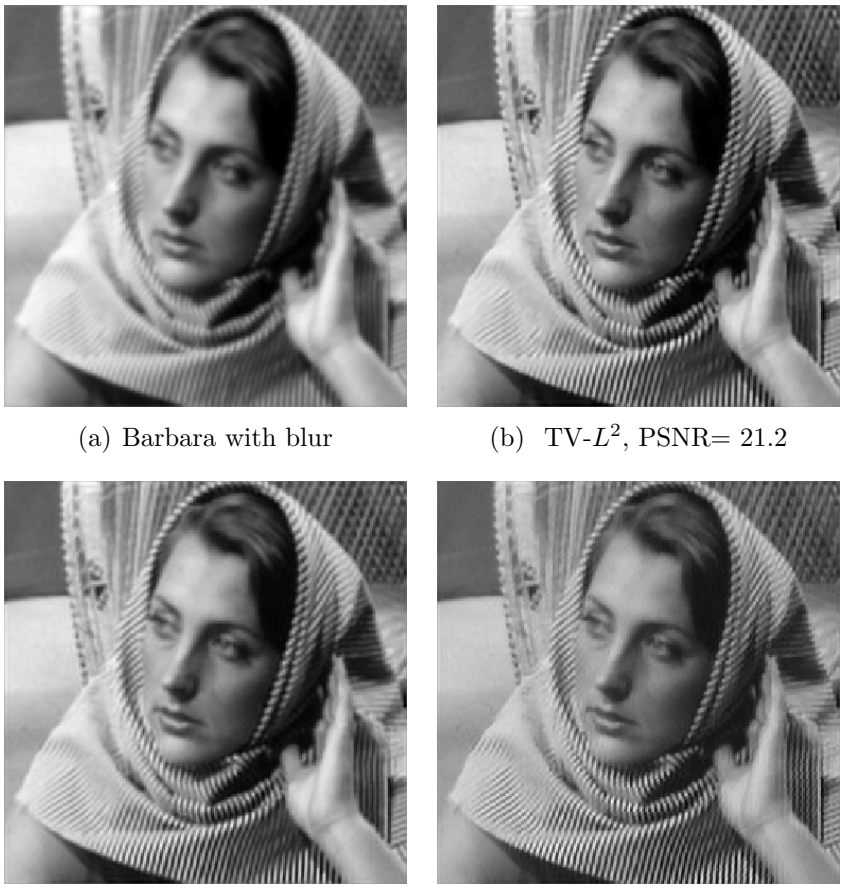
intensity sparsity, the better the texture. To measure the amount of noise that is removed by a method, we compute the entropy of the noise component and compare it to the known entropy of the added noise. The entropy is also applied to the error term to measure the amount of structure and texture that is lost to the noise component. A small error entropy means that the recovered image is closer to the original in terms of small-scale features (since

oscillations may appear as larger entropy). Lastly, for reference, the patch-rank is computed for each method. We compare our proposed texture method to standard ones using various texture and noise metrics.

In Table 6.1, the Synthetic image from Figure 6.2 is denoised by various methods. Each method has the same variance on their corresponding noise component. Since the true cartoon and true texture is known, we only compare the SNR of each component. Notice that both of the entropy metrics order the methods in a similar (but not exactly equivalent) way as the SNR. In Table 6.2, a noisy version of the Brodatz-Wood image is denoised. Our method does as well as the non-local methods in terms of SNR, but better in terms of removing the appropriate noise distribution (measured by entropy). Compared to the other standard cartoon-texture methods, our proposed model outputs a less noisy texture component. In the next two tables, we only compare our method with other cartoon-texture decomposition methods. In both cases, it is clear that our method better handles texture-noise separation. Using these results, we can partial conclude that this pattern-based interpretation is a more appropriate definition for texture than those that are based on oscillations.

6.5.4 Deblurring

Blurry images tend to have severe texture loss and require methods which reconstruct the texture well. In Figure 6.16, the Barbara image is blurred by a Gaussian kernel with a standard deviation of 1.1. We compare our recovered image with the TV-L^2 and NLTV-L^2 [77] deblurring methods. In terms of PSNR, our method better reconstructs the image with the Non-local total variational method getting very close results. Visually, the textures on the left and right ends of the scarf are sharper in our recovered image than in the others. Since the production of sharp edges are the key to deblurring methods, we can also quantitatively compare the results by measuring the percent of edge pixels in the image (using a gradient based edge detector). The blurred image has 15.4% edge pixels, the TV method has 16.3% edge pixels, the NLTV method has 18.6% edge pixels, while ours has the most at 19.2% edge pixels. Using this metric, we can conclude that our method better produces features



(a) Barbara with blur (b) TV-L^2, PSNR= 21.2

(c) NLTV-L^2, PSNR= 22.6 (d) Our Model, PSNR=22.7

Figure 6.16: Deblurring Comparison

of sharp contrast. On a side note, in practice the texture regularized images do not seem to have the typical ringing effect associated with deconvolution problems. This leads us to believe that texture regularized models are more stable for deblurring.

6.5.5 Inpainting: Sparse Reconstruction

In this and the next section, we discuss texture-regularized inpainting. There are two main types of inpainting: missing regions and sparse reconstruction. For inpainting missing regions, the image is first separated into its components, with the cartoon part recovered using a "structure" based inpainting, while the texture part is inpainted by texture-based techniques (for more on this methodology see [9]). The numerical results presented here focus

(a) Corrupted Image (b) Recovered

Figure 6.17: Inpainting Example

on sparse reconstruction. As an example, our method recovers an image almost perfectly (RMSE is less than 0.08) with more than 50% of the pixels randomly removed. In Figure 6.18 and 6.19, we compare our method to $TV - L^2$ and $NLTV - L^2$ inpainting on an image with more than 65% of the pixels randomly removed. Our method is comparable in speed to the $TV - L^2$ and faster than the $NLTV - L^2$. Our method recovers the image better (PSNR=44.0) than the $TV - L^2$ (PSNR=39.6), while being comparable to $NLTV - L^2$ (PSNR=43.4).

6.5.5.1 Sparse Reconstruction and Denoising: Quantitative Comparisons

Statistic	**Recovered RMSE**	**Original RMSE**	**RMSE Decrease**	**Patch Rank**
Minimum	0.035	0.091	-33.4%	5.5
Maximum	0.159	0.333	-83.0%	50.5
Mean	0.088	0.228	-61.0%	29.4
Median	0.085	0.231	-59.3%	30.9

Table 6.5: Statistics on our algorithm applied to 100 images from the database in [45].

Lastly, we investigate the type of textures which can be well-represented within this framework. To do so, we generate a set of corrupt highly textured images by adding a fixed amount of noise to 100 images from [45] and then sparse sampling these images (removing half of the number of pixels). Our algorithm was applied to all of the corrupt images with parameters set at around $\mu = 1$, $\lambda = \lambda_1 = \lambda_2 = 1$, 3 GS sweeps, 5 inner iterations, and

(a) Original Image (b) Corrupted Image, less than 35% retained (c) $TV - L^2$, PSNR=39.6

(d) $NLTV - L^2$, PSNR=43.4 (e) Our Method, PSNR=44.0

Figure 6.18: Inpainting Comparison

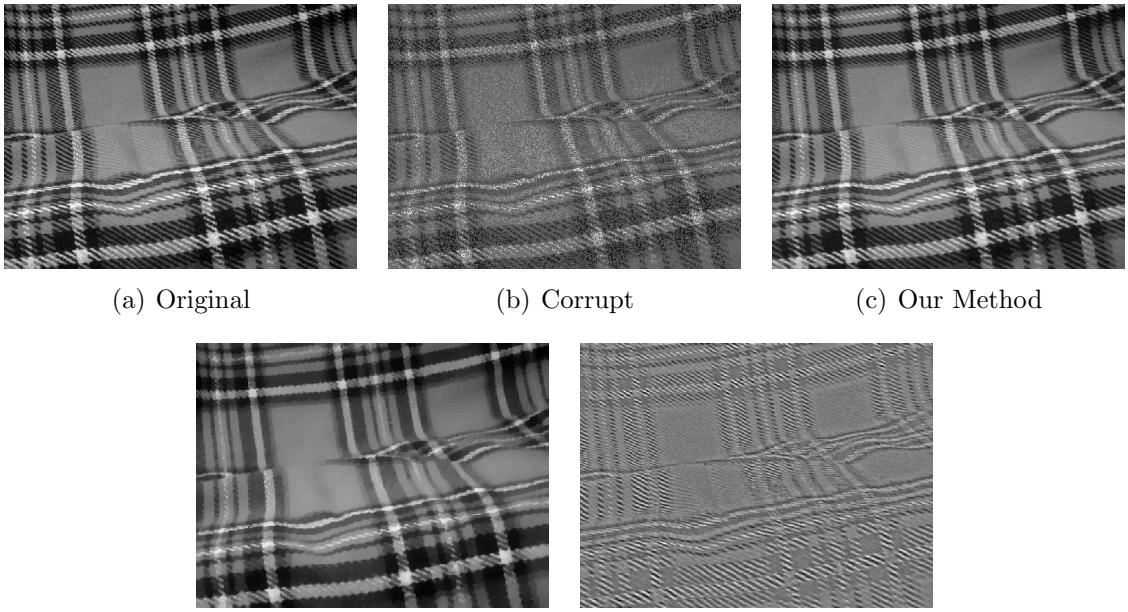
15 outer iterations. The patch size varied from 10 by 10 to 20 by 20 to match the texture scale of each individual image. The parameters were choosen to yield patch-ranks under 50. Table 6.5 displays statistics on the RMSE between the recovered image and the original, the RMSE between the original and the corrupt image, the precentage decrease in RMSE after the recovery, and the patch-rank (which is normalized out of 100). By considering the RMSE, the amount the RMSE changed, the patch rank, and visual metrics, we can evaluate which textures are well-represented by our method.

Although the patches are formed on a grid, the resulting textures do not have to be grid like in structure nor in pattern. Based on this experiment, our method is not sensitive to the angle or directionality of the texture pattern, the texture plane (*i.e.* frontal verse non-frontal textures), the geometry of the pattern, nor the geometry of the texture plane. For example, in Figure 6.20, the texture resides on a warped or non-uniform geometry, which

(a) $TV - L^2$ (b) $NLTV - L^2$ (c) Our Method

Figure 6.19: Inpainting Comparison: Zoomed in

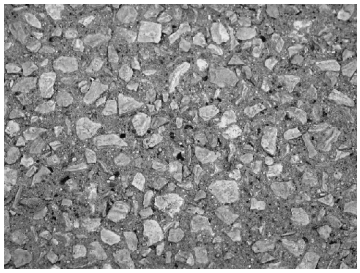
does not align with the patch grid. However, our method preforms well on this textured image and others like this. Our method has difficulty in capturing texture which does not have a regular pattern or has highly deviating structures. An example of this can be found in Figure 6.21, where the patterns have many irregular shapes over many scales. In this example, the texture component does not have a geometrically uniform pattern and we can see loss of texture in the recovered image.



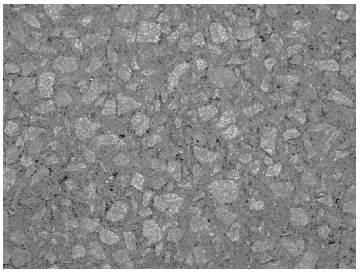
(a) Original (b) Corrupt (c) Our Method

(d) Our Cartoon (e) Our Texture

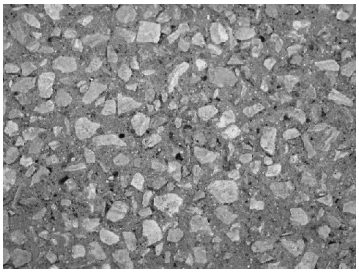
Figure 6.20: Joint Sparse Reconstruction and Denoising: Warped/Non-Uniform Geometry



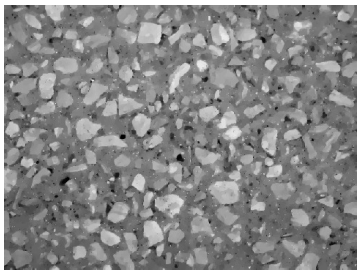
(a) Original



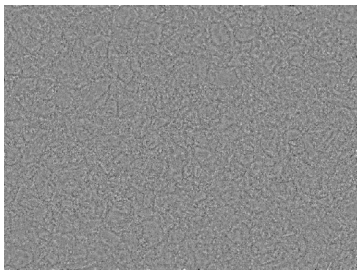
(b) Corrupt



(c) Our Method



(d) Our Cartoon



(e) Our Texture

Figure 6.21: Joint Sparse Reconstruction and Denoising: Multiscale with Irregularities

CHAPTER 7

Some Theoretical Remarks on Low Patch-Rank Textures

7.1 A Tensor Interpretation

As a small remark, since the patch map is a linear map between matrices it can be associated with a tensor. The map $\mathcal{P}$ is a rank 4 tensor (not to be confused with matrix rank) which maps $\mathbb{R}^{n,n}$ into $\mathbb{R}^{r^2,\frac{n^2}{r^2}}$. Associating the map $\mathcal{P}$ with the tensor $P = \left[P_{i,j}^{k,l}\right]$, the operation $\mathcal{P} : v \mapsto w$ is defined by the following (using Einstein notation, where repeated indices are summed):

$$w_{i,j} = P_{i,j}^{k,l}\ v_{k,l} \tag{7.1}$$

where the elements of the tensor are defined as:

$$P_{i,j}^{k,l} = \begin{cases} 1: & \text{if } (i,j) = \left((m(k)-1)r + m(l),\ 1 + \frac{n(k-m(k))}{r^2} + \frac{l-m(l)}{r}\right) \\ 0: & otherwise \end{cases}$$

and $m(\cdot) := mod(\cdot - 1, r) + 1$.

The tensor is bijective since it is a unique invertible mapping. Furthermore, the tensor is a unitary transformation with respect to the element-wise inner product, since for all v_1

and v_2, $\langle \mathcal{P}v_1, \mathcal{P}v_2 \rangle = \langle v_1, v_2 \rangle$ holds.

7.2 Proofs of Theorems

The proofs below are a generalization of the theory from Gilles and Meyer's work [35]. The main difference is that extra care must be given since there is no duality between our cartoon and texture norms.

Lemma 7.1. *Take any Banach space with norm $||\cdot||$ and dual norm $||\cdot||'$. Given an f that is decomposed into $f = u + v$ by minimizing the following:*

$$||u|| + \frac{\lambda}{2}||v||_2^2$$

then the following hold.

1. *If $||f||' \leq \frac{1}{\lambda}$, then $u = 0$ and $v = f$.*
2. *If $||f||' > \frac{1}{\lambda}$, then $||v||' = \frac{1}{\lambda}$ and $\langle u, v \rangle = \frac{1}{\lambda}||u||$.*

Lemma 7.1 will be used in many of the arguments in the various proofs. Similarly to the work [35], the cartoon, texture, and error norms have a proper ordering.

Lemma 7.2. *For all n by n matrices g (with mean zero), we have the following:*

$$||\mathcal{P}g||_* \leq \sqrt{n}||g||_2 \leq \frac{n}{2}||Dg||_1$$

The lemma above provides a comparison between the terms in the energy and allows for the generalization of duality. More precisely, Lemma 7.3 describes the dual pairing relationships.

Lemma 7.3. *For all n by n matrices u and v, we have the following:*

$$\begin{aligned} |\langle u, v \rangle| &\leq ||Du||_1||v||_G \\ |\langle u, v \rangle| &\leq ||\mathcal{P}u||_s||\mathcal{P}v||_* \end{aligned}$$

Proof. For the first inequality, take any (u, v) and let there exist a g such that $v = \text{div}g$, so

$$\begin{aligned} |\langle u, v\rangle| &= |\langle u, \text{div}g\rangle| \\ &= |\langle Du, g\rangle| \\ &\leq ||Du||_1 ||g||_\infty \end{aligned}$$

Then taking the infimum with respect to all possible g yields the inequality. For the second inequality, we use the properties of $\mathcal{P}$.

$$\begin{aligned} |\langle u, v\rangle| &= |\langle \mathcal{P}u, \mathcal{P}v\rangle| \\ &\leq ||\mathcal{P}u||_s ||\mathcal{P}v||_* \end{aligned}$$

Note that the spectral and trace norms are dual.

□

In order to apply Lemma 7.1 to our model, we look at the pair (u, v) simultaneously. To do so, define $w = u + v$ to be the reconstructed image comprised of both components with the following induced norm: $||w|| = \inf\{\mu||Du||_1 + \gamma||\mathcal{P}v||_*\}$. The dual norm is defined as $||g||' = \sup\left\{\frac{1}{\mu}||g||_G, \frac{1}{\gamma}||\mathcal{P}g||_s\right\}$. Writing the energy in terms of w yields:

$$\mathcal{E}(w) = ||w|| + \frac{\lambda}{2}||f - w||_2^2$$

we will consider this in many of the proofs that follow.

Throughout this appendix, the energy $\mathcal{E}$ will be a functional with arguments u, v, and ρ depending on the particular proof; however, all are equivalent by the relationship $u+v+\rho = f$ with f given.

Theorem 7.4. *If $0 < \gamma < \frac{2\mu}{n}$ then the optimal decomposition yields u=0.*

Proof. Examining the energy with respect to the cartoon component u and the residual ρ, the energy can be bounded below by using Lemma 8.2 and positivity of the norms.

$$\begin{aligned}
\mathcal{E}(u,\rho) &= \mu||Du||_1 + \gamma||\mathcal{P}(f-u-\rho)||_* + \frac{\lambda}{2}||\rho||_2^2 \\
&\geq \frac{2\mu}{n}||\mathcal{P}u||_* + \gamma||\mathcal{P}(f-u-\rho)||_* + \frac{\lambda}{2}||\rho||_2^2 \\
&\geq \gamma\left\{||\mathcal{P}u||_* + ||\mathcal{P}(f-u-\rho)||_*\right\} + \frac{\lambda}{2}||\rho||_2^2 \\
&\geq \gamma||\mathcal{P}(f-\rho)||_* + \frac{\lambda}{2}||\rho||_2^2 \\
&= \mathcal{E}(0,\rho)
\end{aligned}$$

For all $u \neq 0$ we have $||\mathcal{P}u||_* > 0$, therefore $u = 0$ is the minimizer. □

Next, using theorem 7.4 with an additional constraint on f yields another trivial minimizer.

Theorem 7.5. *If $0 < \gamma < \frac{2\mu}{n}$ and $||\mathcal{P}f||_s \leq \frac{\gamma}{\lambda}$ then the optimal decomposition yields $u = \rho = 0$ and $v = f$.*

Proof. From theorem 7.4, since $0 < \gamma < \frac{2\mu}{n}$, the energy is equivalent to the following:

$$\mathcal{E}(0,\rho) = \gamma||\mathcal{P}(f-\rho)||_* + \frac{\lambda}{2}||\rho||_2^2$$

By Lemma 7.1, since $||\mathcal{P}f||_s \leq \frac{\gamma}{\lambda}$, then ρ is identically 0. □

The theorems above provide the conditions in which the texture and/or residual component contain all of the information. The theorem below provides the final trivial case.

Theorem 7.6. *If* $||f||_G \leq \frac{\mu}{\lambda}$ *and* $||\mathcal{P}f||_s \leq \frac{\gamma}{\lambda}$ *then the optimal decomposition yields* $u = v = 0$ *and* $\rho = f$.

Proof. Consider the simultaneous energy (with respect to w).

$$\mathcal{E}(w) = ||w|| + \frac{\lambda}{2}||f - w||_2^2$$

The value of f in the dual norm can be calculated by using the assumptions:

$$\begin{aligned} ||f||' &= \sup\left\{\frac{1}{\mu}||f||_G, \frac{1}{\gamma}||\mathcal{P}f||_s\right\} \\ &\leq \sup\left\{\frac{1}{\lambda}, \frac{1}{\lambda}\right\} \\ &= \frac{1}{\lambda}. \end{aligned}$$

Applying Lemma 7.1 yields $w = 0$, which implies both $u = 0$ and $v = 0$. □

For the first non-trivial decomposition, theorem 7.7 characterizes minimizers when f is large with respect to particular norms.

Theorem 7.7. *If* $||f||_G > \frac{\mu}{\lambda}$ *and* $||\mathcal{P}f||_s > \frac{\gamma}{\lambda}$*, then the following holds for all minimizers:* $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$, *and* $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$.

Proof. Calculating the simultaneous dual norm yields the lower bound $||f||' \geq \frac{1}{\lambda}$ and therefore by Lemma 7.1 $||\rho||' = \frac{1}{\lambda}$ and $\langle w, \rho \rangle = \frac{1}{\lambda}||w||$. Since the simultaenous dual norm is defined as $||\rho||' = \sup\left\{\frac{1}{\mu}||\rho||_G, \frac{1}{\gamma}||\mathcal{P}\rho||_s\right\}$, we have the following bounds on the residual ρ:

$$\begin{aligned} ||\rho||_G &\leq \frac{\mu}{\lambda} \\ ||\mathcal{P}\rho||_s &\leq \frac{\gamma}{\lambda} \end{aligned}$$

By the duality principles from lemma 7.3, the following also holds:

$$\begin{aligned} \langle \rho, u \rangle &\leq ||Du||_1 ||\rho||_G = \frac{\mu}{\lambda} ||Du||_1 \\ \langle \rho, v \rangle &\leq ||\mathcal{P}v||_* ||\mathcal{P}\rho||_* = \frac{\gamma}{\lambda} ||\mathcal{P}v||_* \end{aligned}$$

Together, these inequalities give:

$$\langle u + v, \rho \rangle \leq \frac{1}{\lambda} \{\mu ||Du||_1 + \gamma ||\mathcal{P}v||_*\}$$

The inequality above is equivalent to $\langle w, \rho \rangle \leq \frac{1}{\lambda}||w||$, for $w = u + v$. However, by Lemma 7.1 $\langle w, \rho \rangle = \frac{1}{\lambda}||w||$, thus equality holds for all related inequalities: $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$, and $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$.

□

Lastly, when f is small in one dual norm and large in the other many cases occur. The following theorem provides the various optimal solutions and their characterizations.

Theorem 7.8. *If $||f||_G \leq \frac{\mu}{\lambda}$ and $||\mathcal{P}f||_s > \frac{\gamma}{\lambda}$ then three case hold:*

(1) $u = 0$, $||\rho||_G < \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, and $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$

(2) $v = 0$, $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s < \frac{\gamma}{\lambda}$, and $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$

(3) $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$, and $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$

Each of the case above are also the optimal decompositions for the energy.

Proof. Consider the simultaneous energy in terms of w:

$$\mathcal{E}(w) = ||w|| + \frac{\lambda}{2}||f - w||_2^2$$

Under the assumptions of this theorem, the simultaneous dual norm of f is bounded below: $||f||' > \frac{1}{\lambda}$, and by Lemma 7.1, $||\rho||' = \frac{1}{\lambda}$ and $\langle \rho, w \rangle = \frac{1}{\lambda}||w||$. This can occur in three ways. The proof is divided into several subproofs for each of the cases.

1. *First assume* $||\rho||_G < \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$.

   Since $w = u + v$, the result of lemma 7.1 is equivalent to:

$$\langle \rho, u + v \rangle = \frac{1}{\lambda} \{\mu||Du||_1 + \gamma||\mathcal{P}v||_*\}$$

   However, by the duality principles and the assumptions of this case the following inequalities hold:

$$\begin{aligned} \langle \rho, u \rangle &\leq ||Du||_1||\rho||_G < \frac{\mu}{\lambda}||Du||_1 \\ \langle \rho, v \rangle &\leq ||\mathcal{P}v||_*||\mathcal{P}\rho||_* = \frac{\gamma}{\lambda}||\mathcal{P}v||_* \end{aligned}$$

   Combining these statements produces the contradictory strict inequality:

$$\langle \rho, u + v \rangle < \frac{1}{\lambda} \{\mu||Du||_1 + \gamma||\mathcal{P}v||_*\}$$

   Thus $u = 0$ and $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$ to avoid the contradiction.

2. *Next, assume* $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s < \frac{\gamma}{\lambda}$.

   By repeating the argument above, $v = 0$ $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$ must hold.

3. *Assume that* $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$ *then* $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$, *and* $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$ *for the optimal solution.*

   Similarly to the other two cases: By Lemma 7.1, we have

$$\langle \rho, u + v \rangle = \frac{1}{\lambda} \{\mu||Du||_1 + \gamma||\mathcal{P}v||_*\}$$

   The duality principles (Lemma 7.3) and the assumptions of this case yield the following inequalities:

$$\langle \rho, u \rangle \leq ||\rho||_G||Du||_1 = \frac{\mu}{\lambda}||Du||_1$$
$$\langle \rho, v \rangle \leq ||\mathcal{P}\rho||_s||\mathcal{P}v||_* = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$$

   Thus the following equalities $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$ and $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$ must hold.

4. *For the next three cases, we show that the solutions are optimal. Assume* $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s \leq \frac{\gamma}{\lambda}$, *and* $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$, *and* $f = u + \rho$ *then for all* U *and* v *we have*

$$\mu||D(u + U)||_1 + \gamma||\mathcal{P}v||_* + \frac{\lambda}{2}||\rho - U - v||_2^2 \geq \mu||Du||_1 + \frac{\lambda}{2}||\rho||_2^2 \tag{7.2}$$

To show this, first rescale the equation (17) by dividing by $\lambda = \mu||\rho||_G^{-1}$ to get

$$\begin{aligned} ||D(u+U)||_1||\rho||_G &+ \frac{\gamma}{\mu}||\mathcal{P}v||_*||\rho||_G + \frac{1}{2}||\rho - U - v||_2^2 \\ &\geq \langle u+U, \rho\rangle + \frac{\gamma}{\mu}||\mathcal{P}v||_*||\rho||_G + \frac{1}{2}||\rho||_2^2 + \frac{1}{2}||U+v||_2^2 - \langle \rho, U\rangle - \langle v, \rho\rangle \\ &\geq \langle u, \rho\rangle + \frac{\gamma}{\mu}||\mathcal{P}v||_*||\rho||_G + \frac{1}{2}||\rho||_2^2 + \frac{1}{2}||U+v||_2^2 - \langle v, \rho\rangle \\ &= \frac{\mu}{\lambda}||Du||_1 + \frac{\gamma}{\mu}||\mathcal{P}v||_*||\rho||_G + \frac{1}{2}||\rho||_2^2 + \frac{1}{2}||U+v||_2^2 - \langle v, \rho\rangle \\ &\geq \frac{\mu}{\lambda}||Du||_1 + \frac{1}{2}||\rho||_2^2 \end{aligned} \tag{7.3}$$

By the texture duality principle, *i.e.* $\langle v, \rho\rangle \leq ||\mathcal{P}\rho||_s||\mathcal{P}v||_*$, and the assumptions of this case, we have $\langle v, \rho\rangle \leq \frac{\gamma}{\mu}||\rho||_G||\mathcal{P}v||_*$ which is used in the inequalities above.

If equality holds, then $U = -v$ and $\langle v, \rho\rangle = \frac{\gamma}{\mu}||\rho||_G||\mathcal{P}v||_*$. Also $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$ and $\langle v, \rho\rangle = ||\mathcal{P}\rho||_s||\mathcal{P}v||_*$ would hold. Returning to equation (17):

$$\mu||D(u-v)||_1 + \gamma||\mathcal{P}v||_* + \frac{\lambda}{2}||\rho||_2^2 \geq \mu||Du||_1 + \frac{\lambda}{2}||\rho||_2^2$$

by canceling terms, we get:

$$\mu||D(u-v)||_1 + \gamma||\mathcal{P}v||_* \geq \mu||Du||_1$$

Using the equalities we found and the fact that $\gamma = \lambda||\mathcal{P}\rho||_s$, the following inequalities

are equivalent:

$$\begin{aligned}
\mu||D(u-v)||_1 + \lambda||\mathcal{P}\rho||_s||\mathcal{P}v||_* &\geq \mu||Du||_1 \\
\frac{\mu}{\lambda}||D(u-v)||_1 + \langle v, \rho\rangle &\geq \frac{\mu}{\lambda}||Du||_1 \\
||\rho||_G||D(u-v)||_1 + \langle v, \rho\rangle &\geq \frac{\mu}{\lambda}||Du||_1
\end{aligned}$$

However, by the Lemma 7.3:

$$||\rho||_G||D(u-v)||_1 \geq \langle u-v, \rho\rangle = \langle u, \rho\rangle - \langle v, \rho\rangle$$

which implies $\langle u, \rho\rangle = \frac{\mu}{\lambda}||Du||_1$, so if we have equality then $||\rho||_G||D(u-v)||_1 = \langle u-v, \rho\rangle$.

5. *Assume $f = v+\rho$, $||\rho||_G < \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, and $\langle \rho, v\rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$, then the decomposition $f = v + \rho$ is optimal.*

   As before, perturb v by V, with the corresponding energy:

$$\mathcal{E}(u, v+V) = \mu||Du||_1 + \gamma||\mathcal{P}(v+V)||_* + \frac{\lambda}{2}||\rho - u - V||_2^2 \tag{7.4}$$

   By Lemma 7.3 and the assumptions of this case, we have the following inequality:

$$\begin{aligned}
||\mathcal{P}(v+V)||_*||\mathcal{P}\rho||_s &\geq \langle v+V, \rho\rangle \\
&\geq \langle v, \rho\rangle + \langle V, \rho\rangle \\
&= \frac{\gamma}{\lambda}||\mathcal{P}v||_* + \langle V, \rho\rangle
\end{aligned}$$

Since we assume $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, the previous inequality is equivalent to:

$$||\mathcal{P}(v+V)||_* \;\geq\; ||\mathcal{P}v||_* + \frac{\lambda}{\gamma}\langle V, \rho\rangle$$

Note that the decompositions of f, namely $f = u + v + V + \rho$ and $f = \rho + v$, yields:

$$||u+V||_2^2 = ||f-\rho-v||_2^2 = 0$$

Using this fact and expanding the L^2 norm in equation (7.4) gives:

$$||\rho - u - V||_2^2 = ||\rho||_2^2 - 2\langle u, \rho\rangle - 2\langle V, \rho\rangle \tag{7.5}$$

The second term on the right hand side of equation (7.5) can be bounded by using lemma 7.3 and the assumption $||\rho||_G < \frac{\mu}{\lambda}$:

$$\lambda\,|\langle u, \rho\rangle| \leq \lambda ||Du||_1 ||\rho||_G \leq \mu ||Du||_1$$

Returning to the equation (7.4), we have the following lower bound:

$$\begin{aligned}
\mathcal{E}(u, v+V) &\geq \lambda\,|\langle u, \rho\rangle| + \gamma||\mathcal{P}v||_* + \lambda\langle V, \rho\rangle + \frac{\lambda}{2}||\rho||_2^2 - \lambda\langle u, \rho\rangle - \lambda\langle V, \rho\rangle \\
&\geq \gamma||\mathcal{P}v||_* + \frac{\lambda}{2}||\rho||_2^2 \\
&= \mathcal{E}(0, v)
\end{aligned}$$

Thus the decomposition $f = v + \rho$ is optimal.

6. *Lastly, assume* $||\rho||_G = \frac{\mu}{\lambda}$, $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, $\langle \rho, v \rangle = \frac{\gamma}{\lambda}||\mathcal{P}v||_*$, *and* $\langle \rho, u \rangle = \frac{\mu}{\lambda}||Du||_1$ *hold, then the decomposition is optimal.*

   Perturb the functions u by U and ρ by P with the perturbed energy:

$$\mathcal{E}(u + U, \rho + P) = \mu||Du||_1 + \gamma||\mathcal{P}(v - U - P)||_* + \frac{\lambda}{2}||\rho + P||_2^2 \tag{7.6}$$

   Next, by combining the various assumptions in this case, we get $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda} > ||\mathcal{P}f||_s$ and $||\rho||_G = \frac{\mu}{\lambda} \geq ||f||_G$.

   For the texture norm, Lemma 7.3 yields:

$$||\mathcal{P}\rho||_s||\mathcal{P}(v - U - P)||_* \geq \langle \rho, v - U - P \rangle$$

   By using this inequality and the assumption that $||\mathcal{P}\rho||_s = \frac{\gamma}{\lambda}$, the following holds:

$$\gamma||\mathcal{P}(v - U - P)||_* \geq \lambda \langle \rho, v \rangle - \lambda \langle \rho, U \rangle - \lambda \langle \rho, P \rangle \tag{7.7}$$

   With respect to the cartoon norm, Lemma 7.3 yields:

$$||D(u + U)||_1||\rho||_G \geq \langle u + U, \rho \rangle \tag{7.8}$$

   By the assumption $||\rho||_G = \frac{\mu}{\lambda}$, equation (7.8) is equivalent to:

$$\mu||D(u + U)||_1 \geq \lambda \langle u, \rho \rangle + \lambda \langle U, \rho \rangle \tag{7.9}$$

   Combining equation (7.8) and (7.9) with equation (7.6) and canceling terms gives the

following lower bound for the energy:

$$\begin{aligned}
\mathcal{E}(u+U,\rho+P) &= \mu||Du||_1+\gamma||\mathcal{P}(v-U-P)||_*+\frac{\lambda}{2}||\rho+P||_2^2 \\
&\geq \lambda\left(\langle u,\rho\rangle+\langle U,\rho\rangle+\langle\rho,v\rangle-\langle\rho,U\rangle-\langle\rho,P\rangle+\frac{1}{2}||\rho||_2^2+\frac{1}{2}||P||_2^2+\langle P,\rho\rangle\right) \\
&\geq \lambda\left(\langle u+v,\rho\rangle+\frac{1}{2}||\rho||_2^2+\frac{1}{2}||P||_2^2\right) \\
&= \mu||Du||_1+\gamma||\mathcal{P}v||_*+\lambda\left(\frac{1}{2}||\rho||_2^2+\frac{1}{2}||P||_2^2\right) \\
&\geq \mathcal{E}(u,\rho)
\end{aligned}$$

Implying $U=0$ and $P=0$, therefore (u,ρ) is the minimizer.

□

CHAPTER 8

Conclusion

In Part I, an extension to the level set based image segmentation method that detects free endpoint structures was described. By generalizing the curve representation used in [18, 74] to also include free endpoint structures, we are able to segment a larger class of images with a variety of edge structures. This method is able to change its topology by splitting, merging, and now breaking curves without endpoints into free curves and vice versa. The results were tested on both synthetic and real images and, in the examples presented here, were more successful in locating the correct edge set as compared to standard methods. These results show that we can now extract and enhance a wider range of edge types, which could provide a better understanding of jump discontinuities present in any image.

In Part II, a detailed look at two different texture models was presented. The first texture model is described as a weak texture norm, based on duality with the space of Lipschitz functions. To measure the $W^{1,\infty}$ semi-norm, a concentration of measure form of the L^∞ norm was presented. The texture norm was used, in combination with the TV semi-norm, to regularize the deblurring problem. The method was applied to the known and semi-known deblurring problems, both showing improved results (in terms of SNR) and advantages over TV regularization alone.

Lastly, this thesis presented a completely different way to model image texture. Many types of texture can be measured by the number of repetitive patterns present in the image. From the matrix perspective, the nuclear norm on the patch collection is shown to be an appropriated norm. This texture norm was used as a regularizer in models for decomposition, pattern regularization, denoising, deblurring, and sparse reconstruction. For both of the

texture spaces, the various results support the additional use of a texture regularizer rather than only reconstructing the smooth component.

References

[1] L. Ambrosio and V. M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via t-convergence. Communications on Pure and Applied Mathematics, 43(8):999–1036, December 1990.

[2] G. Aubert and P. Kornprobst. Mathematical Problems in Image Processing Partial Differential Equations and the Calculus of Variations. Number 147 in Applied mathematical sciences. Springer, New York, 2nd ed edition, 2006.

[3] J-F. Aujol. Contribution à l'analyse de textures en traitement d'images par méthodes variationnelles et équations aux dérivées partielles. PhD thesis, June 2004.

[4] J-F Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. Journal of Mathematical Imaging and Vision, 22:71–88, January 2005.

[5] J-F. Aujol and T. F. Chan. Combining geometrical and textured information to perform image classification. Journal of Visual Communication and Image Representation, 17(5):1004–1023, October 2006.

[6] J-F. Aujol and S. H. Kang. Color image decomposition and restoration. Journal of Visual Communication and Image Representation, 17(4):916–928, August 2006.

[7] L. Bar and G. Sapiro. Generalized Newton-Type methods for energy formulations in image processing. SIAM Journal on Imaging Sciences, 2(2):508, 2009.

[8] L. Bar, N. Sochen, and N. Kiryati. Semi-blind image restoration via Mumford-Shah regularization. Image Processing, IEEE Transactions on, 15(2):483 –493, February 2006.

[9] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. IEEE Transactions on Image Processing, 12(8):882– 889, August 2003.

[10] A. Buades, B. Coll, and J-M. Morel. On image denoising methods. SIAM Multiscale Modeling and Simulation, 4(2):490–530, 2005.

[11] A. Buades, B. Coll, and J-M. Morel. Neighborhood filters and PDE's. Numerische Mathematik, 105:1–34, September 2006.

[12] J-F. Cai and S. Osher. Fast singular value thresholding without singular value decomposition. UCLA CAM Report, May 2010.

[13] E. J. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? Journal of ACM, 58F, 2009.

[14] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. International Journal of Computer Vision, 22(1):61–79, 1997.

[15] V. Caselles, J.-M. Morel, and C. Sbert. An axiomatic approach to image interpolation. IEEE Trans. Image Processing, 7:376–386, 1999.

[16] A. Chambolle and P-L. Lions. Image recovery via total variation minimization and related problems. Numerische Mathematik, 76:167–188, April 1997.

[17] T. F. Chan, B. Y. Sandberg, and L. A. Vese. Active contours without edges for vector-valued images. Journal of Visual Communication and Image Representation, 11(2):130–141, June 2000.

[18] T.F. Chan and L.A. Vese. Active contours without edges. IEEE Transactions on Image Processing, 10(2):266–277, February 2001.

[19] G. Chung and L. A. Vese. Image segmentation using a multilayer level-set approach. Computing and Visualization in Science, 12(6):267–285, April 2008.

[20] L. D. Cohen and R. Kimmel. Global minimum for active contour models: A minimal path approach. International Journal of Computer Vision, 24(1):57–78, 1997.

[21] M.G. Crandall, L.C. Evans, and R.F. Gariepy. Optimal lipschitz extensions and the infinity Laplacian. Calculus of Variations and Partial Differential Equations, 13(2):123–139, 2001.

[22] G. Dal Maso, J. M. Morel, and S. Solimini. A variational method in image segmentation: Existence and approximation results. Acta Mathematica, 168(1):89–151, 1992.

[23] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising. Applied and Computational Harmonic Analysis, 19(1):1–16, July 2005.

[24] W. Dong, X. Li, L. Zhang, and G. Shi. Sparsity-based image denoising via dictionary learning and structural clustering. In 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 457–464. IEEE, June 2011.

[25] I. Ekeland and R. Témam. Convex Analysis and Variational Problems. SIAM, 1999.

[26] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing, 15(12):3736–3745, December 2006.

[27] L. C. Evans and R. F. Gariepy. Measure Theory and Fine Properties of Functions. CRC Press, 1992.

[28] L. C. Evans and Y. Yu. Various properties of solutions of the Infinity-Laplacian equation. Communications in Partial Differential Equations, 30(9):1401–1428, September 2005.

[29] H. Gao, J-F. Cai, Z. Shen, and H. Zhao. Robust principal component analysis-based four-dimensional computed tomography. Physics in Medicine and Biology, 56:3181–3198, June 2011.

[30] H. Gao, H. Yu, S. Osher, and G. Wang. Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM). Inverse Problems, 27, 2011.

[31] J.B. Garnett, T.M. Le, Y. Meyer, and L.A. Vese. Image decompositions using bounded variation and generalized homogeneous besov spaces. Applied and Computational Harmonic Analysis, 23:25–56, 2007.

[32] G. Gilboa and S. Osher. Nonlocal linear image regularization and supervised segmentation. Multiscale Modeling & Simulation, 6:595, 2007.

[33] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. Multiscale Modeling & Simulation, 7:1005, 2009.

[34] J. Gilles. Noisy image decomposition: A new structure, texture and noise model based on local adaptivity. Journal of Mathematical Imaging and Vision, 28:285–295, August 2007.

[35] J. Gilles and Y. Meyer. Properties of BV-G structures +textures decomposition models. application to road detection in satellite images. IEEE Transactions on Image Processing, 19(11):2793–2800, November 2010.

[36] E. Giorgi, M. Carriero, and A. Leaci. Existence theorem for a minimum problem with free discontinuity set. Archive for Rational Mechanics and Analysis, 108(4):195–218, 1989.

[37] T. Goldstein and S. Osher. The split bregman method for L1-Regularized problems. SIAM Journal on Imaging Sciences, 2(2):323, 2009.

[38] H. Ji, C. Liu, Z. Shen, and Y. Xu. Robust video denoising using low rank matrix completion. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 1791–1798. IEEE, 2010.

[39] M. Jung, G. Chung, G. Sundaramoorthi, L. A. Vese, and A. L. Yuille. Sobolev gradients and joint variational image segmentation, denoising, and deblurring. Proceedings of SPIE, 7246(1):72460I–72460I–13, February 2009.

[40] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. International Journal of Computer Vision, 1(4):321–331, 1988.

[41] Y. Kim and L. Vese. Image recovery using functions of bounded variation and sobolev spaces of negative differentiability. Inverse Problems and Imaging, 3:43–68, February 2009.

[42] R. Kimmel and A. M. Bruckstein. Regularized Laplacian zero crossings as optimal edge integrators. International Journal of Computer Vision, 53:225–243, 2001.

[43] C. Lacoste, X. Descombes, and J. Zerubia. Unsupervised line network extraction in remote sensing using a polyline process. Pattern Recognition, 43(4):1631–1641, April 2010.

[44] C. J. Larsen, C. L. Richardson, and M. Sarkis. A level set method for the Mumford -Shah functional and fracture. Preprint serie A, Instituto Nacional de Matemática Pura e Aplicada, Brazilian Ministry for Science and Technology, 2008.

[45] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27(8):1265 –1278, aug. 2005.

[46] T.M. Le, Lieu L. H., and L.A. Vese. (ϕ, ϕ^*) image decomposition models and minimization algorithms. Journal of Mathematical Imaging and Vision, 33:135–148, 2009.

[47] T.M. Le and L.A. Vese. Image decomposition using total variation and div (BMO). Multiscale Modeling and Simulation, 4:390–423, 2005.

[48] W.-H. Liao, L. Vese, S.-C. Huang, M. Bergsneider, and S. Osher. Computational anatomy and implicit object representation: a level set approach. In Biomedical Image Registration, pages 40–49. Springer, 2003.

[49] L. H. Lieu and L. A. Vese. Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces. Applied Mathematics and Optimization, 58:167–193, May 2008.

[50] G. Lu and P. Wang. Inhomogeneous infinity laplace equation. Advances in Mathematics, 217(4):1838–1868, March 2008.

[51] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In 2009 IEEE 12th International Conference on Computer Vision, pages 2272–2279. IEEE, October 2009.

[52] J. Mairal, J. Bach, F.and Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res., 11:19–60, March 2010.

[53] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. IEEE Transactions on Image Processing, 17(1):53–69, January 2008.

[54] J. Melonakos, E. Pichon, S. Angenent, and A. Tannenbaum. Finsler active contours. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(3):412 –423, March 2008.

[55] Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures. American Mathematical Soc., 2001.

[56] J. M. Morel and S. Solimini. Segmentation of images by variational methods: a constructive approach. Revista Matematica de la Universidad Complutense de Madrid, 1:169–182, 1988.

[57] J. M. Morel and S. Solimini. Segmentation d'images par méthode variationnelle: une preuve constructive d'existence. Comptes rendus de l'Académie des sciences. Série 1, Mathématique, 308(15):465–470, 1989.

[58] 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, July 1989.

[59] J. W. Neuberger. Sobolev Gradients and Differential Equations. Springer, November 2009.

[60] C. Niemann, A.S. Bondarenko, C.G. Constantin, E.T. Everson, K.A. Flippo, S.A. Gaillard, R.P. Johnson, S.A. Letzring, D.S. Montgomery, L.A. Morton, D.B. Schaeffer, T. Shimada, and D. Winske. Collisionless shocks in a large magnetized laser-plasma plume. Plasma Science, IEEE Transactions on, 39(11):2406 –2407, nov. 2011.

[61] A. M. Oberman. A convergent difference scheme for the infinity laplacian: Construction of absolutely minimizing lipschitz extensions. Mathematics of Computation, 74(251):1217–1230., September 2004.

[62] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total Variation-Based image restoration. Multiscale Modeling & Simulation, 4(2):460, 2005.

[63] S. Osher and J. A Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. Journal of Computational Physics, 79(1):12–49, November 1988.

[64] S. Osher, A. Solé, and L. Vese. Image decomposition and restoration using total variation minimization and the H^1. Multiscale Modeling & Simulation, 1:349, 2003.

[65] R J. Renka. A simple explanation of the sobolev gradient method, 2006.

[66] W. B. Richardson. Sobolev gradient preconditioning for image-processing PDEs. Communications in Numerical Methods in Engineering, 24(6):493–504, December 2006.

[67] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1-4):259–268, November 1992.

[68] H. Schaeffer and S. Osher. A low patch-rank interpretation of texture. SIAM Journal on Imaging Sciences, 6(1):226–262, 2013.

[69] H. Schaeffer and L. Vese. Active contours with free endpoints. Journal of Mathematical Imaging and Vision, pages 1–17, 2013.

[70] J. Shen. Piecewise $H^{-1} - H^0 - H^1$ images and the Mumford-Shah-Sobolev model for segmented image decomposition. APPL. MATH. RES. EXP, 4:2005, 2005.

[71] P. Smereka. Spiral crystal growth. Physica D: Nonlinear Phenomena, 138(3–4):282–301, April 2000.

[72] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci. Sobolev active contours. International Journal of Computer Vision, 73(3):345–366, January 2007.

[73] A. Tsai, Jr. Yezzi, A., and A.S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. Image Processing, IEEE Transactions on, 10(8):1169 –1186, August 2001.

[74] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. International Journal of Computer Vision, 50(3):271–293, 2002.

[75] L. A. Vese and S. J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. Journal of Scientific Computing, 19:553—572, 2002.

[76] L. A. Vese and S. J. Osher. Image denoising and decomposition with total variation minimization and oscillatory functions. Journal of Mathematical Imaging and Vision, 20:7–18, January 2004.

[77] X. Zhang, M. Burger, X. Bresson, and S. Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM Journal on Imaging Sciences, 3(3):253–276, 2010.

[78] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin. Non-Parametric bayesian dictionary learning for sparse image representations 1. NIPS, 2009.