

# Multiscale Properties of Weighted Total Variation Flow with Applications to Denoising and Registration

Prashant Athavale\*, Robert Xu, Perry Radau, Graham Wright, and Adrian Nachman

**Abstract**—In this paper we investigate the multiscale nature of weighted total variation (TV) flow. The resulting understanding leads to novel fast denoising and registration algorithms. We show that the weighted TV flow can be derived from a hierarchical decomposition of a given image. We show that the edge-preserving property of the multiscale representation of an input image obtained with the TV flow can be enhanced and localized by appropriate choice of weights. We use this in developing an efficient and edge-preserving denoising algorithm with control on speed and localization properties. We examine analytical properties of the weighted TV flow that give precise information on the rate of decrease of the noise and of the energy of the image. A further contribution of the paper is to use the images obtained at different scales for robust multiscale registration. We demonstrate that on noisy cardiac MRI images the use of weighted TV flow as a preprocessing step in multiscale registration algorithm shows improved performance, compared to other methods such as bilateral or Gaussian filtering.

**Index Terms**—Magnetic resonance imaging (MRI), total variation flow, fast denoising, multiscale registration, normalized gradient field, mutual information.

## I. INTRODUCTION

Image denoising is an important problem, particularly to achieve high-resolution, 3D accelerated MRI acquisitions (see [2], [54], [23]). Denoising via unconstrained optimization, introduced in [39], can be realized as decomposition [31] of a given image into a smooth part that belongs to the space of bounded variation ( $BV$ ) and an  $L^2$  residual. The goal of this paper is to propose a novel multiscale representation based on hierarchical ( $BV, L^2$ ) decomposition. Our approach is motivated by successive application of the ( $BV, L^2$ ) decomposition to the  $BV$  parts obtained at increasing scale. We show that this approach is closely related to a well-studied problem of

*Asterisk indicates corresponding author.*

P. Athavale is with the Department of Mathematics, University of Toronto, and Fields Institute, Toronto, ON, M5S 1A1, Canada, email: prashantva@gmail.com.

R. Xu is with the Department of Medical Biophysics, University of Toronto, Toronto, ON, M5S 1A1, Canada, and Imaging Research, Sunnybrook Research Institute, Toronto, ON, M4N 3M5, Canada, email: robert.xu@sri.utoronto.ca.

P. Radau is with Imaging Research, Sunnybrook Research Institute, Toronto, ON, M4N 3M5, Canada, perry.radau@gmail.com.

A. Nachman is with the Department of Mathematics, the Department of Electrical Engineering, and the Institute of Biomaterials & Biomedical Engineering, University of Toronto, Toronto, ON, M5S 1A1, Canada, email: nachman@math.toronto.edu.

G. Wright is with the Department of Medical Biophysics, and the Institute of Biomaterials & Biomedical Engineering, University of Toronto, ON, M5S 1A1, Canada, email: graham.wright@sri.utoronto.ca.

total variation (TV) flow [4]. In section II we present the mathematical motivation behind the multiscale framework of the weighted TV and its properties. We demonstrate denoising applications in section III and registration applications (cf. [8], [53]) in section IV.

### A. Scale-Space Structure of Images with PDE Based Methods

One of the earliest PDE-based methods for denoising [47] and multiscale representation [29] of a given image  $f : \Omega \rightarrow \mathbb{R}$  for a bounded region  $\Omega$  in a plane is the heat equation,

$$\frac{\partial u}{\partial t} = \Delta u; \quad u \equiv u(x, t) : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \quad \frac{\partial u}{\partial \nu} \Big|_{\Gamma} = 0,$$

where  $u(\cdot, 0) := f$  and  $\frac{\partial u}{\partial \nu} \Big|_{\Gamma}$  denotes the outward normal to the boundary  $\Gamma$  of the region  $\Omega$ . This yields a family of images,  $\{u(\cdot, t) : \Omega \rightarrow \mathbb{R}\}_{t \geq 0}$ , which can be viewed as smoothed versions of  $f$ . In this linear framework, smoothing is implemented by a convolution with the two-dimensional Gaussian kernel,  $G_{\sigma}(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|x|^2}{2\sigma^2}\right)$ , with standard deviation  $\sigma = \sqrt{2t}$ . Hence, details with a scale smaller than  $\sqrt{2t}$  are smoothed out. Here,  $\lambda(t) := \sqrt{2t}$  acts as a *scaling function*. We can say that  $\{u(\cdot, t)\}_{t \geq 0}$  is a multiscale representation of  $f$ , as  $u(\cdot, t)$  diffuses from the small scales in  $f$  into increasingly larger scales.

Image denoising by the heat equation is based on isotropic diffusion, and consequently blurs all edges, which often contain useful information about the image. This drawback was addressed by the Perona-Malik (PM) model [36], which is based on *nonlinear diffusion*

$$\frac{\partial u}{\partial t} = \operatorname{div}(\alpha(|\nabla u|)\nabla u); \quad u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \quad \frac{\partial u}{\partial \nu} \Big|_{\Gamma} = 0, \quad (1)$$

with an initial condition  $u(\cdot, 0) := f$ . Here, the diffusion controlling function,  $\alpha$  is decreasing and vanishing at infinity, so that the amount of diffusion decreases as the gradient  $|\nabla u|$  increases. Thus,  $\alpha$  is responsible for the anisotropic nature of the PM model. The family of PM models are not well-posed [13]. They also pose a problem for noisy images, since noise produces high gradients, which can be confused with relevant edges. These shortcomings were removed by Catté et al. [13] by replacing  $\alpha(|\nabla u|)$  with  $\alpha(|G_{\sigma} * \nabla u|)$ , where  $G_{\sigma} * \nabla u$  denotes convolution of the two-dimensional Gaussian kernel  $G_{\sigma}$ .

The Perona-Malik equation (1) is closely related to the bilateral filtering method [48]. In fact, it has been shown to

be asymptotically equivalent to bilateral filtering [44]. The idea of bilateral filtering is to combine the low-pass filtering with range filtering. Given image  $f$ , the bilateral filtering is described as follows

$$h(x) = k^{-1}(x) \int_{-\infty}^{\infty} f(\zeta) c(\zeta, x) s(f(\zeta), f(x)) d\zeta$$

where  $c(x, \zeta)$  measures the geometric closeness between the neighborhood center  $x$  and a nearby point  $\zeta$ , whereas  $s(f(\zeta), f(x))$  measures the photometric similarity between the pixels at the neighborhood center  $x$  and a nearby point  $\zeta$ , and  $k^{-1}(x)$  is a normalization factor. For the Gaussian filtering case, both  $c$  and  $s$  are Gaussian functions of the Euclidean distance between their arguments. Bilateral filtering is known to preserve edges and produces a 2-dimensional multiscale representation of the function  $f$ , depending on the parameters  $\sigma_d$  and  $\sigma_r$ .

### B. Scale-Space Structure of Images with Variational Methods

Rudin, Osher and Fatemi introduced [39] a variational framework for denoising an image  $f$  which can be formulated as the following  $(BV, L^2)$  decomposition [31], [46] of an image  $f$  into a regular part  $u_\lambda$  and a noisy residual  $v_\lambda$ :

$$f = u_\lambda + v_\lambda, \quad [u_\lambda, v_\lambda] := \operatorname{arginf}_{f=u+v} \{ |u|_{TV(\Omega)} + \lambda \|v\|_{L^2(\Omega)}^2 \}, \quad (2)$$

where  $|u|_{TV(\Omega)}$  is the total variation or the  $BV$  seminorm of  $u$  (see [20]), and  $\lambda$  determines the relative weighing of the regularization and the fidelity term. This minimization problem has a unique minimizer,  $u_\lambda \in BV(\Omega)$  (see [1], [14], [49]). The solution to the minimization problem (2) depends on the scale parameter  $\lambda$ . More precisely  $\|v\|_* = \frac{1}{2\lambda}$ , where the star-norm,  $\|\cdot\|_*$ , is defined as the dual of the  $BV$  seminorm with respect to the  $L^2$  inner product [31]. Thus, the solution  $u$  of (2) has different scale for different  $\lambda$ . The best value of the parameter  $\lambda$  is difficult to determine a priori. Instead, in a seminal paper [46], the authors proposed to work with multiscale representations of images. While the emphasis in [46] was to progressively extract details of the image, we treat the multiscale approach as progressive removal of small-scale structures. Furthermore, we extend the results of [46] to continuous multiscale decompositions. We also (in parallel with [7] and [45]) allow for inhomogeneous weights in the regularization functional. This will be helpful for both denoising as well as registration methods. On one hand, since noise tends to consist of small-scale structures, removal of such details can yield a denoising algorithm; on the other hand, removal of small features is helpful in obtaining a good initial registration transformation, which can be progressively improved by working at different scales. We were able to obtain fast denoising and registration algorithms by considering the continuous limit of multiscale representations. For the inverse multiscale representation this limit was first studied in [45].

## II. RELATIONSHIP OF MULTISCALE DECOMPOSITION AND THE TOTAL VARIATION (TV) FLOW

It is very helpful to consider more general decompositions than (2), with the  $BV$  seminorm replaced by a convex func-

tional  $J(u)$  which will allow the flexibility of inhomogeneous weighing:

$$f = u_\lambda + v_\lambda, \quad [u_\lambda, v_\lambda] := \operatorname{arginf}_{f=u+v} \{ J(u) + \lambda \|v\|_{L^2}^2 \}. \quad (3)$$

Motivated by [46] we start by decomposing the original  $f$  using the parameter  $\lambda_1$  to obtain the initial decomposition,  $f = u_{\lambda_1} + v_{\lambda_1}$ . However, we choose the initial parameter  $\lambda_1$  large enough so that  $u_{\lambda_1}$  preserves most of the features of  $f$ . Using another scaling parameter  $\lambda_2 < \lambda_1$  we obtain the decomposition of  $u_{\lambda_1}$  as  $u_{\lambda_1} = u_{\lambda_2} + v_{\lambda_2}$ . If we continue this process iteratively, each time decomposing the  $u_{\lambda_i}$  part with  $\lambda_i < \lambda_{i-1}$ , we obtain the following nonlinear multiscale decomposition

$$\begin{aligned} f &= u_{\lambda_1} + v_{\lambda_1} \\ &= u_{\lambda_2} + v_{\lambda_2} + v_{\lambda_1} \\ &= \dots \\ &= u_{\lambda_N} + \sum_{i=1}^N v_{\lambda_i}. \end{aligned} \quad (4)$$

The idea is to terminate the iteration at the  $N^{\text{th}}$  step, when the cumulative residual,  $\sum_{i=1}^N v_{\lambda_i}$ , reaches a desired level. We next derive the continuous limit of the above multiscale decomposition. The Euler-Lagrange differential equation for (3) is

$$0 \in \partial J(u_\lambda) + 2\lambda(u_\lambda - f). \quad (5)$$

For a lower semicontinuous, proper convex functional  $J$  the subgradient  $\partial J(u) = A(u)$ , is a maximal monotone operator [11], and we can rewrite (5)

$$f = u_\lambda + \frac{1}{2\lambda} A(u_\lambda) \Leftrightarrow u_\lambda = \left( \operatorname{Id} + \frac{1}{2\lambda} A \right)^{-1} f.$$

Going towards the continuous model, in the hierarchical decomposition (4), we replace the scale parameter sequence  $\{\lambda_i\}$  with  $\{\frac{\lambda_i}{\tau}\}$  cf. [45], where  $\tau$  denotes the time-step. With this notation (4) can be written as

$$u_{\lambda_N} = \left[ \prod_{i=1}^N \left( \operatorname{Id} + \frac{\tau}{2\lambda_i} A \right)^{-1} \right] f.$$

Let  $\lambda(\cdot)$  denote a positive real-valued function and the scale parameter  $\lambda_i = \lambda(i\tau)$ . We know from the semigroup generation theory (see [18], Theorem A) that the limit

$$\lim_{N \rightarrow \infty} \left[ \prod_{i=1}^N \left( \operatorname{Id} + \frac{\tau}{2\lambda_i} A \right)^{-1} \right] f = u(\cdot, t)$$

exists for all  $t \in [0, \infty)$ , with  $\tau = \frac{t}{N}$  and it solves the following differential equation

$$\frac{\partial u}{\partial t} = -\frac{1}{2\lambda(t)} A(u), \quad u(\cdot, 0) = f, \quad \frac{\partial u}{\partial \nu} \Big|_{\Gamma} = 0. \quad (6)$$

### A. Weighted TV Flow

While the derivation above is quite general, we will work with a functional  $J(u)$  that is suitable for image processing. The choice of  $J(u)$  as the weighted bounded variation ( $BV$ ) seminorm was previously considered in [42], [43] for its edge-preserving properties. Properties of the weighted  $BV$  spaces and the corresponding minimization with  $L^2$  fidelity term were investigated in [15]. Motivated by the edge-preserving properties of the weighted TV minimization, we consider the weighted  $BV$  seminorm,  $J(u) = \int_{\Omega} \alpha |\nabla u|$ . Following [5] and [6], the operator  $-\operatorname{div}\left(\frac{\alpha \nabla u}{|\nabla u|}\right) \in A(u) = \partial J(u)$  is understood in the sense that for any  $u \in BV_{\alpha}(\Omega) \cap L^2(\Omega)$  there exists a bounded vector field  $z$  with the properties:  $\operatorname{div}(z) \in L^2(\Omega)$ ,  $\|z\|_{L^2} \leq 1$ ,  $A(u) = -\operatorname{div}(\alpha z)$  in  $\mathcal{D}'(\Omega)$ , satisfying

$$\int_{\Omega} \alpha z \cdot \nabla u = |u|_{TV_{\alpha}(\Omega)},$$

where the weighted  $BV$  seminorm,  $|u|_{TV_{\alpha}(\Omega)}$  is defined as

$$|u|_{TV_{\alpha}(\Omega)} := \sup \left\{ \int_{\Omega} u \operatorname{div}(\alpha \xi) dx : \xi \in C_c^1(\Omega, \mathbb{R}^2), \|\xi\|_{L^2} \leq 1 \right\}.$$

The Neumann boundary condition will be written in the form  $[z, \nu] = 0$  on  $\partial\Omega$ . With  $J(u) = |u|_{TV_{\alpha}(\Omega)}$  we say that the decomposition (3) is a  $(BV_{\alpha}, L^2)$  decomposition and, with  $A(u) = -\operatorname{div}\left(\frac{\alpha \nabla u}{|\nabla u|}\right)$  in (6), leads to the weighted TV flow

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mu(t) \operatorname{div} \left( \frac{\alpha \nabla u(\cdot, t)}{|\nabla u(\cdot, t)|} \right); \\ u : \Omega \times \mathbb{R}_+ &\mapsto \mathbb{R}, \quad \frac{\partial u}{\partial \nu} \Big|_{\Gamma} = 0, \quad u(\cdot, 0) = f, \end{aligned} \quad (7)$$

where we define  $\mu(t) := \frac{1}{2\lambda(t)}$  for convenience. Note that, the weighted TV flow (7) is closely related to the standard TV flow

$$\begin{aligned} \frac{\partial \hat{u}}{\partial t} &= \operatorname{div} \left( \frac{\nabla \hat{u}(\cdot, t)}{|\nabla \hat{u}(\cdot, t)|} \right); \\ \hat{u} : \Omega \times \mathbb{R}_+ &\mapsto \mathbb{R}, \quad \frac{\partial \hat{u}}{\partial \nu} \Big|_{\Gamma} = 0, \quad \hat{u}(\cdot, 0) = f, \end{aligned} \quad (8)$$

which has been extensively studied and is known to have a unique solution for  $f \in L^1(\Omega)$  (see [3], [4], [5]). Indeed, if  $\alpha \equiv 1$  and  $\gamma(t) = \int_0^t \mu(s) ds$ , then  $u(\cdot, t) = \hat{u}(\cdot, \gamma(t))$ . See [35], [28] for applications of TV flow in medical imaging.

### B. Designing the Edge-preserving Weighted TV Filter

The method based on the weighted functional has the flexibility to allow for weights  $\alpha$  chosen according to the image processing problem at hand. If the objective is to preserve edges, then the weight  $\alpha$  can be chosen to be small when the gradient of the image is large. To avoid mischaracterization of noise as edges, we often work with the convolution of the original image by a Gaussian kernel (or any smoothing kernel) with small variance. Thus we define

$$\alpha(x) = \frac{1}{\sqrt{1 + |\nabla G_{\sigma} * f|^2 / \beta^2}},$$

where  $\sigma$  denotes the standard deviation of the Gaussian kernel  $G_{\sigma}$  and for some  $\beta \in \mathbb{R}$ . The value of  $\beta$  is chosen so that  $\alpha$  attains a small value at the most prominent edge in the image. These types of edge detecting functions [36], [9] are common in the image processing literature.

We observe that for  $f \in [0, 1]$ , the maximum possible jump in the image (i.e.  $\max |\nabla G_{\sigma} * f|$ ) can not exceed 1. To ensure that the weight  $\alpha = \alpha_0$  at this jump, we set  $\beta = \alpha_0$ . In this paper, the value of  $\beta$  was set at 0.07, and the speed function is chosen as  $\mu(t) = 1.01^t$  unless otherwise mentioned.

Alternatively, the weight  $\alpha$  can be used to select the region where the noise is a major concern. For example, one can choose  $\alpha = G_{\sigma} * \chi_A$ , where  $A$  indicates the region of interest, where denoising is needed.

### C. Comparison Between Weighted TV Flow and Standard TV Flow

In this section we compare the edge-preserving property of the weighted TV flow (7) and the standard TV flow (8). The weight  $\alpha(x)$  in (7) can be thought of as a local adaptivity control (cf. [43]). With an appropriate choice of the weight  $\alpha$  we can improve the edge preservation. We demonstrate this property in Fig. 1. We added a Gaussian noise to a one dimensional synthetic image shown in Fig. 1(b). Then we performed denoising using standard TV flow and the weighted TV flow ( $\beta = 0.05$ ) with the same stopping criterion (estimated peak signal-to-noise ratio<sup>1</sup> (PSNR) = 31 dB). To solve (7) and (8) numerically, we used a semi-implicit scheme [46] to discretize the divergence term in (7) (see also [33], [38], [10]). We see in Fig. 1(c) and (d) that the edges are better preserved with the weighted TV flow. We observe in Fig. 2 that the intensity of the peak is also better preserved in the case of weighted TV flow in comparison to the standard TV flow. This feature is important MR images as the intensity is related to the tissue type [19], [26], and changes in intensity may result in tissue mischaracterization.

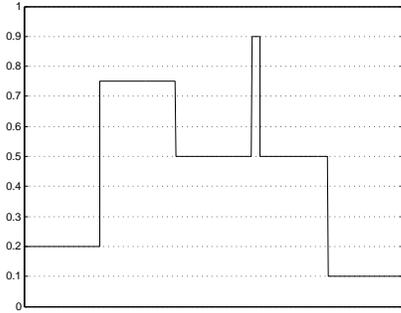
### D. Convergence to the Average Value in Finite Time

We observe that for the weighted TV flow (7) the function  $u(\cdot, t)$  has the same average value as that of the original image  $f$  at any time. This property is inherited from the standard TV flow [4]. Moreover, we know from [4] that for the standard TV flow (8),  $\hat{u}(\cdot, t)$  approaches the average of the function  $f$  i.e.  $\hat{u}(\cdot, t) \rightarrow f_{avg} := \int_{\Omega} f$ , in finite time and the flow  $u(\cdot, t)$  in (7) also exhibits this property.

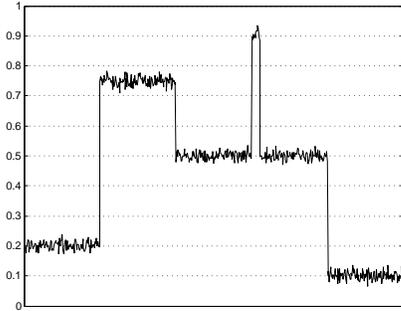
## III. MULTISCALE DENOISING APPLICATIONS

Noise is present in many imaging techniques and its nature differs from one modality to another. In an additive noise model, the given image  $f$  is the sum of the clean image,  $u$ , and some additive noise,  $v$ . Generally, we do not know the structure of the noise a priori. Nevertheless in most cases, noise is a *small-scale component* compared to the scale of other components in the clean image,  $u$ . The proposed

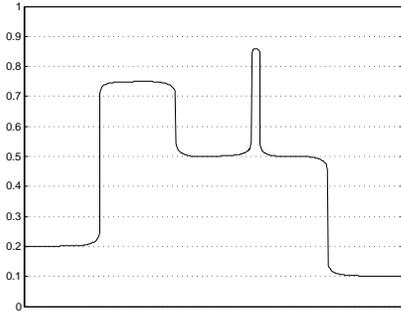
<sup>1</sup>PSNR( $u, f$ ) =  $10 \log_{10} \frac{MN}{\sum_{j=1}^N \sum_{i=1}^M (u_{i,j} - f_{i,j})^2}$ , where  $u$  and  $f$  are images of size  $M \times N$ .



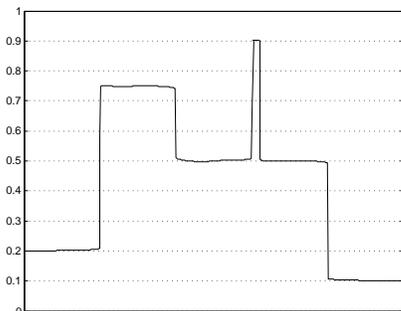
(a) Original synthetic image



(b) Original image with noise



(c) Denoising with standard TV flow



(d) Denoising with weighted TV flow

Fig. 1. The original image is shown in (a) and the image with additive Gaussian noise is depicted in (b). The noisy image in (b) is then denoised with the standard TV flow, shown in (c). Compare it with the results of the weighted TV flow in (d), where the edges are better preserved. Also note that the intensity of the flat regions is also better preserved with the weighted TV flow.

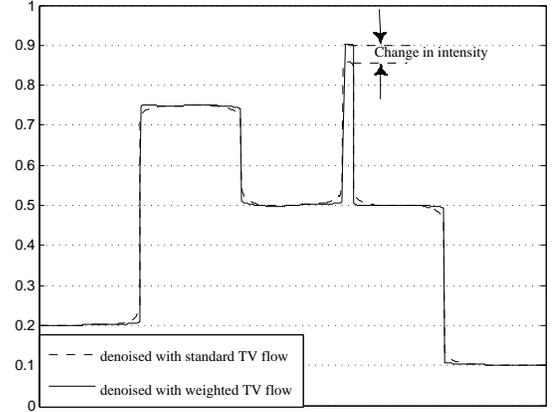


Fig. 2. Denoising result of standard TV flow and weighted TV flow superimposed. The dotted line shows the result of standard TV flow, and the solid line shows the result of weighted TV flow. We see that the weighted TV flow better maintains the intensity profile than the TV flow.

multiscale weighted TV flow (7) can be a valuable tool in denoising, where for small times  $t$ , the flow  $u(\cdot, t)$  is obtained by removing only finer scales from the image  $f$ . Thus, the weighted TV flow (7) can be effectively applied for denoising an image, while edge preservation is enhanced due to the weight  $\alpha$ .

#### A. Stopping Criterion

The multiscale family  $\{u(t)\}_{t \geq 0}$ , produced by the weighted TV flow is a single-parameter family of progressively smoother images. Indeed, we can show (see Proposition B.3) that the weighted TV seminorm of the flow  $u(t)$  decreases monotonically with time, indicating that  $u(t)$  becomes smoother over time. In fact  $u(t)$  reaches a constant in finite time. For the purpose of denoising we need to decide on the stopping time  $t_c$  such that the  $u(t_c)$  provides us with a denoised version of the given image  $f$ . We also observe in Proposition B.4 that the  $L^2$  energy of the flow  $u(t)$  also decreases monotonically with time and that the  $\alpha^*$  norm of the residual  $v(t)$  is bounded by  $\int_0^t \mu(s) ds$  (Proposition B.4).

Nevertheless, it is difficult to prescribe the stopping time based on the  $L^2$  energy of  $u(t)$  or the  $\alpha^*$  norm of the residual  $v(t)$ . Instead we propose to choose the stopping time,  $t_c$ , as the least time when the peak signal-to-noise ratio,  $\text{PSNR}(u(t), f)$  falls below the estimated PSNR (see [22], [17] for noise estimation), thus making ‘estimated PSNR’ as the only denoising parameter.

#### B. Fast Denoising

In this section we show that the selection of the speed of denoising function  $\mu(t)$  in (7) determines the speed of denoising. To this effect we define the weighted star-norm [5] as the dual of the weighted TV seminorm, i.e.

$$\|v\|_{\alpha^*} := \sup_{\substack{\varphi \in BV_\alpha(\Omega) \\ |\varphi|_{TV_\alpha(\Omega)} \neq 0}} \left\{ \frac{(v, \varphi)_{L^2(\Omega)}}{|\varphi|_{TV_\alpha(\Omega)}} \right\}. \quad (9)$$

We show in Proposition B.1 that the speed of the weighted TV flow, measured in  $\alpha^*$  norm is precisely given by the speed function  $\mu(t)$ . This suggests that we can perform fast denoising by selecting a rapidly increasing speed function. Indeed, this can be observed in experiments with different speed functions, summarized in Table I. We note that the speed is limited only by time discretization. We used the standard Shepp-Logan phantom (scaled in the range  $[0, 1]$ ) with additive Gaussian noise of zero mean and variance 0.005. The stopping criterion was  $\text{PSNR}(u(t), f) < 25$  dB.

### C. Comparative Study Between Weighted TV Flow and Bilateral Filter

The edge-preserving feature and speed of denoising are the main properties of the weighted TV flow. The edge detecting function  $\alpha$  that we chose in Sec. II-B is similar in nature to a function used by the Perona-Malik model [36], [13], which is an anisotropic diffusion method. We found that it performs qualitatively similar to the bilateral filter. In fact, it is shown in [44] that the bilateral filter is asymptotically equivalent to Perona-Malik filter. Thus, in this paper we chose to compare our method with bilateral filtering.

### D. Phantom Experiments Comparing to Bilateral Filter

We performed denoising experiments with the Shepp-Logan phantom, with an additive Gaussian noise of zero mean and varying standard deviations  $\sigma$ . The estimated  $\text{PSNR} = 10 \log(1/\sigma^2)$  was used as a stopping criterion for denoising with the weighted TV flow. Then we used bilateral filtering for denoising. To have a fair comparison, we adjusted the parameters of bilateral filters,  $\sigma_1$  and  $\sigma_2$  through an extensive 2D search, to obtain denoised images with the same PSNR as that obtained with weighted TV flow.

To examine image quality quantitatively, we looked at the contrast-to-noise ratio (CNR) [41] of the denoised images with respect to the clean Shepp-Logan phantom image. The  $\text{CNR}_{AB}$  is computed<sup>2</sup> for the regions  $A$  and  $B$ , as annotated in Fig. 3(a). The experiment was repeated 50 times for each variance. The average values of the CNR for bilateral filter and the weighted TV flow are shown in Table II. We observe that the weighted TV flow yields higher CNR images than bilateral filtering.

TABLE I  
FUNCTION  $\mu(t)$  DETERMINES THE SPEED OF THE WEIGHTED TV FLOW.

$\mu(t)$	PSNR (dB)	time (ms)
$0.5^t$	24.45	199
1	24.29	160
$10^t$	24.12	110
$100^t$	23.98	98
$10^{1+t}$	23.91	87

<sup>2</sup>CNR is computed with the formula:  $\text{CNR}_{AB} = |\mu_A - \mu_B|/\sigma$ , where  $\mu_A, \mu_B$  are the mean values of the intensities in regions  $A$  and  $B$  respectively, and  $\sigma = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$  with  $\sigma_A, \sigma_B$  indicating the standard deviations in regions  $A$  and  $B$  respectively.

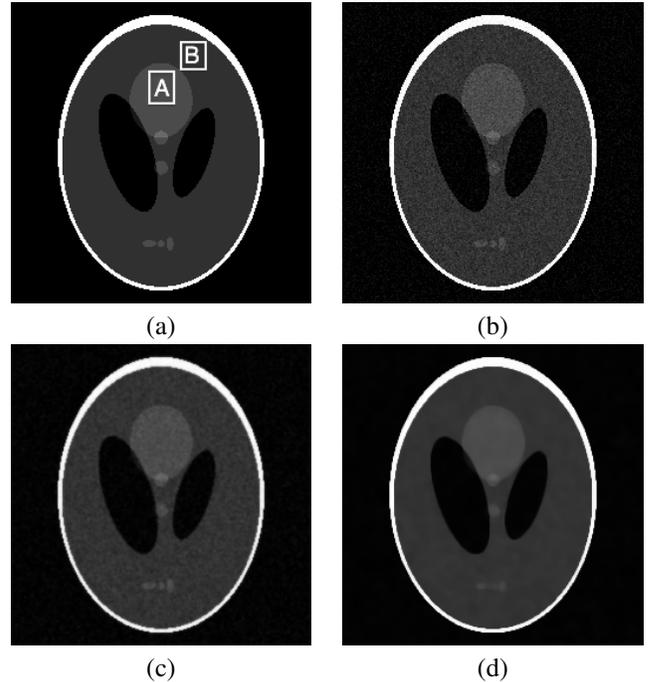


Fig. 3. Comparison between the bilateral filter and the proposed weighted TV flow in denoising. (a) Shepp-Logan phantom, (b) Noisy image,  $\sigma^2 = 0.002$ , (c) Denoising with bilateral filter (average  $\text{CNR}_{AB} = 6.62$ ), (d) Denoising with weighted TV flow (average  $\text{CNR}_{AB} = 21.08$ ).

TABLE II  
DENOISING RESULTS FOR THE SHEPP-LOGAN PHANTOM

Noise ( $\sigma^2$ )	PSNR dB	Avg. $\text{CNR}_{AB}$ with bilateral filter	Avg. $\text{CNR}_{AB}$ with weighted TV flow
0.0008	30.96	8.50	24.42
0.0010	30.00	8.22	23.17
0.0020	26.98	6.62	21.08
0.0025	26.02	6.33	21.04
0.0050	23.01	5.34	19.40

Image denoising with edge preservation could be used as an important preprocessing step in the problem of image registration. We discuss this application in the following section.

## IV. APPLICATIONS TO REGISTRATION

We have seen in section II that the weighted TV flow is obtained as a limiting case of a hierarchical multiscale decomposition, which makes it suitable for extracting different scales from given images. In this section, we demonstrate the use of the multiscale nature of the weighted TV flow in image registration or morphing problems where the goal is to obtain a geometrical transformation between two images. The registration transformations are usually obtained in a hierarchical manner, where an initial registration transformation is obtained by comparing coarse scales from given images; this map is then used as an initial guess to register finer scales, and so on. In this approach to we need different scales of the given images, which are obtained by the weighted TV flow.

Furthermore, the edge-preservation property of the weighted TV flow as demonstrated in Fig. 1 is especially useful in

techniques such as normalized gradient field (NGF) based registration [24]. The NGF based approach was proposed as an alternative to the mutual information (MI) based registration [51]. We note that NGF based methods share some of the same drawbacks of the MI based registration, namely non-convexity and multiple local minima; nevertheless, we argue that NGF based methods are better suited in the presence of shading artifacts which are common in magnetic resonance images. This is demonstrated by an explicit example in Appendix A where NGF based approach gives a correct global solution, whereas the MI based approach fails completely.

We will show in Sec. IV-D that the multiscale approach with weighted TV flow produces better results when used in NFG based technique, compared to other filtering techniques such as Gaussian, or bilateral filtering (Fig. 7). This is due to the fact that edges are important in NGF based approach, which are better preserved with the weighted TV flow than with other filtering methods. We also show in section IV-E that using the weighted TV flow in conjunction with MI also produces results comparable to other filtering methods (Fig. 9).

#### A. Registration as a Minimization Problem

Mathematically, the registration of a template image  $f$  and a reference image  $g$  can be posed as a minimization problem as follows:

$$\min_w \{ \mathcal{D}(f[w], g) \}, \quad (10)$$

where  $w : \mathbb{R}^n \rightarrow \mathbb{R}^n$  represents the geometric transformation, and  $\mathcal{D}$  represents a dissimilarity measure between the transformed template image  $f[w]$  and the reference image  $g$ . Possible distance measures  $\mathcal{D}$  include sum of squared differences (SSD), cross correlation, mutual information (MI) [50], [16], and normalized gradient fields (NGF) [24].

Various papers [37], [21], [12], [27], [40] have shown the use of mutual information as an effective similarity measure. If the images  $f$  and  $g$  take values from finite sets  $\{\alpha_i\}_{i=1}^m$ , and  $\{\beta_j\}_{j=1}^m$  respectively, then mutual information (MI) between two images  $f$  and  $g$ , is defined as follows [50]:

$$\begin{aligned} MI(f, g) = & - \sum_i p(\alpha_i) \log p(\alpha_i) - \sum_j p(\beta_j) \log p(\beta_j) \\ & + \sum_i \sum_j p(\alpha_i, \beta_j) \log p(\alpha_i, \beta_j), \end{aligned} \quad (11)$$

where  $p(\alpha_i)$ ,  $p(\beta_j)$  represent the marginal probability distribution of grey values  $\alpha_i$ ,  $\beta_j$  in images  $f$  and  $g$  respectively, and  $p(\alpha_i, \beta_j)$  represents the joint probability distribution of grey values  $(\alpha_i, \beta_j)$  in the overlapped region between images  $f$  and  $g$ . Note, to formulate the registration as a minimization problem, we seek to minimize the negative of the mutual information, i.e. we use the following dissimilarity measure:

$$\mathcal{D}_{MI}(f[w], g) = -MI(f[w], g).$$

Mutual information is especially useful for multimodal images [51], where matching of corresponding intensity patterns is more important than the actual intensity values. Nevertheless, there are well-known problems with this approach [24], such

as non-convexity of the registration problem and discrete structure of the mutual information. An alternative dissimilarity metric was suggested in [24], which makes use of normalized gradient fields (NGF) of images. The NGF metric is defined as follows:

$$\mathcal{D}_{NGF}(f, g) = - \int_{\Omega} \langle \nu(f, x), \nu(g, x) \rangle^2 dx, \quad (12)$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product, and  $\nu$  represents the normalized gradient of the image,

$$\nu(I, x) := \frac{\nabla I(x)}{\|\nabla I(x)\|_{\rho}}, \quad (13)$$

and  $\|\cdot\|_{\rho}$  for a vector  $\xi \in \mathbb{R}^n$  is defined as

$$\|\xi\|_{\rho} = \sqrt{\sum_{i=1}^n \xi_i^2 + \rho^2}, \quad \text{with } \rho := \frac{\eta}{|\Omega|} \int_{\Omega} |\nabla I(x)| dx,$$

where  $\eta$  is estimated noise level and  $|\Omega|$  is the volume of the domain. The concept behind the gradient based registration is derived from the observation that if two images are similar, many intensity changes should occur at the same spatial location, which can be detected by a large value of the inner product between the normalized gradients of the two images.

Nevertheless, the NGF based approach depends on the faithful preservation of prominent edges, which are often lost due to noise.

#### B. Hierarchical Registration Framework

Presence of noise is one of the main problems in image registration. This is usually addressed by using denoising as a preprocessing step. Furthermore, to improve the speed, multiscale registration approach is used [34]. The multiscale representation of the template and target images is produced by some smoothing filter. Depending on the smoothing filter used, the smoothed images may not preserve the structure of images. For example, isotropic filtering such as Gaussian smoothing would smooth the edges; whereas, anisotropic filtering such as bilateral filtering produces a 2-dimensional multiscale family [48], which makes it difficult to use in practice.

On the other hand, being obtained as a limiting case of a multiscale decomposition, the weighted TV flow is inherently multiscale in nature. Indeed, the flow  $\{u(\cdot, t)\}_{t \geq 0}$  can be realized as a family of images in which small-scale features are successively removed. Thus, for  $t_1 < t_0$ , the image  $u(\cdot, t_0)$  is a coarser image than  $u(\cdot, t_1)$ , producing a single-parameter multiscale family,  $\{u(\cdot, t)\}_{t \geq 0}$ . This phenomenon is also seen from the proposition B.3, which states that the total variation in the image  $u(\cdot, t)$  is always decreasing.

These properties enable us to introduce the hierarchical multiscale registration framework using weighted TV flow, shown in Fig. 4 (see [30], [32] for registration based on  $(BV, L^1)$  decomposition). For this section, the multiscale families generated by the weighted TV flow with  $f$  and  $g$  as the initial images are denoted by  $\{u^f(\cdot, t)\}_{t \geq 0}$  and  $\{u^g(\cdot, t)\}_{t \geq 0}$ , respectively.

The initial unregistered images  $f$  and  $g$  are inputs to the registration framework along with an initial estimate  $w^{(0)}$

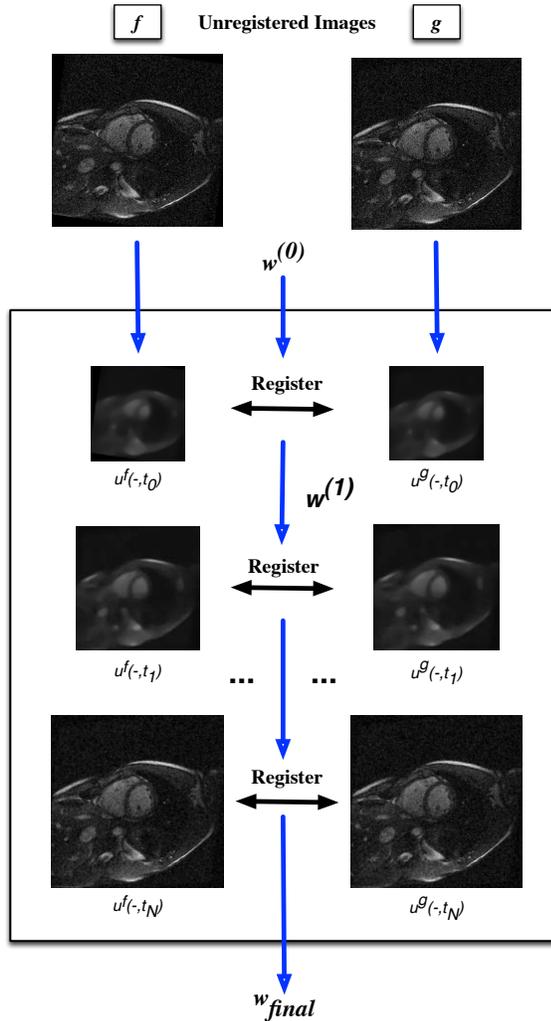


Fig. 4. Multiscale registration framework is demonstrated. Original input images are down-sampled and registered at lower scales, before the estimated registration parameters are propagated as the initial guess to images at higher resolutions.  $w^{(0)}$  represents the initial guess of the optimal registration parameters, and  $w_{final}$  corresponds to the final estimate of the registration parameters.

for the registration transformation. The images  $u^f(\cdot, t_0)$  and  $u^g(\cdot, t_0)$  are registered, and the registration transformation  $w^{(1)}$  is used as an initial estimate for the registration of finer scale images  $u^f(\cdot, t_1)$  and  $u^g(\cdot, t_1)$  where  $t_1 < t_0$ . This process is repeated, and the optimal transformation  $w^{(i+1)}$  is used as the initial estimate for the next finer scale registration until  $w_{final}$  is obtained at the finest level.

### C. Methods

To quantitatively evaluate the performance of the registration framework with weighted TV flow, a controlled experiment was carried out in this study. Although the hierarchical framework presented in section IV-B is applicable for any type of registration, for the validation purpose we present simulations for rigid registration. To this effect, a mid short axis MR (magnetic resonance) cardiac image  $g_0$  was acquired from each of 10 healthy volunteers. The physical locations

of the landmarks within the left and right ventricles were recorded for each of these 10 images. One of the acquired cardiac images with the corresponding landmarks is shown in Fig. 5.

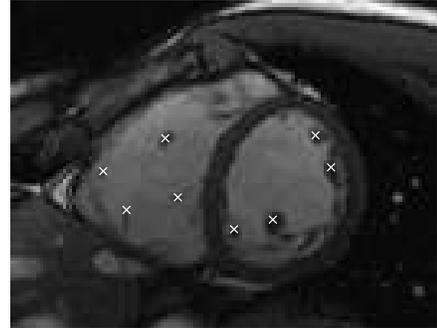


Fig. 5. Manually delineated landmarks in the left and right ventricles, as indicated by the 'x' markings.

Subsequently, we generated a series of known rigid transformations of  $g_0$  consisting of rotation  $\theta^\circ$ , translation  $t_{x_1}$  mm in the  $x_1$ -direction and translation  $t_{x_2}$  mm in the  $x_2$ -direction, resulting in the transformed image  $f_0$ . We identify this transformation with the vector of the transformation parameters  $\omega := \langle \theta, t_{x_1}, t_{x_2} \rangle$ .

For each transformation  $\omega := \langle \theta, t_{x_1}, t_{x_2} \rangle$  of the image  $g_0$ , the mean displacement of the landmarks in Fig. 5 is denoted by  $d$ . Fixing the mean displacement  $d$  of the landmarks at  $d = 5, 10, 15$  and  $20$  mm, we generated 100 transformations  $\omega := \langle \theta, t_{x_1}, t_{x_2} \rangle$  for each value of  $d$ , creating 400 distinct transformed versions  $f_0$  for each of the 10 acquired images  $g_0$ , resulting in 4000 simulations. The parameters  $\theta, t_{x_1}$ , and  $t_{x_2}$  are randomly generated with independent and identically uniform distributions.

To simulate low signal-to-noise ratio (SNR) conditions, Rician noise  $R \sim \text{Rice}(\nu, \sigma^2)$  with  $\nu = 0$  and  $\sigma = 0.1$  was added to images  $f_0$  and  $g_0$  to obtain noisy images  $f$  and  $g$  respectively. An example of a low SNR image (i.e. SNR = 3.16) is illustrated in Fig. 6 (a). The transformed and noisy image  $f$  serves as the template image in the multiscale registration framework 4, where we solve the minimization problem (10) to obtain the optimal transformation  $\omega_{final}$ . Since the physical landmark locations are known, the target registration error (TRE) can be computed as the mean distance between the landmark locations of the registered template image  $f[\omega_{final}]$  and the reference image.

### D. Experiments with Hierarchical NGF Based Registration

The results of the hierarchical registration using the NGF based metric (12) are shown in Fig. 7, where the target registration errors are shown for four different hierarchical registration approaches. Specifically, from left to right, each group of TREs corresponds to hierarchical registration using NGF without preprocessing, and preprocessing using Gaussian filtering, bilateral filtering, and weighted TV flow to smooth the input images respectively. Within each group, 4 bar plots

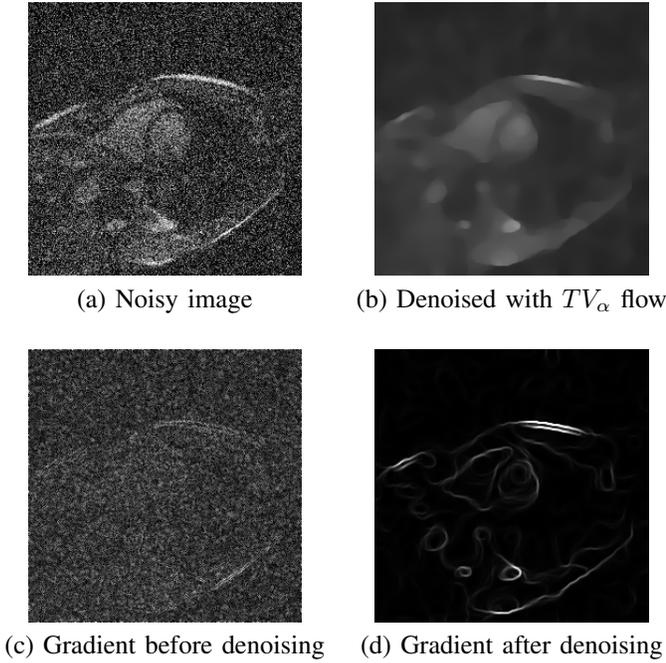


Fig. 6. A noisy image  $f$  and a weighted TV flow smoothed image  $u^f$  are shown in (a) and (b) respectively. The corresponding magnitude of their gradient images  $|\nabla f|$  and  $|\nabla u^f|$  are shown in (c) and (d) respectively.

are shown, the height of each bar represents the mean TRE for all 10 volunteer datasets after registration recovery from the transformations that caused an initial displacement of 5 mm, 10 mm, 15 mm or 20 mm from their respective original landmark locations.

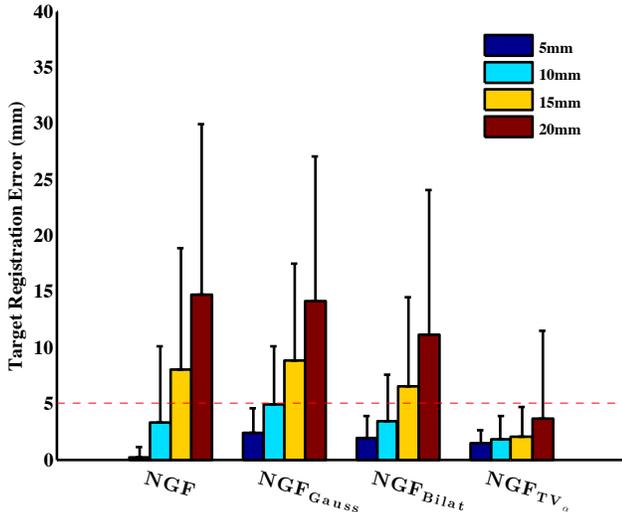


Fig. 7. Registration results are shown for different preprocessing techniques.  $NGF$ ,  $NGF_{Gauss}$ ,  $NGF_{Bilat}$ , and  $NGF_{TV_\alpha}$  represents no preprocessing, Gaussian filtering, bilateral filtering, and weighted TV flow filtering for each scale of the hierarchical registrations respectively.

As expected, the registration accuracy improved using each of the 3 denoising methods. However, the best results were achieved using weighted TV flow for preprocessing. The Wilcoxon signed-rank test [52] was performed with signif-

icance at the 1% level. Specifically, the TRE results using each denoising method was compared in turn to the weighted TV flow ( $NGF_{TV_\alpha}$ ) method. In each test, the null hypothesis stated that the  $NGF_{TV_\alpha}$  method and the other preprocessing method performed equally well; whereas, the alternative hypothesis stated that  $NGF_{TV_\alpha}$  performed better by achieving lower TRE. It was shown at the 1% significance level that the  $NGF_{TV_\alpha}$  method achieved better registration accuracy than  $NGF_{Gauss}$  ( $p < 10^{-5}$ ), and  $NGF_{Bilat}$  ( $p < 10^{-5}$ ).

Furthermore, in this experiment, weighted TV flow is the only method that achieved a mean TRE of less than 5 mm for all four initial displacement categories. This improvement is due to the fact that weighted TV flow was able to suppress the Rician noise 6 (b), at the same time preserving the structural edges in the input images (see Fig. 6 (c)-(d)).

An example of a successful NGF based registration with weighted TV flow is illustrated in Fig. 8, where an initially misaligned template image  $f$  and a reference image  $g$  are shown in Fig. 8 (a) and (b) respectively. Due to the initial misalignment, the difference image shows misregistered chest wall structures in Fig. 8 (c). After using the proposed registration framework, the template image is registered to the reference image and their difference image  $|f[\omega_{final}] - g|$  (Fig. 8 (d)) no longer displays prominent structures, indicating successful registration.

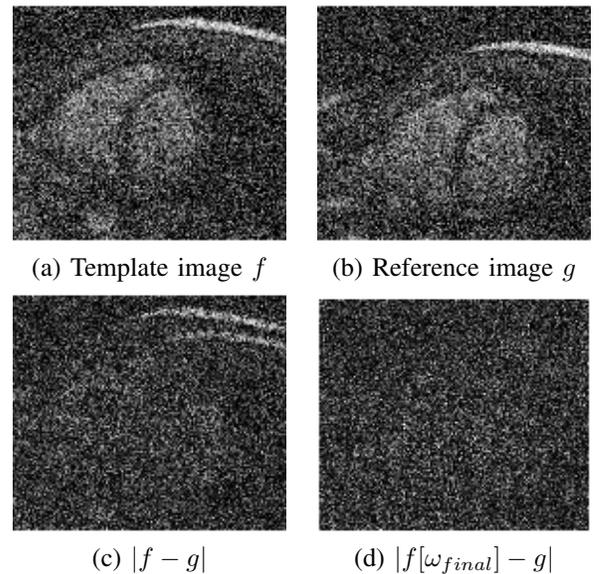


Fig. 8. Registration example: (a) initially misaligned template image  $f$ ; (b) reference image  $g$  (c) absolute difference image between the initially misaligned template and the reference images  $|f - g|$ ; (d) absolute difference image between registered template and reference images  $|f[\omega_{final}] - g|$ .

### E. Experiments with Hierarchical MI Based Registration

Moreover, the same experiment is performed using a hierarchical registration framework with the MI based dissimilarity measure (11) instead of the NGF based measure (12). The registration results are shown in Fig. 9. Again, all 3 denoising methods improved the registration accuracy compared to registration using MI without any preprocessing procedure

( $p < 0.01$ ). Further comparison of the 3 denoising methods using the Wilcoxon signed-rank test showed no significant difference between them at the 1% level. However, we observed that for experimental trials with a large initial misalignment of  $d = 20$  mm, the chance of converging to a wrong solution was higher in the case of preprocessing with Gaussian or bilateral filters as opposed to using weighted TV flow. Specifically, the percentage of trials that achieved a TRE  $> 10$  mm after registration were 5.2%, 4.5%, and 2.5% for Gaussian, bilateral, and weighted TV flow preprocessing respectively. As a result, reduced standard deviations of the TREs are observed in Fig. 9. Therefore, weighted TV flow has been demonstrated to improve the robustness of mutual information (MI) based registration in cases of substantial initial misalignment.

We remark that even where the MI based registration outperforms NGF based registration (due to absence of clear edges), the use of weighted TV flow for preprocessing is beneficial.

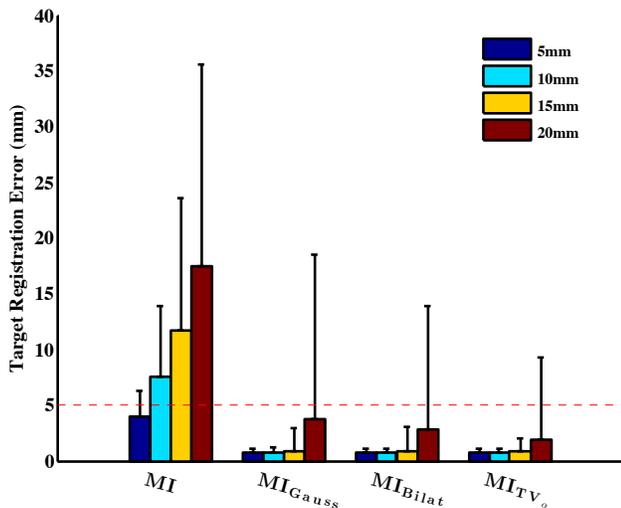


Fig. 9. Registration results are shown for different preprocessing techniques.  $MI$ ,  $MI_{Gauss}$ ,  $MI_{Bilat}$ , and  $MI_{TV_\alpha}$  represents no preprocessing, Gaussian filtering, bilateral filtering, and weighted TV flow filtering for each scale of the hierarchical registrations respectively.

## V. CONCLUSION

We realized the weighted TV flow as a limiting case of a multiscale decomposition. In this setting the flow can be viewed as successive removal of the small-scale structures. This feature was successfully used in image denoising. The weighted TV flow was also shown to be better at edge preservation than the standard TV flow and the bilateral filtering. Moreover, the speed of the flow could be controlled through the speed function, which could be effectively used for real time denoising.

The inherently multiscale nature of the weighted TV flow enables us to extract different scales from an image. Indeed the weighted TV flow generates a single parameter multiscale family. This feature was successfully employed in a multiscale registration application. Due to the edge-preserving nature of the weighted TV flow, we observed statistically significant

improvement in gradient based registration. We also showed that the weighted TV flow produced comparable results to other smoothing methods when used in conjunction with the mutual information based registration.

## APPENDIX A MUTUAL INFORMATION VERSUS NORMALIZED GRADIENT FIELD

In this section, we argue that in presence of a phenomenon such as signal intensity shading artifact in magnetic resonance imaging [25], registration using mutual information can produce an incorrect solution, whereas gradient based registration prevails. To this effect, we produce an explicit example below.

Let  $\Omega$  be the domain,  $\Omega = [-1, 1] \times [-1, 1]$ . Let  $A \subset \Omega$  and  $B \subset \Omega$  be two square regions defined by  $A = [-\frac{1}{6}, \frac{1}{6}] \times [-\frac{1}{6}, \frac{1}{6}]$ ,  $B = [-\frac{5}{6}, -\frac{1}{2}] \times [\frac{1}{2}, \frac{5}{6}]$ . We construct a simple image function  $f$ ,

$$f = \chi_\Omega + \chi_A + \frac{1}{2}\chi_B,$$

where  $\chi_\Omega$ ,  $\chi_A$  and  $\chi_B$  indicate the characteristic functions on  $\Omega$ ,  $A$  and  $B$  respectively. We let  $g$  be the image obtained by rotating  $f$  counterclockwise through  $90^\circ$ . Now let us construct a multiplier function

$$\psi(x_1, x_2) = 1 + \frac{(x_2 + 1)^2}{16}$$

to obtain images  $\psi f$  and  $\psi g$  as shown below in Fig. 10. Such type of multiplicative function can be observed in MR imaging due to coil sensitivity drop-off.

For such function  $\psi$ , with  $1 \leq \psi \leq \frac{5}{4}$ , we would expect that a robust registration metric would succeed in registering the images  $\psi f$  and  $\psi g$  shown in Fig. 10. Nevertheless, as we rotate  $\psi f$ , and calculate the distance  $\mathcal{D}_{MI}$  between the rotated  $\psi f$  and  $\psi g$ , a global minimum is observed at  $\theta = 0^\circ$ , which is clearly incorrect; in fact,  $\mathcal{D}_{MI}$  attains a maximum at the correct solution,  $\theta = 90^\circ$  (see Fig. 11). The failure of the MI based registration is due to the fact that it attempts to minimize the entropy of the joint histogram, which is distorted due to the sensitivity function  $\psi$ .

Now we look at the registration using normalized gradient fields (NGF) metric (12). As opposed to the mutual information, we observe that registration using NGF has a global minimum at  $\theta = 90^\circ$  (see Fig. 11). This is due to the fact that the NGF emphasizes matching of edges in two images. (Note that we need to use  $\rho$  in (13) such that the normalized gradient  $\nu$  at prominent edges is much higher compared to the smoother region in the image. For  $\psi f, \psi g \in [0, 1]$  we used  $\rho = 0.01$ .)

## APPENDIX B SPEED AND SCALE RELATED PROPERTIES OF THE WEIGHTED TV FLOW

In this Appendix we give intuitive derivations of the speed and scale related properties of the multiscale weighted TV flow. We use the characterization of the subdifferential  $\partial J(u)$  discussed in section II-A and the boundary condition  $[z, \nu] = 0$

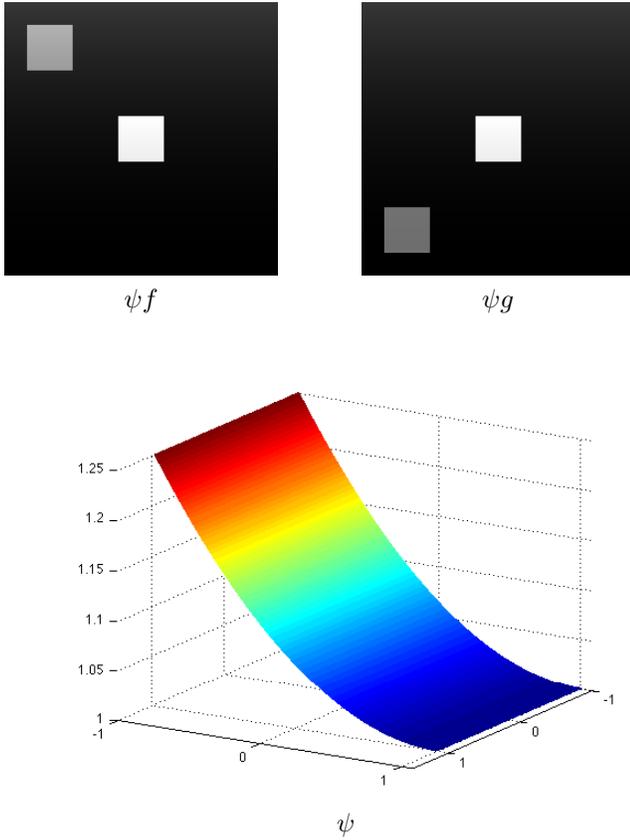


Fig. 10. Synthetic images  $\psi f$ ,  $\psi g$  and the multiplier function  $\psi$  representing phenomenon similar to coil shading.

on  $\partial\Omega$  to facilitate the integration in space. In particular, for  $u, \phi \in BV_\alpha(\Omega) \cap L^2(\Omega)$  we have

$$\begin{aligned} - \int_{\Omega} \operatorname{div} \left( \frac{\alpha \nabla u}{|\nabla u|} \right) u &= |u|_{TV_\alpha(\Omega)}, \\ - \int_{\Omega} \operatorname{div} \left( \frac{\alpha \nabla u}{|\nabla u|} \right) \phi &\leq |\phi|_{TV_\alpha(\Omega)}. \end{aligned}$$

First, we compute the  $\alpha^*$  norm (9) of the speed of the weighted TV flow. To this effect we prove the following proposition.

**Proposition B.1.** *For the multiscale weighted TV flow (7) the  $\alpha^*$  norm of the rate of change of the flow  $u(\cdot, t)$  is:*

$$\left\| \frac{\partial u}{\partial t}(\cdot, t) \right\|_{\alpha^*} = \mu(t). \quad (14)$$

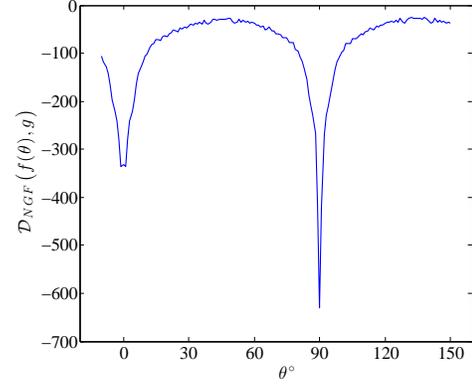
*Proof.* We first observe that integrating (7) with  $\varphi \in BV_\alpha(\Omega) \cap L^2(\Omega)$  in space yields

$$\left| \left( \frac{\partial u}{\partial t}(\cdot, t), \varphi \right)_{L^2(\Omega)} \right| \leq \mu(t) |\varphi|_{TV_\alpha(\Omega)}.$$

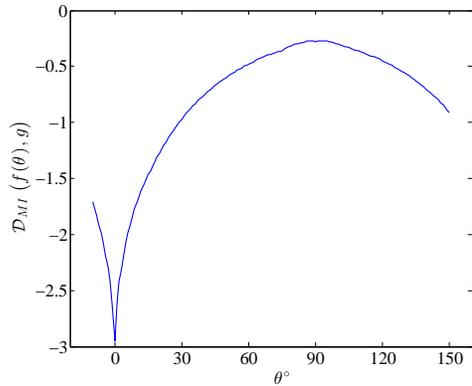
If  $\varphi = u(\cdot, t)$ , we get an equality, and the conclusion follows.  $\square$

The equation (14) asserts that the speed of the multiscale flow (7), measured in  $\alpha^*$  norm is exactly equal to  $\mu(t)$ . This property can be used for fast denoising by manipulating the speed function  $\mu(t)$ . As  $v(\cdot, t) = f - u(\cdot, t)$ , it follows from the above Proposition B.1 that

$$\left\| \frac{\partial v}{\partial t}(\cdot, t) \right\|_{\alpha^*} = \mu(t). \quad (15)$$



a)



b)

Fig. 11. Dissimilarity measures: a)  $\mathcal{D}_{MI}$  and b)  $\mathcal{D}_{NGFF}$  between  $\psi f$  rotated by  $\theta^\circ$  and  $\psi g$ . Correct alignment between  $\psi f$  and  $\psi g$  images is a counterclockwise rotation of  $90^\circ$  applied to  $\psi f$ .

Now, we consider the rate of change of the  $L^2$ -energy and of the  $TV_\alpha$ -seminorm of the multiscale flow  $u(\cdot, t)$ . Integrating (7) with  $u(\cdot, t)$  in space we get

$$\begin{aligned} \int_{\Omega} u \frac{\partial u}{\partial t} dx &\equiv \frac{1}{2} \frac{d}{dt} \|u(\cdot, t)\|_{L^2(\Omega)}^2 \\ &= \mu(t) \int_{\Omega} \operatorname{div} \left( \frac{\alpha \nabla u(\cdot, t)}{|\nabla u(\cdot, t)|} \right) u(\cdot, t) dx \\ &\equiv -\mu(t) |u(\cdot, t)|_{TV_\alpha(\Omega)}. \end{aligned}$$

Thus, we have

$$\frac{d}{dt} \|u(\cdot, t)\|_{L^2(\Omega)}^2 = -2\mu(t) |u(\cdot, t)|_{TV_\alpha(\Omega)}.$$

As  $\mu(t) > 0$  for all  $t$ , we note that the  $L^2$  energy of the  $BV_\alpha$ -part never increases. We state this as a proposition.

**Proposition B.2.** *For the multiscale TV flow (7) the  $L^2$  energy of the function  $u$  is non-increasing for all times  $t$ .*

We also note that the weighted  $BV$  seminorm of the flow  $u(\cdot, t)$  is also non-increasing. To this effect we have the following proposition.

**Proposition B.3.** *For the multiscale weighted TV flow (7) the weighted BV seminorm of  $u(\cdot, t)$  is non-increasing for any time  $t$ .*

*Proof.* For the PDE (7) we look at the time derivative of  $|u(\cdot, t)|_{TV_\alpha(\Omega)}$ .

$$\begin{aligned} \frac{d}{dt}|u(\cdot, t)|_{TV_\alpha(\Omega)} &= \int_{\Omega} \frac{\alpha \nabla u}{|\nabla u|} \frac{\partial}{\partial t}(\nabla u) dx \\ &= \int_{\Omega} \frac{\alpha \nabla u}{|\nabla u|} \nabla \left( \frac{\partial u}{\partial t} \right) dx \\ &= - \int_{\Omega} \operatorname{div} \left( \frac{\alpha \nabla u(\cdot, t)}{|\nabla u(\cdot, t)|} \right) \frac{\partial u}{\partial t} dx \\ &= - \int_{\Omega} \frac{1}{\mu(t)} \left( \frac{\partial u}{\partial t} \right)^2 dx \\ &\leq 0. \end{aligned}$$

The fourth equality follows from (7), and the last inequality follows as  $\mu(t) > 0$  for all  $t$ .  $\square$

Now, we look at the  $\alpha^*$  norm of the residual of the weighted TV flow.

**Proposition B.4.** *For the weighted TV flow (7) we define the residual  $v(\cdot, t) := f - u(\cdot, t)$ . Then we have the following inequality for its star-norm*

$$\|v(\cdot, t)\|_{\alpha^*} \leq \int_0^t \mu(s) ds.$$

*Proof.* For a function  $\varphi \in BV_\alpha(\Omega) \cap L^2(\Omega)$  we have

$$\begin{aligned} |(v(\cdot, t), \varphi)_{L^2(\Omega)}| &= \left| \int_{\Omega} \int_0^t \mu(s) \operatorname{div} \left( \frac{\nabla u(\cdot, s)}{|\nabla u(\cdot, s)|} \right) ds \varphi dx \right| \\ &\leq \left| \int_0^t \mu(s) |\varphi|_{TV_\alpha(\Omega)} ds \right| \\ &= |\varphi|_{TV_\alpha(\Omega)} \int_0^t \mu(s) ds. \end{aligned}$$

The conclusion then follows from taking the supremum over  $\varphi \in BV_\alpha(\Omega)$ .  $\square$

**Remark B.1** (Bound on the star-norm of an image). *We have seen that the proposed multiscale TV flow (7) is a scaled version of the standard TV flow (8) with the scaling function  $\gamma(t) = \int_0^t \mu(s) ds$ . We see from the Proposition B.4 that this scaling function is the upper bound on the star-norm of the residual. Moreover, we know that  $v(\cdot, T) = f - f_{avg}$  for some finite time  $T$ . Hence, we have*

$$\|f - f_{avg}\|_{\alpha^*} = \|v(\cdot, T)\|_{\alpha^*} \leq \int_0^T \mu(s) ds.$$

*In particular, if the function  $f$  is normalized such that  $f_{avg} = 0$ , then*

$$\|f\|_{\alpha^*} = \|v(\cdot, T)\|_{\alpha^*} \leq \int_0^T \mu(s) ds.$$

*This gives us a novel way of looking at the star-norm of a function  $f$ , in terms of the speed function  $\mu(t)$ .*

## REFERENCES

- [1] R. Acar and C. Vogel, *Analysis of bounded variation penalty methods of ill-posed problems*, Inverse Problem **10** (1994), 1217–1229.
- [2] G. Adluru, T. Tasdizen, M. Schabel, and E. DiBella, *Reconstruction of 3D dynamic contrast-enhanced magnetic resonance imaging using nonlocal means*, Journal of Magnetic Resonance Imaging **32** (2010), no. 5, 1217–1227.
- [3] F. Andreu, C. Ballester, V. Caselles, and J. Mazón, *Minimizing total variational flow*, Differential and integral equations **14** (2001), no. 3, 321–360.
- [4] F. Andreu, V. Caselles, J. Diaz, and J. Mazón, *Some qualitative properties for the total variational flow*, Journal of functional analysis **188** (2002), 516–547.
- [5] F. Andreu-Vaillio, V. Caselles, and J. Mazón, *Parabolic quasilinear equations minimizing linear growth functionals*, Progress in mathematics, vol. 223, Birkhäuser, Switzerland, 2004.
- [6] G. Anzellotti, *Pairings between measures and bounded functions and compensated compactness*, Ann. di Matematica Pura ed Appl. **135** (1983), no. 1, 293–318.
- [7] P. Athavale, *Novel integro-differential schemes for multiscale image representation*, Ph.D. thesis, University of Maryland, College Park, 2009.
- [8] P. Athavale, R. Xu, P. Radau, A. Nachman, and G. Wright, *Multiscale TV flow with applications to fast denoising and registration*, SPIE Medical Imaging, International Society for Optics and Photonics, 2013, pp. 86692K–86692K.
- [9] G. Aubert and P. Kornprobst, *Mathematical problems in image processing: Partial differential equations and the calculus of variations (second edition)*, Applied Mathematical Sciences, vol. 147, Springer-Verlag, 2006.
- [10] A. Borsic, B. Graham, A. Adler, and W. Lionheart, *In vivo impedance imaging with total variation regularization*, IEEE Transactions on Medical Imaging **29** (2010), no. 1, 44–54.
- [11] H. Brezis, *Analyse fonctionnelle*, vol. 5, Masson, 1983.
- [12] A. Carrillo, J. Duerk, J. Lewin, and D. Wilson, *Semiautomatic 3-D image registration as applied to interventional MRI liver cancer treatment*, Medical Imaging, IEEE Transactions on **19** (2000), no. 3, 175–185.
- [13] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal. **29** (1992), no. 1, 182–193.
- [14] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math. **76** (1997), no. 2, 167–188.
- [15] Y. Chen and M. Rao, *Minimization problems and associated flows related to weighted  $p$  energy and total variation*, SIAM journal on mathematical analysis **34** (2003), no. 5, 1084–1104.
- [16] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal, *Automated multi-modality image registration based on information theory*, Information processing in medical imaging, vol. 3, 1995, pp. 263–274.
- [17] P. Coupé, J. Manjón, E. Gedamu, D. Arnold, M. Robles, and D. Collins, *Robust rician noise estimation for MR images*, Medical image analysis **14** (2010), no. 4, 483–493.
- [18] M. Crandall and T. Liggett, *Generation of semi-groups of nonlinear transformations on general banach spaces*, American Journal of Mathematics **93** (1971), no. 2, 265–298.
- [19] J. Detsky, G. Paul, J. Alexander, and A. Graham, *Reproducible classification of infarct heterogeneity using fuzzy clustering on multicontrast delayed enhancement magnetic resonance images*, Medical Imaging, IEEE Transactions on **28** (2009), no. 10, 1606–1614.
- [20] L. Evans and R. Gariepy, *Measure theory and fine properties of functions*, CRC press, Boca Raton, FL, 1992.
- [21] B. Fei, A. Wheaton, Z. Lee, J. Duerk, and D. Wilson, *Automatic MR volume registration and its evaluation for the pelvis and prostate*, Physics in Medicine and Biology **47** (2002), no. 5, 823.
- [22] H. Gudbjartsson and S. Patz, *The rician distribution of noisy MRI data*, Magnetic Resonance in Medicine **34** (1995), no. 6, 910–914.
- [23] M. Guerquin-Kern, M. Haberlin, K. Pruessmann, and M. Unser, *A fast wavelet-based reconstruction method for magnetic resonance imaging*, Medical Imaging, IEEE Transactions on **30** (2011), no. 9, 1649–1660.
- [24] E. Haber and J. Modersitzki, *Intensity gradient based registration and fusion of multi-modal images*, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006, Springer, 2006, pp. 726–733.
- [25] J. Haselgrove and M. Prammer, *An algorithm for compensation of surface-coil images for sensitivity of the surface coil*, Magnetic Resonance Imaging **4** (1986), no. 6, 469–472.

- [26] W. Heindel, T. Jernigan, S. Archibald, C. Achim, E. Masliah, and C. Wiley, *The relationship of quantitative brain magnetic resonance imaging measures to neuropathologic indexes of human immunodeficiency virus infection*, Archives of neurology **51** (1994), no. 11, 1129.
- [27] M. Holden, D. Hill, E. Denton, J. Jarosz, T. Cox, T. Rohlfing, J. Goodey, and D. Hawkes, *Voxel similarity measures for 3-D serial MR brain image registration*, Medical Imaging, IEEE Transactions on **19** (2000), no. 2, 94–102.
- [28] L. Jonasson, X. Bresson, J-P Thiran, V. Wedeen, and P. Hagmann, *Representing diffusion MRI in 5-D simplifies regularization and segmentation of white matter tracts*, Medical Imaging, IEEE Transactions on **26** (2007), no. 11, 1547–1554.
- [29] J. Koenderink, *Structure of images*, Biological Cybernetics **50** (1984), 363–370.
- [30] D. Li, H. Li, H. Wan, J. Chen, G. Gong, H. Wang, L. Wang, and Y. Yin, *Multiscale registration of medical images based on edge preserving scale space with application in image-guided radiation therapy*, Physics in Medicine and Biology **57** (2012), no. 16, 5187.
- [31] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations*, University Lecture Series, vol. 22, American Mathematical Society, Providence, RI, 2001, The fifteenth Dean Jacqueline B. Lewis memorial lectures.
- [32] Jan Michálek, Martin Čapek, and Lucie Kubínová, *Nonrigid registration of CLSM images of physical sections with discontinuous deformations*, Microscopy and Microanalysis **17** (2011), no. 6, 923.
- [33] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, and B. Thirion, *Total variation regularization for fMRI-based prediction of behavior*, Medical Imaging, IEEE Transactions on **30** (2011), no. 7, 1328–1340.
- [34] J. Modersitzki, *Numerical methods for image registration (numerical mathematics and scientific computation)*, Oxford university press USA, 2004.
- [35] A. Montillo, J. Udupa, L. Axel, and D. Metaxas, *Interaction between noise suppression and inhomogeneity correction in MRI*, Medical Imaging 2003, International Society for Optics and Photonics, 2003, pp. 1025–1036.
- [36] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, Tech. report, University of California at Berkeley, Berkeley, CA, USA, 1988.
- [37] J. Pluim, J. Maintz, and M. Viergever, *Mutual-information-based registration of medical images: a survey*, Medical Imaging, IEEE Transactions on **22** (2003), no. 8, 986–1004.
- [38] S. Ramani and J. Fessler, *A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction*, Medical Imaging, IEEE Transactions on **31** (2012), no. 3, 677–688.
- [39] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), 259–268.
- [40] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes, *Nonrigid registration using free-form deformations: application to breast MR images*, Medical Imaging, IEEE Transactions on **18** (1999), no. 8, 712–721.
- [41] X. Song, B. Pogue, S. Jiang, M. Doyley, H. Dehghani, T. Tosteson, and K. Paulsen, *Automated region detection based on the contrast-to-noise ratio in near-infrared tomography*, Applied optics **43** (2004), no. 5, 1053–1062.
- [42] D. Strong and T. Chan, *Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing*, Tech. report, UCLA Math Department CAM Report, 1996.
- [43] D. Strong and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse problems **19** (2003), no. 6, S165.
- [44] P. Sylvain, P. Kornprobst, J. Tumblin, and F. Durand, *Bilateral filtering: theory and applications*, Foundations and trends in computer graphics and vision <sup>®</sup>, vol. 4, Now Publishers Inc., Hanover, MA, 2008.
- [45] E. Tadmor and P. Athavale, *Multiscale image representation using integro-differential equations*, Inverse problems and imaging **3** (2009), no. 4, 693–710.
- [46] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical  $(BV, L^2)$  decompositions*, Multiscale Model. Simul. **2** (2004), no. 4, 554–579.
- [47] A. Tikhonov and V. Arsenin, *Solutions of ill-posed problems*, V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977, Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.
- [48] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, Proc. IEEE ICCV (Bombay, India), 1998, pp. 839–846.
- [49] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Appl. Math. Optim. **44** (2001), no. 2, 131–161.
- [50] Paul Viola and William M Wells III, *Alignment by maximization of mutual information*, Computer Vision, 1995. Proceedings., Fifth International Conference on, IEEE, 1995, pp. 16–23.
- [51] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, *Multi-modal volume registration by maximization of mutual information*, Medical image analysis **1** (1996), no. 1, 35–51.
- [52] F. Wilcoxon, *Individual comparisons by ranking methods*, Biometrics bulletin **1** (1945), no. 6, 80–83.
- [53] R. Xu, P. Athavale, Y. Lu, P. Radau, and G. Wright, *Myocardial segmentation in late-enhancement MR images via registration and propagation of cine contours*, IEEE Int. Symp. Biomed. Imag., 2013, pp. 844–847.
- [54] X. Ye, Y. Chen, W. Lin, and F. Huang, *Fast MR image reconstruction for partially parallel imaging with arbitrary k-space trajectories*, IEEE Transactions on Medical Imaging **30** (2011), no. 3, 575–585.