

# Bayesian Video Dejittering By BV Image Model

Jianhong Shen  
School of Mathematics  
University of Minnesota  
Minneapolis, MN 55455, USA \*

*Dedicated to all pioneering mathematicians in image and vision analysis,  
on whose shoulders we the younger generation are standing...*

## Abstract

Line jittering, or random horizontal displacement in video images, occurs when the synchronization signals are corrupted in video storage media, or by electromagnetic interference in wireless video transmission. The goal of intrinsic video dejittering is to recover the ideal video directly from the observed jittered and often noisy frames. The existing approaches in the literature are mostly based on local or semi-local filtering techniques and autoregressive image models, and complemented by various image processing tools. In this paper, based on the statistical rationale of Bayesian inference, we propose the first variational dejittering model based on the bounded variation (BV) image model, which is global, clean and self-contained, and intrinsically combines dejittering with denoising. The mathematical properties of the model are studied based on the direct method in Calculus of Variations. We design one effective algorithm and present its computational implementation based on techniques from numerical partial differential equations (PDE) and nonlinear optimizations.

**Keywords:** Video, line jittering, Bayesian, BV, variational, PDE.

## 1 Introduction

The best way for describing video jittering is to quote from the recent remarkable monograph by Anil Kokaram on motion picture restoration [16]:

“Video signals must contain synchronization information to allow the video display to properly locate lines and frames relative to each other in space and time. Noise in the video signal, or degradation of the storage medium on which the signal is stored (video tape) can cause the synchronization signals to be corrupted. This can cause the loss of ‘lock’ in video

---

\*Research is supported by the Program of Applied Mathematics of NSF under grant number DMS-0202565. Email: jhshen@math.umn.edu. Tel: (612) 625-3570. Fax: (612) 626-2017.

digitizing and playback apparatus. The loss of line synchronization pulses will prevent the video manipulation device from locating the actual start and end of each line thus yielding random line displacements (line jitter) in the observed video images.”

Figure 1 displays a typical jittered video frame, in which horizontal image lines are randomly shifted. In the current paper, we will not consider inter-frame correlation and jittering. Therefore the dejittering problem, i.e., to recover the original ideal image frame  $u$  from the observed jittered (and often noisy due to medium corruption) image frame  $u_0$ , is fundamentally a still image restoration problem.

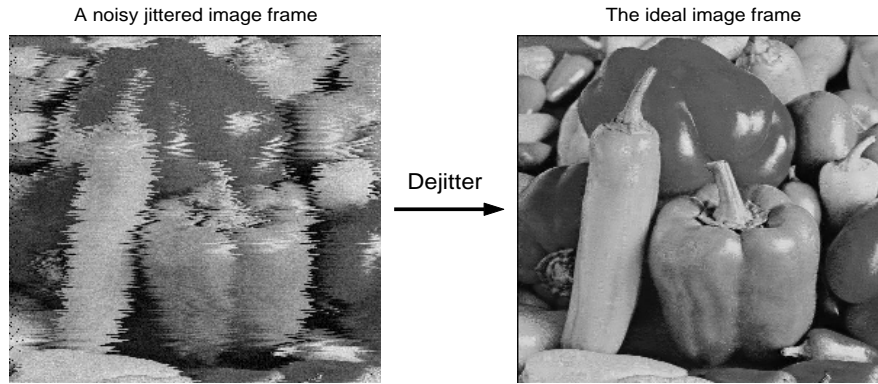


Figure 1: A noisy jittered image frame and the goal of dejittering.

For real analogue videos, it is possible to recover the line synchronization information by cleaning the non-picture parts of the video signals. This is the method of *Time Base Correctors* [16]. It demands the availability of signal information that has nothing to do with the video or image content.

The idea of *intrinsic* video dejittering [17, 18], on the other hand, is to restore the ideal image frames directly from the observed jittered data. That is, as in the classical denoising or deblurring problems, one attempts to recover the ideal images solely based on intrinsic image structures (i.e. image prior models).

Thus in comparison, the intrinsic approach is more flexible, and applicable in more general settings. For instance, due to electromagnetic interference in the environment (especially intentional interfering as in battle fields), wireless image signals can experience very similar jittering problems, completely or partially. It is impossible to reconstruct the random and dynamic process of environmental interfering, but it is still feasible to recover the original images based on their intrinsic structures.

In the pioneering works by Kokaram et al. [17, 18], intrinsic dejittering models are developed based on autoregressive image models, line registration, image interpolation, parameter estimation, and the filtering technique. The goal of the current paper is to address the dejittering problem with the help of Calculus of Variations and partial differential equations (PDEs), two novel modern tools in mathematical image and vision analysis (Miva).

Our main contributions are highlighted below.

- (a) Treating dejittering as an image restoration problem, we propose the first variational dejittering model in the literature. The rationale of this deterministic model is the general statistical framework of Bayesian inference.
- (b) The model is clean and self-contained, meaning that no other pre- or post-processing steps are needed. Furthermore, it intrinsically combines the two processes of denoising and dejittering.
- (c) We propose to apply the BV (bounded variation) image prior model for faithfully recovering the regularity of the jittered boundaries of image objects. The BV image model was first applied by Rudin, Osher, and Fatemi [22, 23] for image denoising and deblurring. The most beautiful attribute of the BV image model is that it takes care of image boundaries automatically, without the pain of separating them from the *interior* homogeneous regions, which greatly simplifies both the theory and computation.
- (d) By applying the direction method in Calculus of Variations, we attempt to reveal some important properties of our dejittering model, including uniqueness, existence, and convergence.
- (e) For the nonlinear and non-convex objective of our model, we design an iterative algorithm which alternately optimizes the image estimation and jittering estimation. This algorithm is then numerically implemented by techniques from computational PDEs, and nonlinear optimization.

The organization goes as follows. The statistical assumptions on the jittering and intensity noisy models are stated in Section 2. In Section 3, we derive our BV based dejittering model from the Bayesian rationale. In Section 4, we study the associated admissible conditions and some fundamental properties of the model. We explain in details the algorithm and its computational implementation in Section 5, which are followed by two typical numerical examples. Section 6 concludes the paper with a brief summary.

## 2 The Statistical Jittering Model

We begin with some necessary statistical assumptions or models for the line jittering process. For convenience, the image domain is assumed to be a horizontal stripe:  $\Omega = \mathbb{R} \times (0, H)$ , in which each point  $(x, y)$  is called a pixel.

Let  $s = s(y)$  be a Gaussian homogeneous white noise on  $y \in (0, H)$ . That is, there exists no correlation between the jitters of two distinct horizontal lines:

$$\mathbf{E}[s(y_1)s(y_2)] = \sigma_s^2 \delta_{y_1, y_2},$$

where  $\delta_{y_1, y_2} = 1$  if  $y_1 = y_2$  and 0 otherwise, and  $\sigma_s^2$  denotes the shared jittering variance. We shall always assume that  $s(y)$  has zero mean.

Let  $u(x, y), (x, y) \in \Omega$  denote the original ideal image to be displayed or transmitted. The horizontal line jittering process is modeled by  $u \rightarrow u_s$ :

$$u_s(x, y) = u(x + s(y), y), \quad (x, y) \in \Omega.$$

Thus generally  $u_s$  becomes a random field on  $\Omega$ . In practice, due to electromagnetic or medium noise, the jittered image  $u_s$  is further polluted to  $u_0$ :

$$u_0(x, y) = u_s(x, y) + n(x, y), \quad (x, y) \in \Omega.$$

Here we have adopted an additive noise model, and  $n$  denotes a Gaussian homogenous white noise with mean 0 and variance  $\sigma_n^2$ .

We shall further assume that the jittering  $s$  and intensity noise  $n$  are independent, since their physical causes are often uncorrelated in applications.

The dejittering problem can then be stated as follows. Suppose only one *single* observation  $u_0$  is made. Find a suitable way to restore the original image  $u$  (to certain commercially acceptable standards, say).

It naturally falls into the scope of Bayesian inference: decision making or feature (in our case  $u$ ) extraction based on observed data. Thus the spirit of our approach is tightly attached to the Bayesian restoration framework.

### 3 Bayesian Dejittering for BV Images

The goal of dejittering is to estimate both the original ideal image  $u$  and the particular jitter  $s$  involved in the given one single observation  $u_0$ . In the Bayesian framework, we are to maximize the posterior probability

$$p(u, s|u_0) = \frac{p(u_0|u, s)p(u, s)}{p(u_0)}.$$

Jittering is caused by the corruption of the synchronizing signatures, and is therefore independent of the image  $u$ , which leads to  $p(u, s) = p(u)p(s)$ . In addition, once the observation  $u_0$  is given,  $p(u_0)$  is simply a normalization constant and has no influence in terms of probability maximization. Taking either the logarithm likelihood function, or formally in terms of statistical mechanics, the Gibbs' ensemble energy  $E[\cdot] = -\frac{1}{\beta} \ln p(\cdot)$  ( $\beta$  denoting the reciprocal of the absolute temperature), we are to minimize the posterior "energy"

$$E[u, s|u_0] = E[u_0|u, s] + E[s] + E[u].$$

The equality holds up to an additive "grounding" energy level. Throughout this paper, the notation  $E[A|B]$  always denotes a functional of  $A$ , which depends on the given  $B$  as well. Both  $A$  and  $B$  can contain more than one objects.

The *data model*  $E[u_0|u, s]$  easily follows from the Gaussian intensity noise  $n$ :

$$u_0 = u_s + n = u(x + s(y), y) + n(x, y).$$

That is,

$$E[u_0|u, s] = \lim_{R \rightarrow \infty} \frac{1}{2\sigma_n^2 |\Omega_R|} \int_{\Omega_R} (u_0 - u_s)^2 dx dy, \quad (1)$$

where  $\Omega_R = (-R, R) \times (0, H)$ , and  $|\Omega_R| = 2RH$  denotes its Lebesgue measure.

We have two priors: the line jittering model  $E[s]$  and the image model  $E[u]$ . Since the former has been assumed Gaussian, we must have

$$E[s] = \frac{1}{2\sigma_s^2 H} \int_0^H s^2(y) dy. \quad (2)$$

Therefore, the key to our Bayesian dejittering approach is to adopt an appropriate image prior  $E[u]$ .

Identification of effective image priors is probably the most fundamental problem in the entire field of Miva. There are a number of valuable image priors in the literature, which can be coarsely classified into two categories: stochastic and deterministic.

Stochastic image priors are typically based on either the lattice model and Gibbs' fields in statistical mechanics (Geman and Geman [12]), or statistical learning via techniques like multiscale filtering and the maximum entropy principle (Zhu, Wu, and Mumford [27, 28]).

Deterministic image priors emphasize more the property of regularities, as contrast to the focus on various spectral or multiscale statistical features in stochastic priors.

Well known examples include (a) the Sobolev image model  $E_2[u] = \alpha \int_{\Omega} |\nabla u|^2 dx dy < \infty$  in the classical linear filtering theory; (b) the bounded variation (BV) [13] model  $E_{tv}[u] = \alpha \int_{\Omega} |Du| < \infty$ , first introduced to image restoration by Rudin, Osher, and Fatemi [22, 23]; and (c) the Mumford-Shah [21] object-edge free boundary model

$$E_{ms}[u, \Gamma] = E[u|\Gamma] + E[\Gamma] = \alpha \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx dy + \beta H^1(\Gamma) < \infty,$$

where  $\Gamma$  denotes the jump set or the collection of "edges,"  $H^1$  the one-dimensional Hausdorff measure, and  $\alpha$  and  $\beta$  two tunable weights. Recently, motivated by the image inpainting problem, we have also investigated high order geometric image models such as the elastica model (Chan, Kang, and Shen [4]) and the Mumford-Shah-Euler image model (Esedoglu and Shen [11]), where high order geometric information such as the mean curvature is also taken into account.

In the current paper, we choose to work with Rudin-Osher-Fatemi's BV image model

$$E_{tv}[u] = \alpha \int_{\Omega} |Du| < 0. \quad (3)$$

Notice that here  $|Du|$  denotes the Radon measure of a BV function [13]. For the more restricted Sobolev  $W^{1,1}$  images,  $E_{tv}[u]$  is simply the ordinary  $L^1$  norm of  $\nabla u$ . Numerous applications have demonstrated that the BV image model is well balanced in terms of faithfulness in approximating generic images (especially those mainly containing man-made objects), theoretical accessibility, and computational complexity [1, 2, 3, 5,

6, 7, 9, 14, 22, 23, 24]. Such advantages can also be witnessed in the rest of the paper. We refer to our most recent survey paper [8] for a concise overview on the role of the BV image model in Miva.

The combination of all the three models (1) (2) (3) leads to the complete Bayesian posterior “energy” to be minimized:

$$E[u, s|u_0] = \lim_{R \rightarrow \infty} \frac{\lambda_R}{2} \int_{\Omega_R} (u_0 - u_s)^2 dx dy + \frac{\mu}{2} \int_0^H s^2(y) dy + \alpha \int_{\Omega} |Du|, \quad (4)$$

where  $\lambda_R = 1/(\sigma_n^2 |\Omega_R|)$  and  $\mu = 1/(\sigma_s^2 H)$ . Notice that the total variation weight  $\alpha$  is the only tunable parameter. The impossibility of having a universally working  $\alpha$  is closely connected to the undefinability of a scale-invariant probability measure over “all” images (see Mumford and Gidas’ most recent work [20]).

The rest of the paper focuses on the properties and computation of this Bayesian de jittering model. In what follows, we first argue that model (4) has to be modified for allowing non-trivial solutions.

## 4 Properties of the Model

So far the de jittering model (4) has been purely inspired by the statistical jittering model and the Gibbs’ energy rationale. In this section, (a) we first rigorously define its admissible space; (b) we then argue that model (4) has to be modified for yielding meaningful solutions; and finally (c) we study some important properties of the modified model.

### 4.1 Admissible Conditions and Correcting Model (4)

From the squared integration term in (4) defining the jittering variance  $s$ , it is necessary that  $s = s(y) \in L^2(0, H)$ .

The BV prior naturally requires that  $u \in \text{BV}(\Omega)$ . By the Sobolev Embedding Theorem [13] (and its extension to the BV space),  $\text{BV}(\Omega)$  is embedded in  $L^2(\Omega)$ , implying that  $\int_{\Omega} u^2(x, y) dx dy < \infty$ .

For any  $s \in L^2(0, H)$ , define the jittering transform  $T_s : \Omega \rightarrow \Omega$  by

$$T_s : (x, y) \rightarrow (x + s(y), y). \quad (5)$$

Then  $T_s^{-1} = T_{-s}$ . We now show that  $T_s$  is a Lebesgue isomorphism. Let  $E \subset \Omega$  be any measurable set, and  $|E|$  its Lebesgue measure. Denote the characteristic function or indicator of  $E$  by  $1_E(x, y)$ . Then it is easy to see that

$$1_{T_s E}(x, y) = 1_E \circ T_{-s}(x, y).$$

By Fubini’s Theorem,

$$\begin{aligned} |T_s E| &= \int_{\Omega} 1_E \circ T_{-s}(x, y) dx dy = \int_0^H dy \int_{\mathbb{R}} 1_E(x - s(y), y) dx \\ &= \int_0^H dy \int_{\mathbb{R}} 1_E(x, y) dx = \int_{\Omega} 1_E(x, y) dx dy = |E|. \end{aligned}$$

The jittering transform  $T_s$  is therefore a Lebesgue isomorphism. In particular, for any  $u \in \text{BV}(\Omega) \subset L^2(\|\cdot\|, \Omega)$ ,

$$u_s = u \circ T_s \in L^2(\Omega) \quad \text{and} \quad \|u_s\| = \|u\|. \quad (6)$$

Finally, the data model in (4) has been formally motivated by the *Law of Large Numbers*:

$$\sigma_n^2 = \lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \int_{\Omega_R} n^2(x, y) dx dy.$$

Due to the averaging over the entire infinite domain, we now show that model (4) leads to an unexpected dead end. The problem is fixed in the next subsection.

**Theorem 1** *Let  $u_0$  be a given noisy jittered image in  $L^2_{\text{loc}}(\Omega)$ . Suppose that there exists at least one  $w(x, y) \in \text{BV}(\Omega)$ , such that*

$$\lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \int_{\Omega_R} n^2(x, y) dx dy = \sigma_n^2 \quad \text{exists,} \quad \text{with} \quad n = u_0 - w_s = u_0 - w \circ T_s.$$

Then

$$(u = 0, s = 0) = \text{argmin} E[u, s | u_0], \quad E \text{ as in model (4).}$$

*Proof.* The conclusion follows directly from

$$\lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \int_{\Omega_R} (u_0 - w_s)^2 dx dy = \lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \int_{\Omega_R} (u_0 - u_s)^2 dx dy,$$

for any  $u \in \text{BV}(\Omega)$ , which we are now to prove. Define

$$\langle f, g \rangle_R = \int_{\Omega_R} f(x, y) g(x, y) dx dy, \quad \|f\|_R^2 = \langle f, f \rangle_R.$$

Notice that

$$\|u_0 - u_s\|_R^2 = \|u_0 - w_s\|_R^2 - 2\langle u_0 - w_s, u_s - w_s \rangle_R + \|u_s - w_s\|_R^2. \quad (7)$$

Now that  $u, w \in \text{BV}(\Omega)$ , we must have  $u_s - w_s = (u - w)_s \in L^2(\|\cdot\|, \Omega)$ . Therefore,

$$\lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \|u_s - w_s\|_R^2 = \lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \|u_s - w_s\|^2 = 0.$$

For the cross term in (7), by the Schwartz inequality,

$$\begin{aligned} \frac{1}{|\Omega_R|} |\langle u_0 - w_s, u_s - w_s \rangle_R| &\leq \frac{1}{|\Omega_R|} \|u_0 - w_s\|_R \|u_s - w_s\|_R \\ &= \left[ \frac{1}{|\Omega_R|} \|n\|_R^2 \right]^{\frac{1}{2}} \left[ \frac{1}{|\Omega_R|} \|u_s - w_s\|_R^2 \right]^{\frac{1}{2}}. \end{aligned}$$

As  $R \rightarrow \infty$ , the first term converges to the standard deviation  $\sigma_n$  according to the assumption, but the second term vanishes as just shown. Therefore,

$$\lim_{R \rightarrow \infty} \frac{1}{|\Omega_R|} \langle u_0 - w_s, u_s - w_s \rangle_R = 0.$$

In combination, we are able to conclude that

$$\lim_{R \rightarrow \infty} \frac{1}{\|\Omega_R\|} \|u_0 - w_s\|_R^2 = \lim_{R \rightarrow \infty} \frac{1}{\|\Omega_R\|} \|u_0 - u_s\|_R^2.$$

This completes the proof.  $\square$

## 4.2 The Corrected Model and Its Properties

The problem is caused by averaging near infinity. In applications, images are given only on a bounded domain  $\Omega_R$ , and can be extended over the infinite stripe domain  $\Omega$  by zero-padding. Thus to diminish the over reaction of the infinity, we propose to modify model (4) to

$$E[u, s|u_0] = \frac{\lambda}{2} \int_{\Omega} (u_0 - u_s)^2 dx dy + \frac{\mu}{2} \int_0^H s^2(y) dy + \alpha \int_{\Omega} |Du|, \quad (8)$$

where  $\mu = 1/(\sigma_s^2 H)$  and  $\lambda = \beta/\sigma_n^2$ . Besides  $\alpha$ , a new tunable parameter  $\beta$  is thus introduced. It is now easy to collect all the admissible conditions:

- (a) the given jittered image  $u_0 \in L^2(\Omega)$ ; and
- (b)  $u \in \text{BV}(\Omega)$ , and  $s \in L^2(0, H)$ .

As in the statistical framework, we define the corresponding ‘‘conditional’’ energies for  $E[u, s|u_0]$ :

$$E[u|u_0, s] = \frac{\lambda}{2} \int_{\Omega} (u_0 - u_s)^2 dx dy + \alpha \int_{\Omega} |Du|, \quad \text{when } s \text{ is known;} \quad (9)$$

$$E[s|u_0, u] = \frac{\lambda}{2} \int_{\Omega} (u_0 - u_s)^2 dx dy + \frac{\mu}{2} \int_0^H s^2(y) dy, \quad \text{when } u \text{ is known.} \quad (10)$$

Since  $s \in L^2(0, H)$  is to at least model the rapid oscillations of random jittering, differential constraints (such as Sobolev norms or the BV norm) are generally inappropriate. (Of course, one is able to introduce them with small weights merely for the sake of Tikhonov regularization in inverse problems.) This makes the energy  $E[u, s|u_0]$  in (8) lack the necessary compactness properties for establishing a general existence theorem. Uniqueness is jeopardized as well because of the lack of convexity (especially in  $s$ ).

However, it is indeed possible to say more about the two ‘‘conditional’’ energies just defined, which will also be important for our algorithm in the next section.

**Theorem 2** *For any given jittering  $s$ , the minimizer in  $\text{BV}(\Omega)$  for  $E[u|u_0, s]$  is unique. Furthermore if there exists a minimizing sequence which is bounded in  $L^1(\Omega)$ , then the minimizer indeed exists.*

We shall explain right after the proof why we need the extra boundedness condition for the existence part, which is often unnecessary for image processing problems



on finite domains (see Chambolle and Lions [3] for example)

*Proof.* The jittering operator is linear:  $(u + v)_s = u_s + v_s$  for any  $u, v \in \text{BV}(\Omega)$ . Therefore it is straightforward to establish the strict convexity of  $E[u|u_0, s]$ , which leads to the uniqueness. We now prove the existence.

Let  $(u^n)_n \subset \text{BV}(\Omega)$  be a minimizing sequence of  $E[u|u_0, s]$ :

$$\lim_{n \rightarrow \infty} E[u^n|u_0, s] = \inf_{u \in \text{BV}(\Omega)} E[u|u_0, s] < \infty.$$

In addition, assume that the sequence is bounded in  $L^1(\Omega)$ . There must exist an upper bound  $M$ , so that  $e_n = E[u^n|u_0, s] \leq M, n = 1, 2, \dots$ , and

$$\int_{\Omega} |Du^n| \leq \frac{1}{\alpha} e_n \leq \frac{M}{\alpha}.$$

Therefore,  $(u^n)_n$  is bounded in  $\text{BV}(\Omega)$ . By the weak compactness property, there exist some  $w = w(x, y) \in \text{BV}(\Omega)$  and a subsequence, still denoted by  $(u^n)_n$  for convenience, such that

$$u^n \rightarrow w \text{ in } L^1(\Omega) \quad \text{and} \quad \int_{\Omega} |Dw| \leq \liminf_{n \rightarrow \infty} \int_{\Omega} |Du^n|.$$

The second half follows from the  $L^1$  lower semi-continuity in  $\text{BV}(\Omega)$ .

Possibly with another step of subsequence refinement, we can assume that

$$u^n(x, y) \rightarrow w(x, y), \quad a.e. \text{ on } \Omega.$$

Since the jittering transform  $T_s : \Omega \rightarrow \Omega$  is a Lebesgue isomorphism, we must have as well

$$u_s^n(x, y) \rightarrow w_s(x, y), \quad a.e. \text{ on } \Omega.$$

Application of Fatou's Lemma to the nonnegative sequence  $g_n = (u_0 - u_s^n)^2$  leads to

$$\int_{\Omega} (u_0 - w_s)^2 dx dy \leq \liminf_{n \rightarrow \infty} \int_{\Omega} (u_0 - u_s^n)^2 dx dy.$$

Eventually the above results enable us to conclude that

$$E[w|u_0, s] \leq \liminf_{n \rightarrow \infty} E[u^n|u_0, s] = \inf_{u \in \text{BV}(\Omega)} E[u|u_0, s].$$

Thus  $w$  has to be the (unique) minimizer of  $E[u|u_0, s]$  in  $\text{BV}(\Omega)$ . □

**Remark.** We now explain why the  $L^1$  boundedness condition seems to be necessary. It is true that we can bound the  $L^2$  norms of any minimizing sequence similar to what has been done for the total variation norms in the proof:

$$\|u^n\|^2 = \|u_s^n\|^2 \leq 2\|u_0\|^2 + \frac{4}{\lambda} e_n \leq 2\|u_0\|^2 + \frac{4M}{\lambda}, \quad n = 1, 2, \dots$$

If the domain  $\Omega$  is finite, the Schwartz inequality immediately leads to a common bound on the  $L^1$  norms:

$$\left[ \int_{\Omega} |u(x, y)| dx dy \right]^2 \leq \|u\|^2 |\Omega|.$$

But in our case,  $\Omega = \mathbb{R} \times (0, H)$  is an infinitely long stripe domain, for which the  $L^1$  norms can indeed be unbounded. For example, define

$$u^n(x, y) = \frac{1}{(1+x^2)^{\frac{n+1}{2n}}}, \quad n = 1, 2, \dots,$$

which are translation invariant along the  $y$  direction. Then it is easy to show that

- (a) the  $L^2$  norms are bounded:  $\|u^n\|^2 \leq \pi H$ ;
- (b) all  $u^n \in \text{BV}(\Omega)$  and their total variations are always  $2H$  exactly; but
- (c) their  $L^1$  norms diverge to  $\infty$  by the *Monotone Convergence Theorem*:

$$\int_{\Omega} |u^n(x, y)| dx dy = H \int_{\mathbb{R}} \frac{1}{(1+x^2)^{\frac{n+1}{2n}}} dx \rightarrow H \int_{\mathbb{R}} \frac{1}{(1+x^2)^{\frac{1}{2}}} dx = \infty.$$

Thus it cannot be a Cauchy sequence in  $L^1(\Omega)$ . By the *Monotone Convergence Theorem* again, it is not compact even in the topology of  $L^1_{\text{loc}}(\Omega)$ . This closes the remark.  $\square$

We can say much less about the other conditional energy  $E[s|u_0, u]$ , due to the lack of convexity (for uniqueness) and enough regularity (for existence).

One nice property of  $E[s|u_0, u]$  is that it is *separable*. That is, it can be written as

$$E[s|u_0, u] = \int_0^H e(s(y), y) dy, \quad \text{with}$$

$$e(s(y), y) = \frac{\mu}{2} s(y)^2 + \frac{\lambda}{2} \int_{\mathbb{R}} (u_0(x, y) - u(x + s(y), y))^2 dx.$$

Having  $y$  fixed, we define  $f_0(x) = u_0(x, y)$ ,  $f(x) = u(x, y)$ , and  $e(t) = e(t, y)$ . Then the minimization of  $E[s|u_0, u]$  is reduced to the minimization of every such single variable function  $e(t)$  associated to each  $y$ . Notice that

$$e(t) = \frac{\mu}{2} t^2 + \frac{\lambda}{2} \int_{\mathbb{R}} (f_0(x) - f(x + t))^2 dx, \quad (11)$$

which is well defined for *a.e.*  $y \in (0, H)$  following Fubini's Theorem. Notice that  $e(t)$  is generally non-convex, which leads to the uncertainty of uniqueness.

However, we are still able to establish the existence theorem and give an a priori bound on the minimizers.

**Theorem 3** *Suppose that  $f_0, f \in L^2(\mathbb{R})$  in (11). Then*

- (1)  $e(t)$  is a continuous function. In particular, the minimizer exists.

(2) Suppose  $t = s$  is one minimizer of  $e(t)$ . Then,

$$|s| \leq \sqrt{\lambda/\mu} (\|f_0\| + \|f\|).$$

*Proof.* For (a), take the continuity at  $t = 0$  for example. Notice that

$$| \|f_0(x) - f(x+t)\| - \|f_0(x) - f(x)\| | \leq \|f(x+t) - f(x)\|.$$

It is well known from Lebesgue integration theory that  $f(x+t)$  converges to  $f(x)$  in  $L^p(\mathbb{R})$  for any  $p \in [1, \infty)$  (excluding  $L^\infty$ ) as  $t \rightarrow 0$ . ( $p = 2$  in our case.) Therefore  $e(t)$  is indeed a continuous function. Since  $e(t) \rightarrow +\infty$  as  $t \rightarrow \pm\infty$ , the global minima must be attainable at some finite locations. (b) follows easily from  $\frac{\mu}{2}s^2 \leq e(s) \leq e(0)$ .  $\square$

## 5 The Algorithm and Numerical Results

In this section, we present an iterative algorithm to minimize  $E[u, s|u_0]$  in (8).

The plan is to alternately minimize the two conditional energies  $E[u|u_0, s]$  and  $E[s|u_0, u]$  in (9) and (10), just being studied above in theory. Starting from a pair of initial guess  $(u^0, s_0)$ , we generate  $(u^{n+1}, s_{n+1})$  from  $(u^n, s_n)$ ,  $n = 0, 1, \dots$ , by

$$u^{n+1} = \operatorname{argmin} E[u|u_0, s_n]; \quad \text{followed by} \quad (12)$$

$$s_{n+1} = \operatorname{argmin} E[s|u_0, u^{n+1}]. \quad (13)$$

The motivation is clear: the current best jittering estimation  $s_n$  leads to an improved estimation of the target image  $u^{n+1}$ , which in return contributes to the updating of the jittering itself.

Since the energy  $E[u, s|u_0]$  may have many local minima, convergence to the global minimum is generally unguaranteed. However, we do be able to show that the sequence  $(u^n, s_n)_n$  is consistently “down-hill.”

**Theorem 4** *Let  $(u^n, s_n)_n$  be the sequence derived from the above iterative algorithm. Then for  $n = 0, 1, \dots$ ,*

$$E[u^{n+1}, s_{n+1}|u_0] \leq E[u^n, s_n|u_0].$$

*Proof.* Following the energy definitions (1), (2), (9), (10), and the iteration formulae (12),(13), we have

$$\begin{aligned} E[u^{n+1}, s_{n+1}|u_0] &= E[u^{n+1}] + E[s_{n+1}|u_0, u^{n+1}] \\ &\leq E[u^{n+1}] + E[s_n|u_0, u^{n+1}] && \text{by (13)} \\ &= E[u^{n+1}] + E[u_0|s_n, u^{n+1}] + E[s_n] && \text{by (10)} \\ &= E[u^{n+1}|u_0, s_n] + E[s_n] && \text{by (9)} \\ &\leq E[u^n|u_0, s_n] + E[s_n] && \text{by (12)} \\ &= E[u^n, s_n|u_0]. \end{aligned}$$

This verifies the claim.  $\square$

The convergence of the sequence  $(u^n, s_n)_n$  is still unclear, although our numerical results always seem to confirm it. (That is, numerically, the sequence invariably converges to some pair  $(u, s)$ , which appears to be visually meaningful judged by human observers.) What we are able to establish is the following weak theorem on convergence.

**Theorem 5** *Suppose the jittered image  $u_0 \in L^2(\Omega)$  and  $u_0(\cdot, y)$  is a continuous function in  $x$  for a.e.  $y \in (0, H)$ . Let  $(u^n, s_n)$  be the sequence generated by the iterative scheme (12) and (13). Suppose that  $(s_n)_n$  a.e. converges to some  $s(y) \in L^2(0, H)$  and  $(\|u^n\|_{L^1})_n$  does not converge to  $\infty$ . Then there must exist a subsequence  $(n_k)_k$  and some  $u \in \text{BV}(\Omega)$  so that  $u^{n_k} \rightarrow u$  in  $L^1(\Omega)$  and*

$$E[u, s|u_0] \leq \lim_{n \rightarrow \infty} E[u^n, s_n|u_0].$$

*Proof.* By the preceding theorem, for any  $n$ ,

$$E[u^n|u_0, s_n] \leq E[u^n, s_n|u_0] \leq E[u^0, s_0|u_0].$$

Thus the bound on the total variations is immediate:

$$\int_{\Omega} |Du^n| \leq \frac{1}{\alpha} E[u^0, s_0|u_0], \quad n = 1, 2, \dots$$

Since  $(\|u^n\|_{L^1})_n$  does not converge to  $\infty$ , there must exist a subsequence  $(n'_k)_k$  so that  $(u^{n'_k})_k$  is a bounded sequence in  $L^1(\Omega)$ . Therefore,  $(u^{n'_k})_k$  is bounded in  $\text{BV}(\Omega)$ . By the properties of weak compactness and  $L^1$  lower semi-continuity, there must exist a refined subsequence  $(n_k)_k$  and some  $u \in \text{BV}(\Omega)$ , so that  $u^{n_k} \rightarrow u$  in  $L^1(\Omega)$  as  $k \rightarrow \infty$ , and

$$\int_{\Omega} |Du| \leq \liminf_{k \rightarrow \infty} \int_{\Omega} |Du^{n_k}|. \quad (14)$$

Possibly with an extra step of subsequence refinement, we can also assume that  $u^{n_k} \rightarrow u$ , a.e. on  $\Omega$ .

Secondly, since  $s_n(y) \rightarrow s(y)$  a.e., by Fatou's Lemma,

$$\int_0^H s^2(y) dy \leq \liminf_{k \rightarrow \infty} \int_0^H s_{n_k}^2(y) dy. \quad (15)$$

As discussed in Section 4.1, the jittering transform

$$T_s : (x, y) \rightarrow (x + s(y), y), \quad (x, y) \in \Omega$$

is a Lebesgue isomorphism, and  $T_s^{-1} = T_{-s}$ , which implies that

$$\int_{\Omega} (u_0 - u^n \circ T_{s_n})^2 dx dy = \int_{\Omega} (u_0 \circ T_{-s_n} - u^n)^2 dx dy.$$

Now that  $s_n \rightarrow s$  *a.e.* on  $(0, H)$ , we must have, as  $n \rightarrow \infty$ ,

$$T_{-s_n}(x, y) = (x - s_n(y), y) \rightarrow (x - s(y), y) = T_{-s}(x, y), \quad \textit{a.e.} \quad \text{on } \Omega.$$

Since it is assumed that the observed data  $u_0(\cdot, y)$  is a continuous function in  $x$  for *almost every*  $y \in (0, H)$ , we must have

$$u_0 \circ T_{-s_n} \rightarrow u_0 \circ T_{-s}, \quad \textit{a.e.} \quad \text{on } \Omega.$$

In combination with the *a.e.* convergence condition on  $(u^{n_k})_k$ , this implies that

$$u_0 \circ T_{-s_{n_k}} - u^{n_k} \rightarrow u_0 \circ T_{-s} - u, \quad \textit{a.e.} \quad \text{on } \Omega.$$

Therefore, Fatou's Lemma again leads to

$$\int_{\Omega} (u_0 \circ T_{-s} - u)^2 dx dy \leq \liminf_{k \rightarrow \infty} \int_{\Omega} (u_0 \circ T_{-s_{n_k}} - u^{n_k})^2 dx dy,$$

or equivalently,

$$\int_{\Omega} (u_0 - u_s)^2 dx dy \leq \liminf_{k \rightarrow \infty} \int_{\Omega} (u_0 - u_{s_{n_k}})^2 dx dy. \quad (16)$$

The combination of the three bounds (14), (15), and (16) eventually completes the proof:

$$E[u, s | u_0] \leq \liminf_{k \rightarrow \infty} E[u^{n_k}, s_{n_k} | u_0] = \lim_{n \rightarrow \infty} E[u^n, s_n | u_0].$$

The last equality follows from the monotonicity in the proceeding theorem.  $\square$

Notice in the proof how the Lebesgue isomorphism property of the jittering transform allows us to stay away from the touchy issue of the convergence of  $u_{s_n}^n$  to  $u_s$ . From the proof it is clear that the condition “ $u_0(\cdot, y)$  is continuous in  $x$ ” could be relaxed to “ $u_0(\cdot, y)$  is continuous for *a.e.*  $x \in \mathbb{R}$ ”.

Next we discuss the computational strategies for the two “conditional” optimization problems (9) and (10).

## 5.1 Algorithm for Minimizing $E[u | u_0, s]$

Recall that given  $u_0$  and the current available jittering estimation  $s$ , the conditional energy for  $u$  is given by

$$E[u | u_0, s] = \frac{\lambda}{2} \int_{\Omega} (u_0 - u_s)^2 dx dy + \alpha \int_{\Omega} |Du|,$$

which is similar to the classical TV denoising model of Rudin, Osher, and Fatemi [23], except that now the jittering  $u_s = u \circ T_s$  is involved.

Formally, or assuming that  $u \in W^{1,1}(\Omega)$  and  $\int_{\Omega} |Du| = \int_{\Omega} |\nabla u| dx dy$ , the first order variation on total variation leads to the well known curvature derivative [23] in the distributional sense:

$$-\alpha \nabla \cdot \left[ \frac{\nabla u}{|\nabla u|} \right].$$

Secondly, the variation on the data model term gives

$$\lambda T_s^*(T_s u - u_0),$$

where  $T_s u = u \circ T_s$ , and  $T_s^*$  denotes the adjoint of  $T_s$ . Notice that  $T_s^* = T_{-s} = T_s^{-1}$ . Eventually we obtain the (formal) Euler-Lagrange equation

$$\frac{\partial E}{\partial u}[u|u_0, s] = \lambda(u - u_{0,-s}) - \alpha \nabla \cdot \left[ \frac{\nabla u}{|\nabla u|} \right] = 0, \quad (17)$$

where  $u_{0,-s} = (u_0)_{-s} = u_0 \circ T_{-s}$ , with the natural Neumann boundary condition  $\frac{\partial u}{\partial \vec{n}} = 0$ .

Define  $v_0 = u_{0,-s}$  for a given jitter  $s$ . To our great surprise, the Euler-Lagrange equation here is identical to that of Rudin-Osher-Fatemi's denoising model when applied to  $v_0$ .

Consequently, we are entitled to applying all computational techniques from the rich literature of TV denoising. First, instead of solving the singular equation (17), we are to solve its viscosity approximation:

$$\lambda(u - u_{0,-s}) - \alpha \nabla \cdot \left[ \frac{\nabla u}{|\nabla u|_\epsilon} \right] = 0, \quad |a|_\epsilon = \sqrt{a^2 + \epsilon^2}, \quad (18)$$

for some small regularizing positive parameter  $\epsilon$ . Furthermore, this nonlinear elliptic equation is solved by *lagged diffusivity iterations* [10], which is a natural linearization technique. Let  $u^n$  be the current estimation for (18). Then  $u^n$  is updated to  $u^{n+1}$  by

$$\lambda(u^{n+1} - u_{0,-s}) - \alpha \nabla \cdot \left[ \frac{\nabla u^{n+1}}{|\nabla u^{n+1}|_\epsilon} \right] = 0,$$

with the associated Neumann boundary conditions. To  $u^{n+1}$ , the original ‘‘diffusivity coefficient’’  $1/|\nabla u^{n+1}|_\epsilon$  is replaced by that from the previous step  $1/|\nabla u^n|_\epsilon$ . Convergence of the algorithm is well studied in [3, 10].

## 5.2 Algorithm for Minimizing $E[s|u_0, u]$

Given the current best image estimation  $u$ , the ‘‘conditional’’ energy for the jittering  $s$  is given by

$$E[s|u_0, u] = \frac{\lambda}{2} \int_{\Omega} (u_0 - u_s)^2 dx dy + \frac{\mu}{2} \int_0^H s^2(y) dy. \quad (19)$$

As in the last part of Section 4, for almost every (in the Lebesgue sense) given  $y \in (0, H)$ , define  $f_0(x) = u_0(x, y)$  and  $f(x) = u(x, y)$ . Then the minimization of the functional  $E[s|u_0, u]$  is reducible to 1-D energy functions in the form of

$$e(s) = \frac{\mu}{2} s^2 + \frac{\lambda}{2} \int_{\mathbb{R}} (f_0(x) - f(x + s))^2 dx. \quad (20)$$

It is a nonlinear function well defined for any given  $f_0, f \in L^2(R)$ . From the viscosity approximation, we can assume that  $u$  belongs to the Sobolev space  $H^1(\Omega)$  [13]. By Fubini's Theorem, for almost every  $y \in (0, H)$ ,  $f(x) = u(x, y) \in H^1(R)$ .

The optimal line jittering  $s$  must satisfy

$$e'(s) = \mu s - \lambda \int_{\mathbb{R}} (f_0(x) - f(x+s)) f'(x+s) dx = 0.$$

Notice that the integration is indeed well defined since  $f_0, f(x+s), f'(x+s)$  are all square integrable. Let  $\langle f, g \rangle$  denote the inner product in the Hilbert space  $L^2(R)$ . Then we can rewrite it by

$$e'(s) = \mu s - \lambda \langle f_0 - f(x+s), f'(x+s) \rangle. \quad (21)$$

If furthermore, we assume that the current estimation  $u \in H^2(\Omega)$ , then Fubini's Theorem again implies that  $f'' \in L^2(R)$ . We can then take the second order derivative following (21):

$$e''(s) = \mu + \lambda \langle f'(x+s), f'(x+s) \rangle - \lambda \langle f_0 - f(x+s), f''(x+s) \rangle. \quad (22)$$

Assume either the Neumann condition or vanishing condition at  $\pm\infty$  :  $f f'(\pm\infty) = 0$ . Then integration by parts gives

$$\langle -f(x+s), f''(x+s) \rangle = \langle f'(x+s), f'(x+s) \rangle,$$

and eventually

$$e''(s) = \mu + \lambda \langle f_0, -f''(x+s) \rangle. \quad (23)$$

Our algorithm for minimizing  $e(s)$  is then based on the Newton-Raphson method. Starting from an initial guess  $s_0$ , we update  $s_n$  to  $s_{n+1}$  by

$$s_{n+1} = s_n - \frac{e'(s_n)}{e''(s_n)} = \frac{s_n \langle f_0, -f''(x+s_n) \rangle + \langle f_0 - f(x+s_n), f'(x+s_n) \rangle}{\langle f_0, -f''(x+s_n) \rangle + (\mu/\lambda)}. \quad (24)$$

Notice that  $\lambda \propto 1/\sigma_n^2$  (see Section 2). In the absence of intensity noise (corresponding to  $\sigma_n = 0$ ), the ratio  $\mu/\lambda = 0$ . Then we have a much simpler formula:

$$s_{n+1} = s_n + \frac{\langle f_0 - f(x+s_n), f'(x+s_n) \rangle}{\langle f_0, -f''(x+s_n) \rangle}. \quad (25)$$

The feasibility of the algorithm relies on how robust the denominator  $e''(s)$  stays away from 0 (i.e., pure convexity or concavity), at least when  $s$  is close to the optimal jittering. We now argue heuristically that indeed  $e''(s)$  is reasonably well behaved, which has been observed from our numerical implementation as well. Suppose that the observed jittered 1-D image  $f_0(x)$  has been generated from an image  $g(x) \in H^1(R)$  with a jittering  $t$ , i.e.  $f_0(x) = g(x+t) + n(x)$ , where  $n$  denotes the Gaussian intensity white noise. Since  $n$  is independent of both  $g$  and  $f$ , or has rapid oscillatory behavior [19], we have

$$\langle f_0(x), -f''(x+s) \rangle = \langle g(x+t), -f''(x+s) \rangle = \langle g'(x+t), f'(x+s) \rangle,$$

where the last equality follows from integration by parts and the vanishing conditions at  $\pm\infty$ . Therefore, as the estimation pair  $(f, s)$  gets close to the genuine pair  $(g, t)$ ,  $e''(s)$  is in the order of

$$\mu + \lambda \langle g'(x+t), g'(x+t) \rangle = \mu + \lambda \|g'(x)\|^2,$$

which certainly robustly stays away from the zero. As a byproduct, we also see that the quantity  $\|g'(x)\|$  functions like an information measure: larger values mean richer variations in the image and more clues for robustly recovering the jittering.

### 5.3 Numerical Simulation and Results

We now briefly discuss some issues in the implementation of the above algorithms.

#### 5.3.1 Neumann boundary jittering model

In numerical simulation and real applications, images are given on a finite square domain  $\Omega_R = (-R, R) \times (0, H)$ . We therefore need a boundary jittering model. In our simulation, we adopt what we have called the *Neumann boundary jittering model*. For any  $s(y)$  and  $(x, y) \in \Omega_R$ : if  $|x + s(y)| \leq R$ , we define  $u_s(x, y) = u(x + s(y), y)$ ; otherwise, suppose  $\pm(x + s(y)) > R$ , then we define  $u_s(x, y) = u(\pm R, y)$  in the sense of traces for BV functions [13]. The intensity noise model remains untouched:

$$u_0(x, y) = u_s(x, y) + n(x, y), \quad \text{with Gaussian white noise } n(x, y). \quad (26)$$

#### 5.3.2 Parameter tuning

On the fixed finite image domain  $\Omega_R$ , the data model in both (4) and (8) bears the exact form following (26):

$$E[u_0|u, s] = \frac{\lambda_R}{2} \int_{\Omega_R} (u_0 - u_s)^2 dx dy,$$

with  $\lambda_R = 1/(\sigma_n^2 |\Omega_R|)$ . Therefore in terms of numerical simulation, the dejittering model becomes

$$E[u, s|u_0] = \frac{\lambda_R}{2} \int_{\Omega} (u_0 - u_s)^2 dx dy + \frac{\mu}{2} \int_0^H s(y)^2 dy + \alpha \int_{\Omega} |Du|.$$

It allows only one tunable parameter  $\alpha$ , since both  $\lambda_R$  and  $\mu$  are completely determined by the noise model ( $\sigma_n^2$ ) and the jittering model ( $\sigma_s^2$ ):

$$\lambda = \frac{1}{2\sigma_n^2 RH}, \quad \mu = \frac{1}{\sigma_s^2 H}.$$

Numerically the variances are obtained from any statistical estimators.



### 5.3.3 Random jittering generation

In digital implementation or real television applications, the image domain  $\Omega_R$  becomes a matrix of pixel dots. Therefore, practically the jittering  $s$  must be an integer. We could simulate  $s$  by the continuous Gaussian noise, followed by a quantization step. In this paper, however, we directly take the binomial noise  $B(p, N)$  with an even integer  $N = 2n$ :

$$\text{Prob}(s = k) = \binom{N}{n+k} p^{n+k} (1-p)^{n-k}, \quad k = -n, -n+1, \dots, n-1, n.$$

From probability theory, we have

$$\mathbf{E}(s) = N(1-p) - n = n(1-2p), \quad \sigma_s^2 = Np(1-p).$$

Since it is assumed in Section 2 that the jittering has zero mean, we must have  $p = 1/2$  and  $\sigma_s^2 = N/4$ . The *Central Limit Theorem* confirms that Gaussian is still a good approximation. In our simulations,  $N$  is in the order of 100, which brings the standard deviation  $\sigma_s$  close to 5 pixels. Jittering with a serval-pixel magnitude is already severe for real video tapes or television signals (see [16] for example).

### 5.3.4 Simulation examples

Finally, we present two numerical results derived from our model and its algorithm detailed above.

In Figure 2, from left to right are: the initial noisy jittered image  $u_0$ , the first image estimation  $u^1 = \text{argmin} E[u|u_0, s_0]$  with zero jittering  $s_0(y) \equiv 0$ , and the final dejittered output.

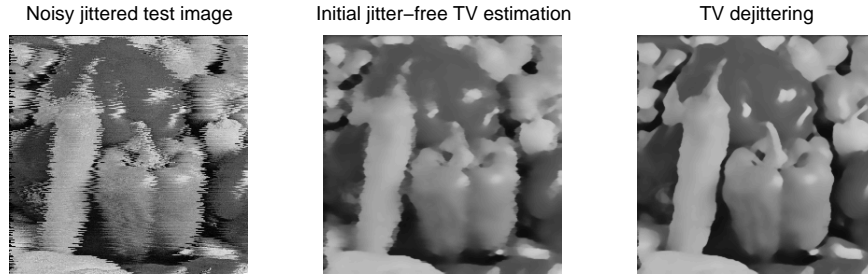


Figure 2: TV based Bayesian dejittering.

Ideally, we would expect that the standard deviation  $\text{std}(s_n)$  for each intermediate jittering estimation  $s_n$  is exactly  $\sigma_s$ . But computationally there is always some deviation, possibly due to the random number generator being used, the limited number of samples in digital implementation, and the Newton-Raphson algorithm itself. Therefore, we could enforce  $\text{std}(s_n) = \sigma_s$  by introducing an extra normalization step. The quality of the output indeed improves (see Figure 3).

In Figure 3, from left to right are: the initial noisy jittered image  $u_0$ , the image estimation after the fifth round  $u^6 = \operatorname{argmin} E[u|u_0, s_5]$ , and the similar image estimation based on the normalized jittering  $\hat{s}_5 = s_5/\operatorname{std}(s_5) \times \sigma_s$ . Here  $\operatorname{std}$  denotes the statistical standard deviation. If one compares the dejittered edges, the normalization step clearly improves their qualities.

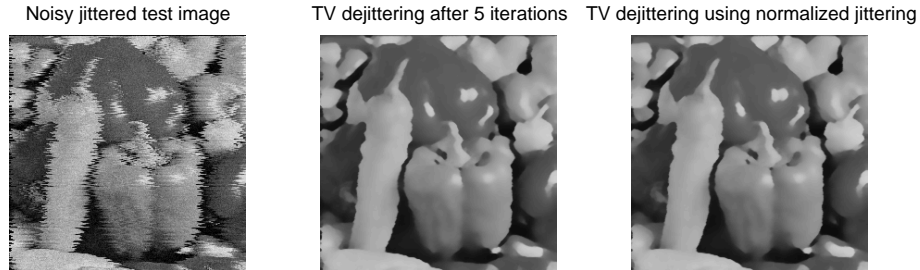


Figure 3: TV based Bayesian dejittering with jittering normalization.

## 6 Conclusion

In this paper, based on the Bayesian rationale, we have proposed a novel variational model for video dejittering. The image model of bounded variations that we have applied, first introduced into image processing by Rudin, Osher, and Fatemi [22, 23], is a powerful tool for restoring the regularity of the randomly jittered image objects.

We have studied the mathematical properties of the model based on the direct method in Calculus of Variations, the theory of functions with bounded variations, and various convergence tools from analysis.

For the nonlinear and non-convex energy functional, we have designed an algorithm based on alternately optimizing the image and jittering estimations. The algorithm is then numerically implemented by solving nonlinear partial differential equations (for image estimations) and by Newton-Raphson iterations. Typical numerical results are demonstrated.

This work again demonstrates the power of a good image model in image analysis. We expect that if the BV image model is replaced by the Mumford-Shah [21] image model, many results should remain similar. (For example, such exchange has been very successful in image inpainting [11].)

If the random jitterings are correlated among different video frames, we expect that dynamic tools such as the Kalman filter [15, 25, 26] can play an important role in the process of modeling and computation.

## Acknowledgments

The author would like to thank Gilbert Strang, Tony Chan, Stan osher, and David Mumford for their invaluable teaching and inspirations.

## References

- [1] F. Andreu, V. Caselles, J. I. Diaz, and J. M. Mazon. Some qualitative properties for the total variation flow. *J. Func. Anal.*, 188:516–547, 2002.
- [2] G. Aubert and L. Vese. A variational method in image recovery. *SIAM J. Numer. Anal.*, 34:1948–1979, 1997.
- [3] A. Chambolle and P. L. Lions. Image recovery via Total Variational minimization and related problems. *Numer. Math.*, 76:167–188, 1997.
- [4] T. F. Chan, S.-H. Kang, and J. Shen. Euler’s elastica and curvature based inpaintings. *SIAM J. Appl. Math.*, 63(2):564–592, 2002.
- [5] T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, 2001.
- [6] T. F. Chan and J. Shen. Variational restoration of non-flat image features: models and algorithms. *SIAM J. Appl. Math.*, 61(4):1338–1361, 2000.
- [7] T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2001.
- [8] T. F. Chan and J. Shen. On the role of the BV image model in image restoration. *Amer. Math. Soc. Contemporary Mathematics*, volume on Computational PDEs and Image Processing, accepted, 2002.
- [9] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, 6:298–311, 1997.
- [10] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.
- [11] S. Esedoglu and J. Shen. Digital inpainting based on the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13:353–370, 2002.
- [12] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.
- [13] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser, Boston, 1984.

- [14] K. Ito and K. Kunisch. An active set strategy based on the augmented Lagrangian formulation for image restoration. *Math. Model. Numer. Anal.*, 33:1–21, 1999.
- [15] R. E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.*, 82:34–45, 1960.
- [16] A. Kokaram. *Motion Picture Restoration*. Springer, London, 1998.
- [17] A. Kokaram and P. Rayner. An algorithm for line registration of TV images based on a 2-D AR model. *Signal Processing VI, Theories and Applications*, pages 1283–1286, 1992.
- [18] A. Kokaram, P. M. B. Roosmalen, P. Rayner, and J. Biemond. Line registration of jittered video. *IEEE Int’l Conference on Acoustics, Speech, and Signal Processing*, pages 2553–2556, 1997.
- [19] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. AMS, Providence, 2001.
- [20] D. Mumford and B. Gidas. Stochastic models for generic images. *Quarterly of Applied Mathematics*, 59:85–111, 2001.
- [21] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, 42:577–685, 1989.
- [22] L. Rudin and S. Osher. Total variation based image restoration with free local constraints. *Proc. 1st IEEE ICIP*, 1:31–35, 1994.
- [23] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [24] J. Shen. On the foundations of vision modeling I. Weber’s law and Weberized TV restoration. *Physica D*, in press, 2002.
- [25] G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, MA, 1993.
- [26] G. Strang. *Wavelet Analysis: Twenty Years’ Developments*, volume 1 of *Series in Analysis*, chapter “Block Tridiagonal Matrices and the Kalman Filter”, pages 266–280. World Scientific, New Jersey, 2002.
- [27] S. C. Zhu and D. Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(11):1236–1250, 1997.
- [28] S. C. Zhu, Y. N. Wu, and D. Mumford. Minimax entropy principle and its applications to texture modeling. *Neural Computation*, 9:1627–1660, 1997.