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

Prashant Athavale[a,b,*], Robert Xu[c,d], Perry Radau[c,1], Adrian Nachman[a,e,f], Graham A. Wright[c,d]

[a]*Department of Mathematics, University of Toronto, Toronto, ON, M5S 1A1, Canada.*
[b]*The Fields Institute, 222 College Street, Toronto, ON, M5T 3J1, Canada.*
[c]*Schulich Heart Program and Physical Sciences Platform, Sunnybrook Research Institute, Toronto, ON, M4N 3M5, Canada.*
[d]*Department of Medical Biophysics, University of Toronto, Toronto, ON, M5G 2M9, Canada.*
[e]*The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, 10 King's College Road, Toronto, ON, M5S 3G4, Canada.*
[f]*Institute of Biomaterials & Biomedical Engineering, University of Toronto, 164 College Street, Toronto, ON, M5S 3G9, Canada.*

Abstract

Images consist of structures of varying scales, large scale structures such as flat regions, and small scale structures such as noise, textures, and rapidly oscillatory patterns. In the hierarchical (BV, L^2) image decomposition, Tadmor et al. (2004) start with extracting coarse scale structures from a given image, and successively extract finer structures from the residuals in each step of the iterative decomposition. We propose to begin instead by extracting the finest structures from the given image and then proceed to extract increasingly coarser structures. In most images, noise could be considered as fine scale structure. Thus, starting the image decomposition with finer scales, rather than large scales, leads to fast denoising. We note that our approach turns out to be equivalent to the nonstationary regularization in Scherzer and Weickert (2000). The continuous limit of this procedure leads to a time-scaled version of total variation flow.

Motivated by specific clinical applications, we introduce an image depending weight in the regularization functional, and study the corresponding the weighted TV flow. We show that the edge-preserving property of the multiscale representation of an input image obtained with the weighted TV flow can be enhanced and localized by appropriate choice of the weight. 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 about denoising speed and the rate of change of energy of the images. An additional contribution of the paper is to use the images obtained at different scales for robust multiscale registration. We show that the inherently multiscale nature of the weighted TV flow improved performance for registration of noisy cardiac MRI images, compared to other methods such as bilateral or Gaussian filtering. A clinical application of the multiscale registration algorithm is also demonstrated for aligning viability assessment magnetic resonance (MR) images from 8 patients with previous myocardial infarctions.

Keywords: Total variation flow, fast denoising, multiscale registration, normalized gradient field, mutual information.

1. Introduction

Most images contain both small-scale structures as well as large-scale structures. In this paper we consider PDE based and variational methods. In medical images, noise usually manifests as a small-scale part of an image. As noise is small-scale structure, most denoising methods can be thought of as methods to decompose images into small-scale structure and large-scale structures.

*Corresponding author

Email addresses: prashantva@gmail.com (Prashant Athavale), robert.xu@sri.utoronto.ca (Robert Xu), perry.radau@gmail.com (Perry Radau), nachman@math.utoronto.ca (Adrian Nachman), gawright@sri.utoronto.ca (Graham A. Wright)

[1]Perry Radau is currently affiliated with Straylight Health Solutions Ltd., 104-451 Rosewell Ave. Toronto, ON, M4R 2H8, Canada.

1.1. Scale-Space Structure of Images with PDE Based Methods

The idea of linear scale-space of images using Gaussian convolution dates back to 1950s (Iijima, 1959; Weickert et al., 1997; Koenderink, 1984; Witkin, 1984). This is equivalent to studying the multiscale representation of a given image $f : \Omega \to \mathbb{R}$ for a bounded region Ω using the heat equation,

$$\frac{\partial u}{\partial t} = \Delta u\,; \;\; u \equiv u(x,t) : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \;\; \left.\frac{\partial u}{\partial \nu}\right|_\Gamma = 0, \tag{1}$$

where $u(\cdot\,,0) := f$ and $\left.\frac{\partial u}{\partial \nu}\right|_\Gamma$ denotes the outward normal to the boundary Γ of the region Ω. This yields a family of images, or the "flow" : $\{u(\cdot\,,t) : \Omega \to \mathbb{R}\}_{t\geq 0}$, which can be viewed as smoothed versions of f. In this linear framework, smoothing is implemented by a convolution with the Gaussian kernel, 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) equation (Perona and Malik, 1988), which is based on *nonlinear diffusion*

$$\frac{\partial u}{\partial t} = \operatorname{div}(\mathcal{P}(|\nabla u|)\nabla u); \;\; u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \;\; \left.\frac{\partial u}{\partial \nu}\right|_\Gamma = 0, \tag{2}$$

with an initial condition $u(\cdot\,,0) := f$. Here, the diffusion controlling function, $\mathcal{P}$ is a monotonically decreasing real-valued function, vanishing at infinity, so that the amount of diffusion decreases as the gradient $|\nabla u|$ increases. Thus, $\mathcal{P}$ is responsible for the anisotropic nature of the PM model. The Perona-Malik equation is closely related to the bilateral filtering method (Tomasi and Manduchi, 1998). In fact, it has been shown to be asymptotically equivalent to bilateral filtering (Sylvain et al., 2008). 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 ζ, whereas $s(f(\zeta), f(x))$ measures the photometric similarity between the pixels at the neighborhood center x and a nearby point ζ, 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, with standard deviations σ_d and σ_r respectively. Bilateral filtering is known to preserve edges and produces a 2-dimensional multiscale representation of the function f, depending on the parameters σ_d and σ_r.

1.2. Scale-Space Structure of Images with Variational Methods

Rudin et al. (1992) introduced a variational framework for denoising an image f which can be formulated as the (BV, L^2) decomposition (Meyer, 2001; Tadmor et al., 2004) of an image f into a regular part u_λ and a oscillatory residual v_λ :

$$f = u_\lambda + v_\lambda, \;\; [\,u_\lambda\,, v_\lambda\,] := \operatorname*{arginf}_{f=u+v} \{\,|\,u\,|_{TV(\Omega)} + \lambda\,\|\,v\,\|^2_{L^2(\Omega)}\}, \tag{3}$$

where $|\,u\,|_{TV(\Omega)} := \int_\Omega |\,\nabla u\,|\,dx$ is the total variation or the BV seminorm of u (see Evans and Gariepy (1992)), and λ determines the relative weighing of the regularization and the fidelity term. This minimization problem has a unique minimizer, $u_\lambda \in BV(\Omega)$ (see Acar and Vogel (1994); Chambolle and Lions (1997); Vese (2001)). The solution to the minimization problem (3) depends on the parameter λ. The dependence of the size of the residual on the parameter λ is given precisely by $\|\,v\,\|_* = \frac{1}{2\lambda}$, where the star-norm, $\|\cdot\|_*$, is defined as the dual of the BV seminorm with respect to the L^2 inner product (cf. Lemma 3, and Definition 11 Meyer (2001)), i.e.

$$\|\,v\,\|_* := \sup_{\substack{\varphi \in BV(\Omega) \\ |\,\varphi\,|_{TV(\Omega)} \neq 0}} \left\{ \frac{\langle\, v\,,\, \varphi\,\rangle_{L^2(\Omega)}}{|\,\varphi\,|_{TV(\Omega)}} \right\}. \tag{4}$$

Thus, the solution u of (3) has different scale for different λ. More precisely, the functional $K : L^2 \to \mathbb{R}$,

$$K(f, \lambda) := \inf_{f=u+v} \{ |u|_{TV(\Omega)} + \lambda \|v\|^2_{L^2(\Omega)} \}$$

dictates the interpolated space $(BV, L^2)_\theta$, where $\theta \in [0, 1]$. It measures how well an L^2 function f is approximated by its BV features. Thus λ dictates the scale of the decomposition $f = u_\lambda + v_\lambda$. The best value of the parameter λ for denoising applications is difficult to determine a priori. Instead, in a seminal paper (Tadmor et al., 2004), the authors proposed to work with multiscale representations of images. While the emphasis in Tadmor et al. (2004) was to progressively extract details of the image, we consider instead a multiscale approach that progressively removes small-scale structures. Furthermore, we extend the results of Tadmor et al. (2004) to continuous multiscale decompositions. We also (in parallel with (Athavale, 2009) and (Tadmor and Athavale, 2009)) 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 map, which can be progressively improved by working at different scales. We were thus able to obtain fast denoising and registration algorithms by considering the continuous limit of multiscale representations. Preliminary versions of some of our results were announced in Athavale et al. (2013).

1.3. Structure of the paper

In Sec. 2 we briefly show the derivation of the weighted TV flow as a limiting case of the hierarchical image decomposition. We then propose a method for designing the parameters in the flow and practical implications related to the speed of the weighted TV flow. In Sec. 3 we propose a method for image denoising using the weighted TV flow, which is also better at preserving edges. The edge-preserving denoising and multiscale nature of the weighted TV flow makes it an ideal candidate as a preprocessing step for hierarchical image registration. We show in Sec. 4 that the the weighted TV flow gives better results, when used in conjunction with a dissimilarity metric like 'the normalized gradient field (NGF)' which is sensitive to the edges. We performed registration experiments with simulated data in Sec.4.3 and clinical data in Sec. 4.6 and show that the weighted TV flow produces significantly better results than Gaussian filtering or Bilateral filtering when NGF is used as the dissimilarity metric. We discuss the results of each of these experiments in the discussion Sec. 5. We also describe several interesting mathematical properties of the weighted TV flow in the Appendix (Sec. 7).

2. Mathematical Theory of the Total Variation (TV) Flow and its Relationship to Multiscale Decomposition

It is very helpful to consider more general decompositions than (3), with the BV seminorm replaced by a convex functional $\mathcal{J}(u)$ which will allow the flexibility of inhomogeneous weighting:

$$f = u_\lambda + v_\lambda, \quad [\, u_\lambda \,, v_\lambda \,] := \operatorname*{arginf}_{f=u+v} \{ \mathcal{J}(u) + \lambda \|v\|^2_{L^2} \}. \tag{5}$$

Motivated by (Tadmor et al., 2004) we start by decomposing the original f using the parameter λ_1 to obtain the initial decomposition, $f = u_{\lambda_1} + v_{\lambda_1}$. However, we choose the initial parameter λ_1 large enough so that u_{λ_1} preserves most of the features of f. Using another scaling parameter $\lambda_2 < \lambda_1$ we obtain the decomposition of u_{λ_1} as $u_{\lambda_1} = u_{\lambda_2} + v_{\lambda_2}$. If we continue this process iteratively, each time decomposing the u_{λ_i} part with $\lambda_i < \lambda_{i-1}$, we obtain the following nonlinear multiscale decomposition

$$\begin{aligned} f &= u_{\lambda_1} + v_{\lambda_1} \\ &= u_{\lambda_2} + v_{\lambda_2} + v_{\lambda_1} \\ &= \dots \\ &= u_{\lambda_N} + \sum_{i=1}^{N} v_{\lambda_i}. \end{aligned} \tag{6}$$

The idea is to terminate the iteration at the N^{th} step, when the cumulative residual, $\sum_{i=1}^{N} v_{\lambda_i}$, reaches a desired level. The choice of N is critical for successful denoising applications. We will return to the issue of a stopping criterion when we work with the corresponding continuous model in Sec. 3.1. While we are motivated by Tadmor et al. (2004) the BV components u_{λ_i} in the above multiscale decomposition turns out to be identical to those produced by iterated non-stationary regularization, (see Eq. (1) in Scherzer and Weickert (2000)).

2.1. Derivation of the Weighted Total Variation Flow

The continuous limit of iterated regularization, for $\lambda_i = \frac{1}{2}$ is shown to lead to diffusion filtering via TV flow, see introduction in Burger et al. (2007). We sketch below the derivation to see: (a) how the choice of the scaling parameters correspond to the speed of the denoising process (see Sec. 3.4.1 and Proposition 1), and (b) how we can control the speed locally by introduction of spatially varying weight in Sec. 2.1. The Euler-Lagrange differential equation for (5) is

$$0 \in \partial \mathcal{J}(u_\lambda) + 2\lambda\,(u_\lambda - f). \tag{7}$$

For a lower semicontinuous, proper, convex functional $\mathcal{J}$ the subgradient $\partial\mathcal{J}(u) = A(u)$, is a maximal monotone operator (Brezis, 1983), and we can rewrite (7) as follows

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

Going towards the continuous model, in the hierarchical decomposition (6), we replace the scale parameter sequence $\{\lambda_i\}$ with $\{\frac{\lambda_i}{\tau}\}$ (cf. Tadmor and Athavale (2009)), where τ denotes the time-step. With this notation (6) can be written as

$$u_{\lambda_N} = \left[\prod_{i=1}^{N}\Big(\mathrm{Id} + \frac{\tau}{2\lambda_i}A\Big)^{-1}\right] f. \tag{8}$$

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 Crandall and Liggett (1971), Theorem A) that the limit

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

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

$$\frac{\partial u}{\partial t} = -\frac{1}{2\lambda(t)}A(u),\;\; u(\,\cdot\,,\,0\,) = f,\;\; \frac{\partial u}{\partial \nu}\Big|_\Gamma = 0. \tag{9}$$

While the derivation above is quite general, we will work with a functional $\mathcal{J}(u)$ that is suitable for image processing. The choice of $\mathcal{J}(u)$ as the weighted bounded variation (BV) seminorm was previously considered in Strong and Chan (1996, 2003) for its edge-preserving properties. Properties of the weighted BV spaces and the corresponding minimization with the L^2 fidelity term were investigated in (Chen and Rao, 2003). Motivated by the edge-preserving properties of the weighted TV minimization, we consider the weighted BV seminorm,

$$\mathcal{J}(u) \equiv |\,u\,|_{TV_\alpha(\Omega)} := \int_\Omega \alpha\,|\,\nabla u\,|\,dx,$$

where $\alpha \equiv \alpha(x)$ is a positive, continuous function in Ω. For this choice of $\mathcal{J}$ we have $-\operatorname{div}\left(\frac{\alpha\nabla u}{|\nabla u|}\right) \in \partial\mathcal{J}(u)$. Following (Anzellotti, 1983), this is understood in the sense that for any $u \in BV_\alpha(\Omega)$ there exists a bounded vector field denoted by $\frac{\nabla u}{|\nabla u|}$ such that

$$\int_\Omega \alpha\,|\,\nabla u\,|\,dx = \int_\Omega \frac{\alpha\nabla u}{|\,\nabla u\,|}\cdot\nabla u\,dx.$$

With $\mathcal{J}(u) = |u|_{TV_\alpha(\Omega)}$ we say that the decomposition (5) is a (BV_α, L^2) decomposition and, with $A(u) = -\operatorname{div}\left(\frac{\alpha\nabla u}{|\nabla u|}\right)$ in (9), leads to the weighted TV flow

$$\frac{\partial u}{\partial t} = \mu(t)\operatorname{div}\left(\frac{\alpha\,\nabla u(\cdot\,,t)}{|\,\nabla u(\cdot\,,t)\,|}\right); \tag{10}$$
$$u : \Omega\times\mathbb{R}_+ \mapsto \mathbb{R},\ \left.\frac{\partial u}{\partial \nu}\right|_\Gamma = 0,\ \ u(\cdot\,,0) = f,$$

where we write $\mu(t) := \frac{1}{2\lambda(t)}$ for convenience. Note that, the weighted TV flow (10) is closely related to the standard TV flow

$$\frac{\partial \hat{u}}{\partial t} = \operatorname{div}\left(\frac{\nabla\hat{u}(\cdot\,,t)}{|\nabla\hat{u}(\cdot\,,t)|}\right), \tag{11}$$

which has been extensively studied and is known to have a unique solution (see Andreu et al. (2001, 2002, 2004)). Indeed, if $\alpha(x) \equiv 1$ and $\gamma(t) = \int_0^t \mu(s)\,ds$, then $u(\cdot\,,t) = \hat{u}(\cdot\,,\gamma(t))$. See Montillo et al. (2003); Jonasson et al. (2007) for applications of TV flow in medical imaging.

Remark 1. *We see from* (10) *that the diffusion co-efficient at* (x,t) *is* $\mu(t)\alpha(x)$*. Thus, the weight* $\alpha(x)$ *allows us to selectively change the variable in clinical applications.*

2.2. Designing the Edge-preserving Weighted TV Filter

The method based on the weighted functional has the flexibility to allow for weights α chosen according to the image processing problem at hand. If the objective is to preserve edges, then the weight α 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

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

where σ denotes the standard deviation of the Gaussian kernel G_σ, and β denotes a constant parameter. Such edge detecting functions (Aubert and Kornprobst, 2006) are common in the image processing literature.

2.2.1. Choice of the parameter β

As shown in Fig.1, the value of the parameter β dictates the shape of the function $\frac{1}{\sqrt{1+\frac{J^2}{\beta^2}}}$, thus deciding the edge-preservation at the jump discontinuity of magnitude J. Lower value of β indicates more control of diffusion at edges, and vice versa. As the value of the parameter β increases, the weight $\alpha(x)$ approaches unity and the weighted TV flow (10) behaves akin to the standard TV flow (11). The value of β is chosen so that the weight $\alpha(x_0)$ attains a small value at the most prominent edge at x_0 in the image domain. To design the edge-preserving weighted TV flow, we need the weight $\alpha(x_0)$ to be a small positive value, say α_0. To achieve this, we need

$$\alpha_0 = \frac{1}{\sqrt{1+\frac{\max_x|\nabla(G_\sigma * f)(x)|^2}{\beta^2}}}.$$

We observe that for $f \in [0,1]$, the maximum possible jump in the image (i.e. $\max_x|\nabla(G_\sigma * f)(x)|$) can not exceed 1 (the magnitude of the jump, $|J| = 1$ if the image has an edge where the intensity changes from zero to the maximum value i.e. 1, or from the maximum value of 1 to zero). Thus, we set β such that

$$\alpha_0 = \frac{1}{\sqrt{1+\frac{1}{\beta^2}}},$$

and as $\alpha_0^2 \ll 1$, we get $\beta \approx \alpha_0$. In this paper, the value of β was set at 0.05, which corresponds to 5% regularization at the most prominent edge in the image.

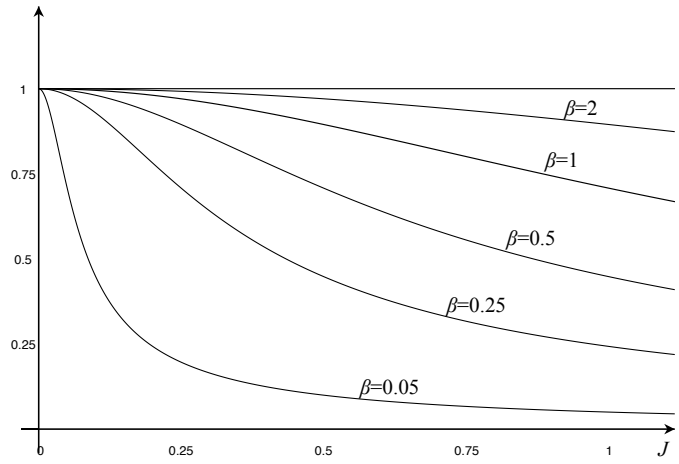


Figure 1: The graphs of the function $\frac{1}{\sqrt{1+\frac{J^2}{\beta^2}}}$. We see that the value of β dictates diffusion at the jump discontinuity with magnitude J.

2.2.2. Choice of the speed function $\mu(t)$

We define the $\alpha*$ norm, $\|\cdot\|_{\alpha*}$, as a dual of the weighted BV-seminorm with respect to the L^2-inner product.

$$\| v \|_{\alpha*} := \sup_{\substack{\varphi \in BV_\alpha(\Omega) \\ |\varphi|_{TV_\alpha(\Omega)} \neq 0}} \left\{ \frac{\langle v , \varphi \rangle_{L^2(\Omega)}}{|\varphi|_{TV_\alpha(\Omega)}} \right\}. \tag{12}$$

We show in section 7 (Proposition 1) that the speed of the weighted TV flow, measured in the $\|\cdot\|_{\alpha*}$ is exactly equal to the speed function $\mu(t)$. Thus, a monotone increasing speed function $\mu(t)$ will progressively increase the speed of the TV flow. In this paper the speed function is chosen as $\mu(t) = 1.1^t$ unless specified otherwise. We further elaborate on this property in section 3.4.

2.3. Comparison Between Weighted TV Flow and Standard TV Flow

In this section we compare the edge-preserving property of the weighted TV flow (10) and the standard TV flow (11). The weight $\alpha(x)$ in (10) can be thought of as a local adaptivity control (cf. Strong and Chan (2003)). With an appropriate choice of the weight $\alpha(x)$ we can improve the edge preservation. To demonstrate this property we performed denoising of an MR image with a myocardial infarct. The image was denoised using the standard TV flow and the weighted TV flow, with the same stopping criteria (estimated peak signal-to-noise ratio[2] (PSNR) = 30 dB). The corresponding denoised images are shown in Fig. 2 (b), and (c). In Fig. 2(d) we see the line profile corresponding to the vertical line in Fig. 2(a) that passes through the infarct. In Fig. 2(e) we see the line profiles of the denoised images corresponding to a vertical line in Fig. 2(a) through the infarct. We see that the intensity level corresponding to the infarct is significantly reduced by the standard TV flow.

To solve (10) and (11) numerically, we used a semi-implicit finite difference scheme (Tadmor et al., 2004) to discretize the divergence term in (10) (see also Michel et al. (2011); Ramani and Fessler (2012); Borsic et al. (2010)). The numerical scheme used to solve (10) is described in Appendix 8.

3. Fast and Edge-Preserving Denoising Experiments

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

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

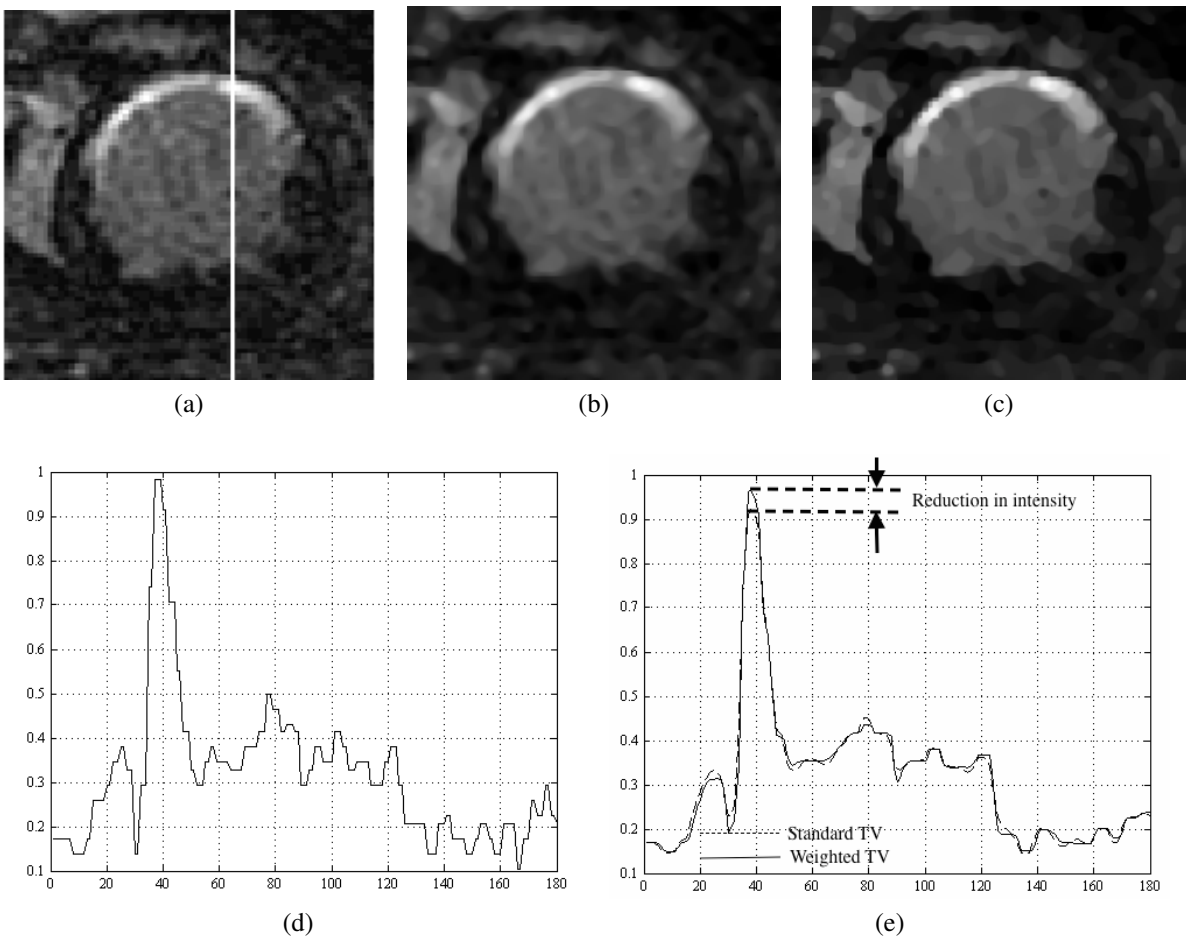
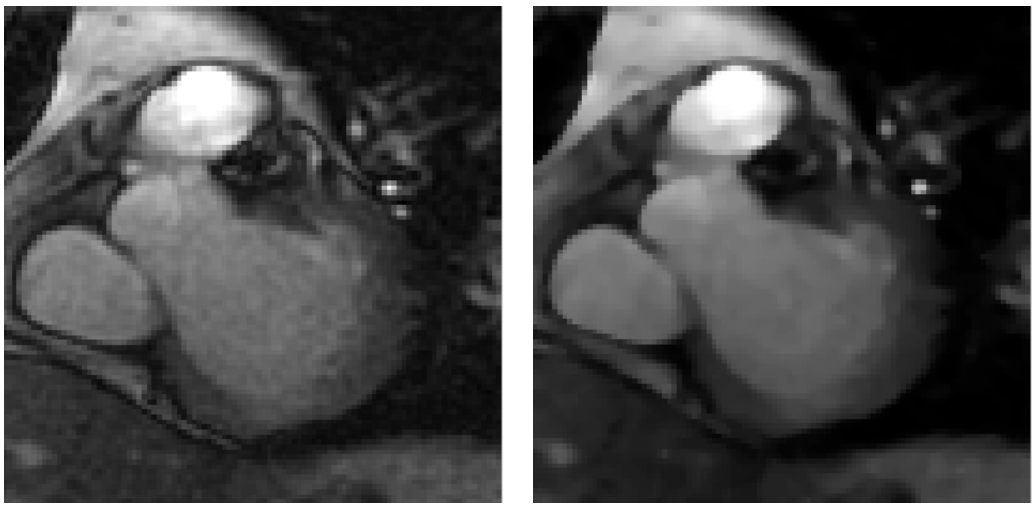


Figure 2: (a) Original MR image with an infarct. (b) Denoised using the standard TV flow, (c) denoised using the weighted TV flow. The line profile of the original and denoised images are shown in Fig. 2. (d) Line profile through infarct in the original image in (a) corresponding to the vertical line. (e) Denoising result of standard TV flow and weighted TV flow superimposed for the vertical line shown in (a). 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.

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 multiscale weighted TV flow (10) 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 (10) can be effectively applied for denoising an image, while edge preservation is enhanced due to the weight $\alpha(x)$.

3.1. 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 4) 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 resulting function, $u(t_c)$, provides us with a denoised version of the given image f. We also observe in Proposition 5 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 5).



(a) (b)

Figure 3: (a) Original cardiac image, (b) denoised with weighted TV flow, stopping criteria: PSNR = 35dB, computational time = 13 ms.

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, PSNR($u(t), f$) falls below the estimated PSNR (see Gudbjartsson and Patz (1995); Coupé et al. (2010) for noise estimation), thus making 'estimated PSNR' as the only denoising parameter. We chose the PSNR as the measure, because it is a direct measure of the extracted noise variance.

Remark 2. *We also considered using a structural similarity index metric (Wang et al., 2004) as a stopping criteria, but due to the lack of ground truth in denoising applications we chose PSNR as the similarity measure.*

The denoising application of the weighted TV flow is demonstrated in Fig. 3 for a MR image of a left ventricle of a human heart. We can see that the denoising is performed without significant blurring of edges.

3.2. Computation of the Stopping PSNR & Denoising Time for Rician Noise

In this section we demonstrate the computation of the stopping criterion i.e. computation of the estimated PSNR used to terminate the weighted TV flow for denoising MR images. For MR images we know that the image intensity is zero in regions of with vacuum or air, such as chest cavity. One can estimate the PSNR using these regions. We demonstrate this using the Shepp-Logan phantom (256×256 pixels). Rician noise, was added to the image for varying $\sigma = 0.04, 0.06, 0.08, 0.1$, and $\nu = 0.005$. The stopping criterion i.e. estimated PSNR was computed using the top left corner of the noisy image of size 50×50 pixels. More precisely, estimated PSNR = PSNR(f_b, z_b), where f_b indicates the upper-left corner of the noisy image, and z_b indicates the zero-image with the same size. For each value of σ, we performed 100 simulations. We report the average values of time taken to estimate the "stopping PSNR", as well as "denoising time" in the Table 1. (All computations in this section and the rest of the paper are performed using MATLAB 2010b, on Macintosh OS X version 10.9.5, with 2.3 GHz Intel Core i5, 4 GB. The speed function for the weighted TV flow was chosen as $\mu(t) = 1.1^t$, with $\beta = 0.05$.) The time taken to compute the estimated PSNR is about 0.8 ms, whereas the actual denoising with the weighted TV flow using the estimated PSNR is of the order of a few hundred ms. As expected, the denoising time decreases with the decrease in noise, whereas the time to compute the stopping criterion depends only on the size of the region used in the estimation. The corresponding noisy and denoised images are shown in Fig. 4. for various values of σ.

Table 1: Computational time for the stopping criterion, and the actual denoising time for Rician noise of varying σ. The averages are computed for 100 simulations for each σ.

σ	Average* estimated PSNR (dB)	Average* time taken to compute Estimated PSNR (ms)	Average* denoising time using weighted TV flow (ms)
0.04	30.64	0.762	149.3
0.06	27.28	0.837	236.4
0.08	24.83	0.821	344.0
0.10	22.94	0.853	485.2

(*Averages are computed for 100 simulations.)

3.3. Effect of inaccurate stopping criteria

In clinical setting, the ground truth about the noise is often unknown. In this section we present experiments with inaccurate estimated PSNR. We performed experiments with additive Rician noise, Rice(ν, σ), with $\sigma = 0.1$, $\nu = 0.05$. We computed the estimated PSNR = P_0 using the method described in 3.2. Then, instead of using this P_0 as a stopping criterion, we used $0.99P_0$, $0.98P_0$, $0.97P_0$, $0.96P_0$, and $0.95P_0$, as the stopping criteria. For these errors the mean percentage reduction in SSIM values were observed to be 2.42%, 4.12%, 5.64%, 6.03%, 7.35% respectively. The means were computed for 50 simulations of the noise.

3.4. Experiments Relating Denoising Speed and $\mu(t)$

In this section we show that the selection of the speed function $\mu(t)$ in (10) determines the speed of denoising. We show in Proposition 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.

3.4.1. Results Experiments Relating Denoising Speed and $\mu(t)$

The results of the experiments with different speed functions are summarized in Table 2. We note that the speed

Table 2: 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
10^{1+t}	23.91	87

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 PSNR($u(t), f$) < 25 dB.

3.5. Experiments comparing Edge Preservation properties of 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 $\alpha(x)$ that we chose in Sec. 2.2 is similar in nature to a function used by the Perona-Malik model (Perona and Malik, 1988; Catté et al., 1992), except that in case of weighted TV flow $\alpha(x)$ is independent of u. We found that the Perona-Malik filter performs qualitatively similar to the bilateral filter. In fact, it is shown in (Sylvain et al., 2008) that the bilateral filter is asymptotically equivalent to Perona-Malik filter. Thus, in this paper we chose to compare our method with bilateral filtering.

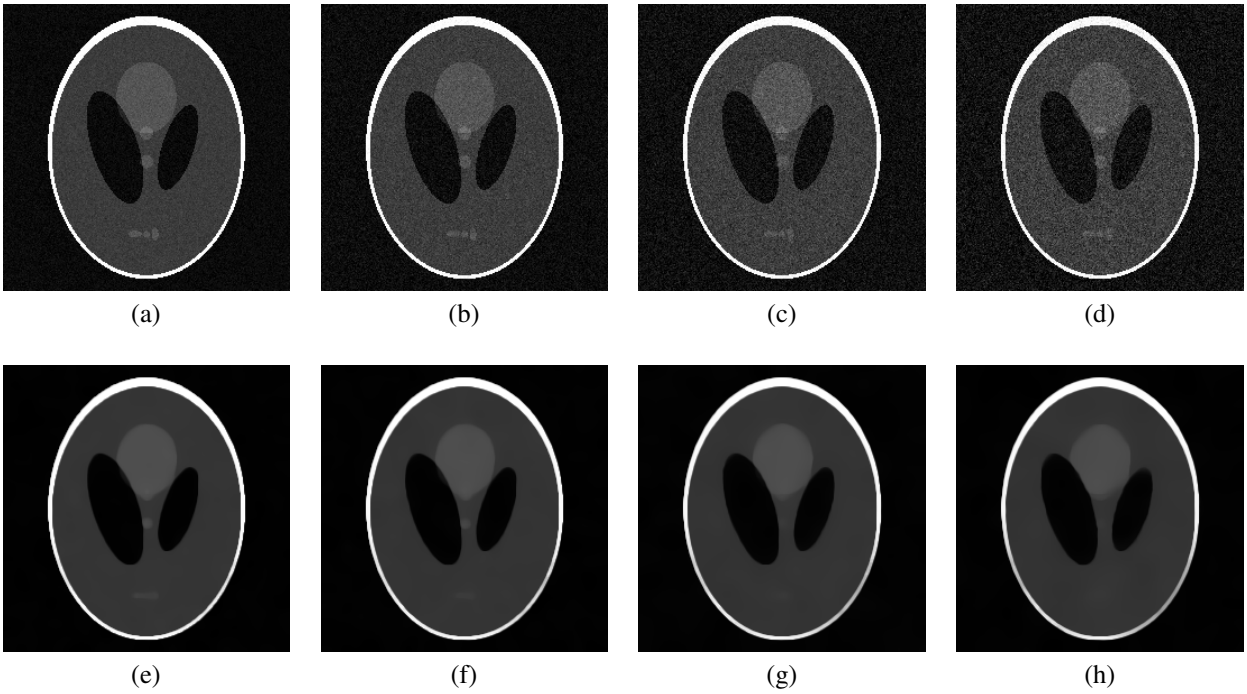


Figure 4: Images (a) to (d) show noisy images with additive Rician noise Rice(ν, σ), with $\sigma = 0.04, 0.06, 0.08, 0.10$ respectively. Images (e) to (h) depict the corresponding denoised images.

We performed denoising experiments with the Shepp-Logan phantom, with an additive Gaussian noise of zero mean and varying standard deviations σ. The estimated PSNR 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, σ_1 and σ_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) of the denoised images (Song et al., 2004) with respect to the clean Shepp-Logan phantom image. The CNR_{AB} is computed for the regions A and B, annotated in Fig. 5(a) with the formula:

$$\mathrm{CNR}_{AB} = \frac{|\mu_A - \mu_B|}{\sigma},$$

where μ_A, μ_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 σ_A, σ_B indicating the standard deviations in regions A and B respectively. The experiment was repeated 50 times for each variance.

3.5.1. Result of Experiments comparing Edge Preservation Properties of Weighted TV Flow and Bilateral Filter

The average values of the CNR for bilateral filter and the weighted TV flow are shown in Table 3. We observe that the weighted TV flow yields higher CNR images than bilateral filtering. We also compute the structural similarity index metric (SSIM) (Wang et al., 2004) of denoised images with respect to the original image.

Remark 3. *The parameters of the bilateral filter were tuned for this study so that both the results have the same PSNR to establish an equivalent denoising level as was achieved with the weighted TV flow. No such fine tuning is needed for the weighted TV flow, as the stopping criteria is itself in terms of the noise measure (see Sec. 3.2).*

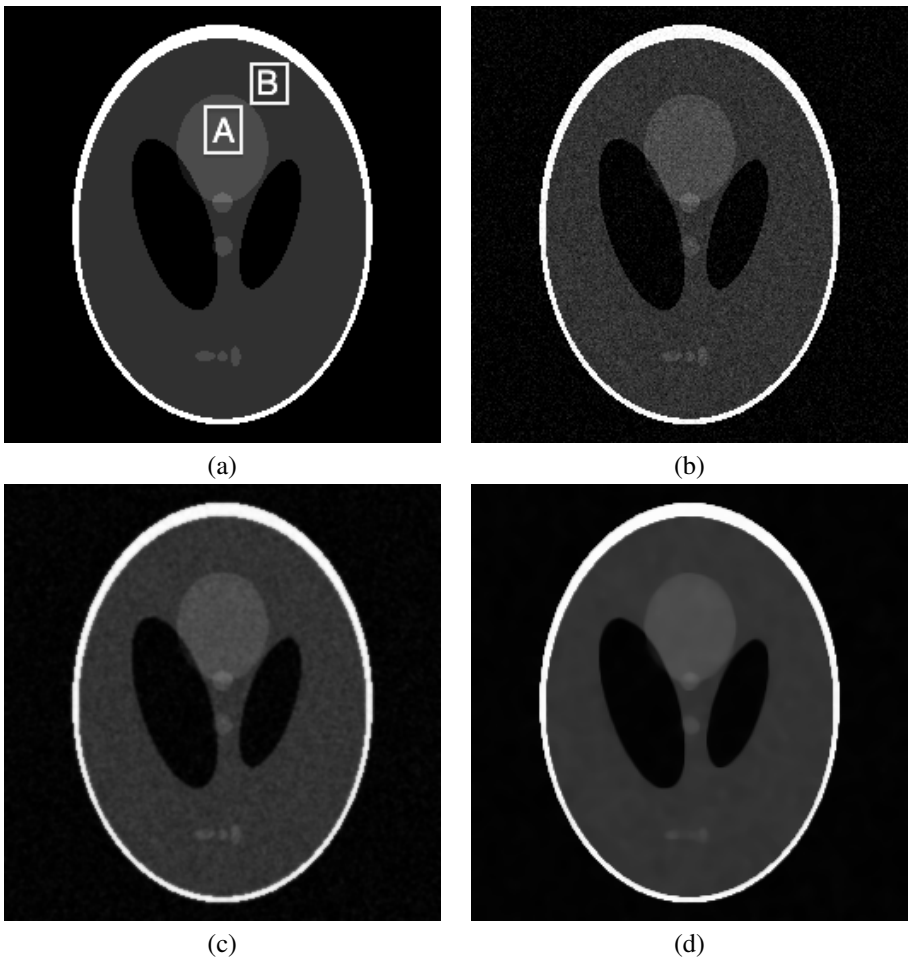


Figure 5: Comparison between the bilateral filter and the proposed weighted TV flow in denoising. (a) Shepp-Logan phantom, (b) noisy image, $\sigma^2 = 0.002$, (c) denoising with bilateral filter (average $\text{CNR}_{AB} = 6.62$), (d) denoising with weighted TV flow (average $\text{CNR}_{AB} = 21.08$).

Table 3: Denoising results for the Shepp-Logan phantom

noise* (σ^2)	average** CNR_{AB}		average** SSIM		average** denoising time bilateral (ms)	average** denoising time TV_α flow (ms)
	bilateral	TV_α flow	bilateral	TV_α flow		
0.0008	8.50	24.42	0.72	0.81	1401	163
0.0010	8.22	23.17	0.70	0.78	1402	186
0.0020	6.62	21.08	0.61	0.69	1414	292
0.0025	6.33	21.04	0.58	0.66	1402	327
0.0050	5.34	19.40	0.43	0.58	1402	582

(*The estimated PSNR=$10\log(1/\sigma^2)$ dB. **The average is computed for 50 experiments for each noise variance.)

Remark 4. *We see that the denoising time for the bilateral filter is around 1400 ms, whereas the denoising time for the weighted TV flow is much less. The denoising time for the bilateral filter depends on the size of the image. Thus, even for small amount of noise the denoising time is unchanged.*

4. Applications of the Multiscale Weighted TV Flow to Image Registration

We have seen in section 2 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 maps are usually obtained in a hierarchical manner, where an initial registration map is obtained by comparing coarse scales from given images; this map is then used as an initial guess to register finer scales. In this approach 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. 2 is especially useful in techniques such as normalized gradient field (NGF) based registration (Haber and Modersitzki, 2006). The NGF based approach was proposed as an alternative to the mutual information (MI) based registration (Wells et al., 1996). 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. We will show in Sec. 4.4 that the multiscale approach with weighted TV flow produces better results when used in NGF based technique, compared to other filtering techniques such as Gaussian, or bilateral filtering (Fig. 9). 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.

4.1. 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:

$$\min_{w} \{\mathcal{D}(f[w], g)\}, \tag{13}$$

where $w : \mathbb{R}^n \to \mathbb{R}^n$ represents the geometric transformation, and $\mathcal{D}$ represents a dissimilarity measure between the transformed template image $f[w] := f(w(x))$ and the reference image g. Possible distance measures $\mathcal{D}$ include sum of squared differences (SSD), cross correlation, mutual information (MI) (Viola and Wells III, 1995; Collignon et al., 1995), and normalized gradient fields (NGF) (Haber and Modersitzki, 2006).

Various papers (Pluim et al., 2003; Fei et al., 2002; Carrillo et al., 2000; Holden et al., 2000; Rueckert et al., 1999) have shown the use of mutual information as an effective similarity measure. 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 (Wells et al., 1996), where matching of corresponding intensity patterns is more important than the actual intensity values. Nevertheless, there are well-known problems with this approach (Haber and Modersitzki, 2006), such as non-convexity of the registration problem and discrete structure of the mutual information. An alternative dissimilarity metric was suggested in (Haber and Modersitzki, 2006), which makes use the NGF metric, defined as follows:

$$\mathcal{D}_{NGF}(f, g) = -\int_{\Omega} \langle \vec{v}(f, x), \vec{v}(g, x) \rangle_{l^2}^2 dx, \tag{14}$$

where $\langle \cdot, \cdot \rangle_{l^2}$ denotes the usual vectorial dot product, and $\vec{v}$ represents the normalized gradient of the image,

$$\vec{v}(I, x) := \frac{\nabla I(x)}{\| \nabla I(x) \|_{\rho}}, \tag{15}$$

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

$$\begin{aligned} \| \vec{\xi} \|_{\rho} &= \sqrt{\sum_{i=1}^{n} \xi_i^2 + \rho^2}, \\ \rho &:= \frac{\eta}{|\Omega|} \int_{\Omega} |\nabla I(x)| \, dx, \end{aligned} \tag{16}$$

where η 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. We note that the NGF based approach depends on the faithful preservation of prominent edges, which are often lost due to noise. This is why the weighted TV with its edge-preserving property is useful as elaborated in Sec. 4.4.

4.2. Hierarchical Registration Framework and Methods

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, a multiscale registration approach is used (Modersitzki, 2004). 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 (Tomasi and Manduchi, 1998), 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. As we have shown, 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 4, 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. 4.2 (see (Li et al., 2012; Michálek et al., 2011) 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. With this notation we describe the steps in the hierarchical registration framework below:

(a) The initial unregistered images f and g are inputs to the registration framework along with an initial guess $\omega^{(0)}$ for the registration map.
(b) We obtain the weighted TV flow images $u^f(\,\cdot\,,\, t_i)$ and $u^g(\,\cdot\,,\, t_i)$ at discrete time intervals $t_N < t_{N-1} < \cdots < t_0$ for a fixed integer N.
(c) As t_0 is the largest value amongst the sequence $\{t_i\}_{i=0}^N$, the images $u^f(\,\cdot\,,\, t_0)$ and $u^g(\,\cdot\,,\, t_0)$ are coarsest scale images. One can either register these images directly or register a suitably downsampled versions of $u^f(\,\cdot\,,\, t_0)$ and $u^g(\,\cdot\,,\, t_0)$. We choose to downsample the images dyadically, i.e. we downsample the images $u^f(\,\cdot\,,\, t_i)$ and $u^g(\,\cdot\,,\, t_i)$ by a factor of 2^{N-i}.
(d) The registration map $\omega^{(i+1)}$ obtained after registering the down sampled versions of $u^f(\,\cdot\,,\, t_i)$ and $u^g(\,\cdot\,,\, t_i)$ is used as an initial guess for the registration of the downsampled finer scale images $u^f(\,\cdot\,,\, t_{i+1})$ and $u^g(\,\cdot\,,\, t_{i+1})$.
(e) This process is repeated, and until $\omega_{final} := \omega^{(N+1)}$ is obtained at the finest level.

4.3. Registration Experiments with Simulated Data

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 4.2 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. 7.

Subsequently, we generated a series of known rigid transformations of g_0 consisting of rotation θ°, 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. 7 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 θ, t_{x_1}, and t_{x_2} are randomly generated with independent and identically uniform distributions.

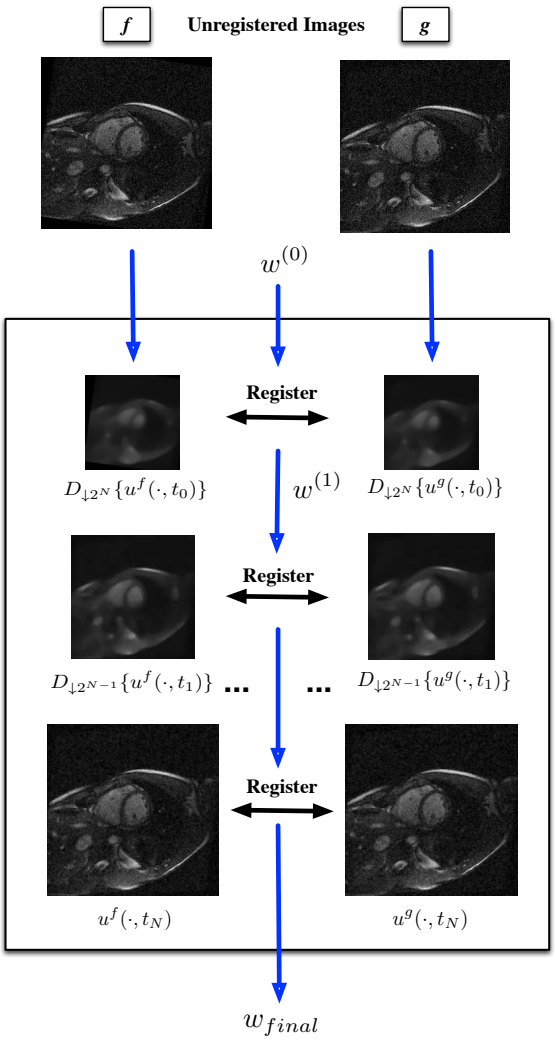


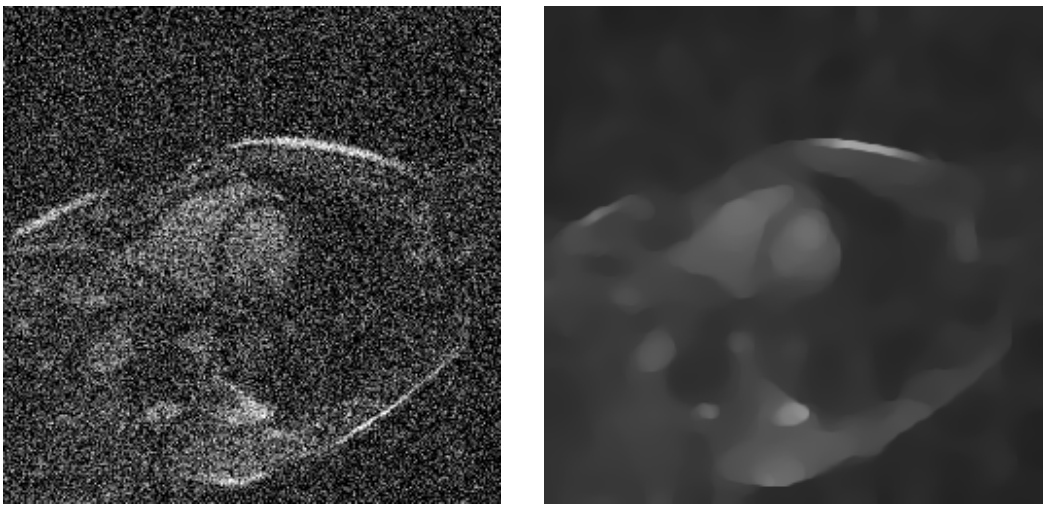
Figure 6: Multiscale registration framework is demonstrated. Original input images are down-sampled and registered at lower scales, before the estimated registration parameters are propagated as the initial guess to images at higher resolutions. $\omega^{(0)}$ represents the initial guess of the optimal registration parameters, and ω_{final} corresponds to the final estimate of the registration parameters.

To simulate low signal-to-noise ratio (SNR) conditions, Rician noise $R \sim \text{Rice}(\nu, \sigma)$ 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. 8 (a). The transformed and noisy image f serves as the template image in the multiscale registration framework 4.2, where we solve the minimization problem (13) to obtain the optimal transformation ω_{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.

4.4. Registration Results with Simulated Data

The results of the hierarchical registration using the NGF based metric (14) are shown in Fig. 9, where the target registration errors are presented for four different hierarchical registration approaches. Specifically, from left to right, each group of TREs corresponds to hierarchical registration using NGF with preprocessing using Gaussian filtering, bilateral filtering, TV flow, and weighted TV flow to smooth the input images respectively. Within each group, 4 box plots are shown, and each box plot shows the median and spread of TRE for all 10 volunteer datasets after registration

Figure 7: Manually delineated landmarks in the left and right ventricles, as indicated by the '×' markings.



(a) Noisy image (b) Large scale structure extracted with TV_α flow

Figure 8: A low SNR image f and a large scale structure obtained using the weighted TV flow smoothed image u^f are shown in (a) and (b) respectively.

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.

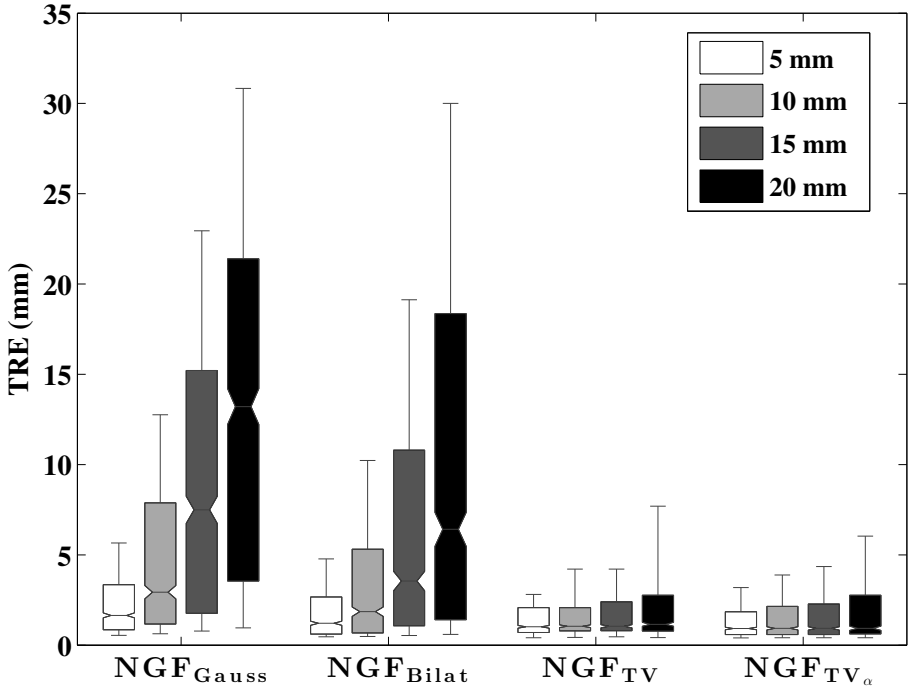


Figure 9: Registration results are shown for different preprocessing techniques. NGF_{Gauss}, NGF_{Bilat}, NGF_{TV} and NGF_{TV_α} represent Gaussian filtering, bilateral filtering, TV flow filtering and weighted TV flow filtering for each scale of the hierarchical registrations respectively. Each box plot illustrates the median (notch), interquartile range (box), and the 10th/90th percentiles (whiskers) of the TRE for all 10 volunteer datasets.

It was observed that the best results were achieved using weighted TV flow for preprocessing. The Wilcoxon signed-rank test (Wilcoxon, 1945) was performed with significance set at the 1% level (i.e. the null hypothesis is rejected if $p < 0.01$). Specifically, the TRE results using each denoising method were compared in turn to the weighted TV flow (NGF_{TV_α}) method. In each test, the null hypothesis stated that the NGF_{TV_α} method and the other preprocessing method performed equally well; whereas, the alternative hypothesis stated that NGF_{TV_α} performed better by achieving lower TRE. It was shown at the 1% significance level that the NGF_{TV_α} method achieved better registration accuracy than NGF_{Gauss}, NGF_{Bilat} and NGF_{TV}, with $p < 0.01$ in all cases. Furthermore, in this experiment, TV flow and weighted TV flow preprocessing are the only methods that achieved a median TRE of less than 5 mm for all four initial displacement categories. However, the 90th percentile TRE for the weighted TV flow is still lower than the TV flow preprocessing method for the 20 mm initial displacement category.

4.5. Registration Experiments with Clinical Data

A clinical application of the proposed registration framework is also demonstrated in this study for aligning MR data with different contrast for patients with previous myocardial infarctions. Two imaging datasets were acquired for each patient. First, a cine cardiac magnetic resonance (CMR) imaging acquisition was used to obtain a multi-slice (10-15 slices) stack of images for evaluating the wall motion of the left ventricle. Then, a late gadolinium enhancement (LGE) MR scan was performed in the same imaging session to acquire a corresponding stack of images, which show the hyper enhanced infarct region (Pennell et al., 2004). The location and size of the infarct region are important factors in viability assessment, as they determine the feasibility and value of revascularization.

To verify the location and size of the infarct, an initial contouring step is usually required to delineate both the endocardial and epicardial borders of the myocardium in the LGE images. One approach to this problem is to segment the cine MR images and then propagate the cine contours to the LGE dataset via image registration (Xu et al., 2013).

This strategy is taken because it is often difficult to segment the LGE images directly, due to intensity inhomogeneity in the myocardium caused by the accumulation of contrast agent. Image registration is required in this approach to correct for potential misalignments between the cine and LGE datasets due to patient or respiratory motions.

In this experiment, an expert manually delineated the endocardial and epicardial borders from 8 patient cine datasets as well as the the corresponding LGE datasets. Then, we applied the proposed hierarchical registration framework to align the cine and LGE images, where each individual imaging slice in the cine and LGE datasets were preprocessed by the weighted TV flow filter. Subsequently, the 3D cine image, consisting of a multi-slice stack, was registered to each 2D LGE imaging slice via a transformation $\omega := \langle t_x, t_y, t_z, \theta_x, \theta_y, \theta_z \rangle$, which involves translations and rotations along the x, y, z imaging axes. This accounts for potential through-plane movement, as well as in-plane movement due to patient or respiratory motion. The obtained transformation ω is subsequently applied to the cine contours, which are propagated and overlaid onto the LGE datasets. To evaluate the registration accuracy, the propagated contours are compared to the manually delineated ground truth LGE contours in terms of the Dice similarity coefficient (DSC) overlap and average perpendicular distance (APD) between the contours for both the endocardial and epicardial borders.

4.6. Registration Results with Clinical Data

Typical registration results using the proposed hierarchical registration framework with weighted TV flow and NGF are illustrated in Fig. 10. The first column illustrates the expert delineated endocardial and epicardial borders on the cine dataset. Second column shows the direct transfer of the cine contours onto the LGE dataset without image registration; notice the arrows that are pointing to the erroneous contour locations at the septal and anterior wall of the LV due to misalignment between the cine and LGE datasets. The third column shows the motion corrected endocardial and epicardial contours after registration between the cine and LGE datasets using the proposed hierarchical registration framework with multiscale weighted TV flow and NGF. Finally, the manually delineated contours for the LGE images are shown in the right most column.

Overall, the quantitative registration results using the clinical viability assessment data are summarized in Fig. 11. Specifically, the registration accuracy in terms of DSC and APD are shown for 4 different registration schemes, as well as a baseline reference, where no registration is applied. In the reference scenario (i.e. *NoReg*), the segmented cine contours are directly overlaid on top of the LGE images without any motion correction. As expected, there is a relatively large discrepancy between these overlaid contours and the manually delineated LGE contours, as indicated by low DSC values and high APD (see Fig. 11). On the other hand, the best registration results were achieved via the proposed multiscale weighted TV flow registration framework with the NGF similarity metric (NGF_{TV_α}). This approach achieved median DSC values of 94.7% and 96.6% for the endocardial and epicardial contour overlaps respectively. Similarly, the proposed approach achieved median APD values of 1.11 mm and 0.94 mm for endo- and epicardial contours respectively. Furthermore, for each registration accuracy criterion, the Wilcoxon signed-rank test showed that the NGF_{TV_α} approach outperformed the weighted TV flow and MI (MI_{TV_α}) approach, the NGF_{Bilat} approach, and the NGF_{TV} approach with $p < 0.05$ in all cases.

5. Discussion

5.1. Discussion of Denoising with Weighted TV flow as compared to Denoising with standard TV flow

We compared the results of weighted TV and the standard TV flow in Fig. 2. We observe that the edges are better preserved with the weighted TV flow compared to the standard TV flow. We also see in Fig. 2(e) that the intensity of the peak is better preserved in the case of weighted TV flow in comparison to the TV flow. This feature is important MR images as the intensity is related to the tissue type (Heindel et al., 1994; Detsky et al., 2009), and changes in intensity may result in tissue mischaracterization.

5.2. Discussion of Fast Denoising Experiments

We showed in Proposition 1 that the speed of the weighted TV flow measured in the $\alpha*$ norm is completely determined by the speed function $\mu(t)$. This property is demonstrated in Sec. 3.4.1, where in Table 2 we see that the denoising speed increases as $\mu(t)$ increases, and therefore the filtering speed can be easily adjusted. It should be noted that increasing speed decreases the accuracy of the stopping criterion.

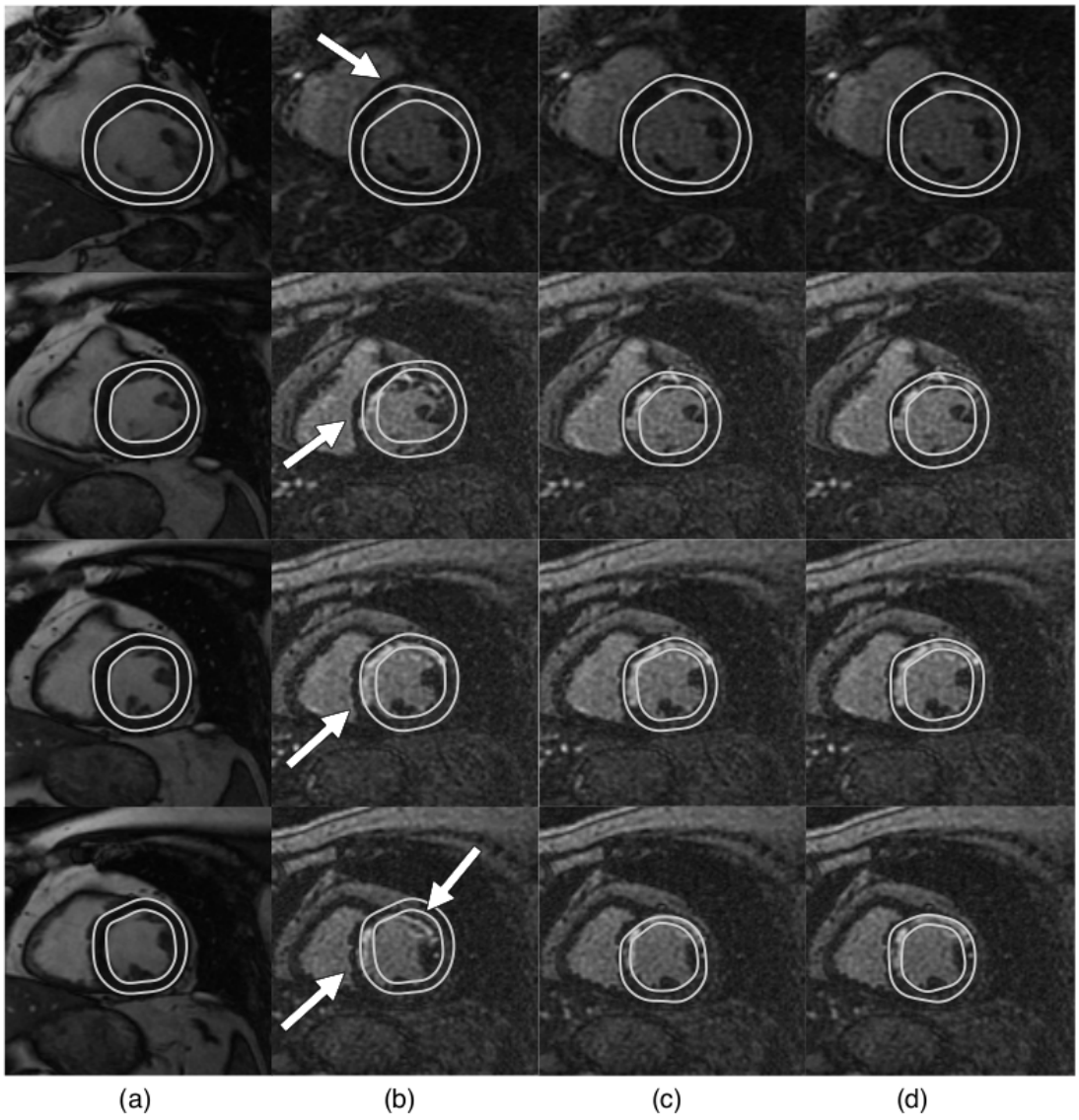


Figure 10: Clinical data registration results. Column (a) shows endo- and epicardial contours delineated on the cine dataset. Column (b) shows the unregistered endo- and epicardial contours directly propagated to the corresponding LGE images from the cine contours in column (a); notice the arrows pointing to erroneous contour locations at the septal and anterior walls of the LV due to misalignment between the cine and FGRE datasets. Column (c) shows the contours that were propagated from the cine dataset to the LGE dataset after hierarchical registration using weighted TV flow with NGF. Column (d) shows the manually delineated ground truth contours.

5.3. Discussion of Comparison of Edge Preserving Properties of Weighted TV Flow and Bilateral Filter

Next we compared the weighted TV flow with the bilateral filter in Sec. 3.5.1. We observe in Table 3 that edges are better preserved with weighted TV flow than the bilateral filter with respect to CNR as well as the structural similarity index metric. This is due to the fact that the bilateral filter can be viewed as a weighted Gaussian filter, as opposed to the weighted TV flow which favors BV functions and especially piecewise constant functions which are observed in MR imaging.

Note that the bilateral filter is computationally independent of its parameters, so its speed is independent of the amount of smoothing. We observed that the Matlab implementation of the bilateral filter for the Shepp-Logan phantom took an average of 1.450 seconds (this excludes the time required to tune the parameters σ_1 and σ_2 which is

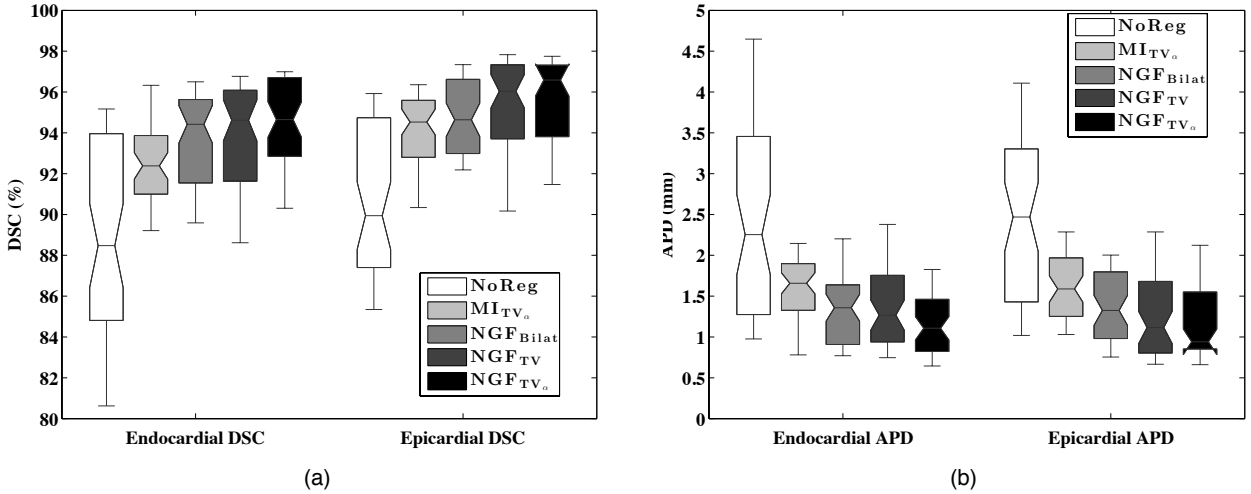


Figure 11: Registration results are shown for different registration schemes; $NoReg$, MI_{TV_α}, NGF_{Bilat}, NGF_{TV}, and NGF_{TV_α} represent no registration, registration with MI and weighted TV flow preprocessing, registration with NGF and bilateral filtering, registration with NGF and TV flow preprocessing, and registration with NGF and weighted TV flow preprocessing respectively. Registration accuracy results are presented for (a) DSC and (b) APD metrics. Each box plot shows the median (notch), interquartile range (box), and 10th/90th percentiles (whiskers) for the 2 evaluation metrics, which are computed from N = 49 slice images extracted from 8 patient cine and corresponding LGE datasets.

approximately 200 seconds), compared to the Matlab implementation of the weighted TV flow was of the order of a few milliseconds as seen in Table 2.

5.4. Discussion of Registration with Simulated Data

In Sec. 4.4 we performed registration experiments with simulated low SNR MR images and focused on the evaluation of various preprocessing techniques for removing noise and improving registration accuracy. Specifically, a hierarchical NGF based registration was chosen to register the misaligned noisy images after they have been preprocessed by either Gaussian, bilateral, TV flow, or the weighted TV flow filter at each scale in the hierarchical framework. It was shown that the TV flow and weighted TV flow methods were able to achieve significantly better registration accuracy compared to Gaussian and bilateral filtering methods (Fig. 9). This improvement is likely due to the fact that TV flow based methods were able to simultaneously suppress the added Rician noise and preserve the prominent structural edges in the input images (see Fig. 8). Moreover, it appears that weighted TV flow approach was able to further improve the TRE results in the case of large initial misalignment (i.e. 20 mm) compared to the standard TV flow approach. This is due to the increased focus on edge preservation associated with weighted TV flow. The edge-preserving feature is critical to registration success, since NGF is an edge-based registration metric.

5.5. Discussion of Registration with Clinical Data

In Sec. 4.6 we performed registration experiments with clinical data consisting of cine and LGE MR images. This is a challenging registration task, as the cine and LGE MR images have inherently different imaging contrast, and the LGE images suffer from noise and intensity inhomogeneity artefacts. We compared registration results using the NGF metric with bilateral, TV flow, or the weighted TV flow preprocessing, as well as registration using the MI metric with weighted TV flow filtering. Note that registration using NGF metric with Gaussian filtering was omitted in this scenario due to its relative poor performance in the previous simulated registration results. The evaluation criteria consisting of DSC and APD showed that registration with NGF outperformed registration with MI, and the preprocessing filter using weighted TV flow was again the best option for optimal registration accuracy (see Fig. 11). The improvements observed using the weighted TV flow preprocessing method were not as drastic in this case compared with the previous experiments with simulated datasets. This is to be expected, as significant noise was added to the simulated datasets to emulate very low SNR conditions, where the enhanced denoising properties of weighted TV flow was clearly demonstrated. In contrast, the clinical datasets are less noisy, and the advantage of

the weighted TV flow method is not as clear. Nevertheless, the improvements of the weighted TV flow approach in terms of the registration evaluation criteria of DSC and APD over the alternatives were still observed, and found to be statistically significant as assessed by the Wilcoxon signed-rank test.

The reason that registration with the NGF metric showed improvement over registration with the MI metric may be due to the fact that NGF is less sensitive to intensity inhomogeneities. By definition, the computed gradients in NGF are normalized according to equation (15), therefore the NGF metric is primarily concerned with the direction of the gradients associated with anatomical edges rather than the magnitude itself. Thus, NGF is less sensitive to intensity inhomogeneity, which is a problem within the clinical LGE dataset. Conversely, registration with MI may fail when intensities associated with one tissue type is homogeneous in the reference image, but suffers from intensity inhomogeneity artefacts in the corresponding region of the template image or vice versa. Furthermore, the observed improvement for the NGF_{TV_α} registration approach compared to the NGF_{Bilat} and NGF_{TV} registration approaches is again likely due to the enhanced denoising and edge preservation property of weighted TV flow.

6. Conclusion

We derived the weighted TV flow as a limit of the multiscale hierarchical decomposition, with weighted TV seminorm as a regularizer. We observed that the resultant weighted TV flow is better in preserving edges in MR images than weighted Gaussian filtering like Bilateral filtering. We also showed theoretically and experimentally that the speed of the weighted TV flow is completely determined by the speed function chosen by the user.

As weighted TV flow is edge-preserving as well as inherently multiscale in nature, it is natural to use it to extract various scales from noisy MR images for hierarchical registration. We showed that this approach is statistically superior when used in combination with normalized gradient field as a dissimilarity metric.

7. Appendix A

In this section we sketch the derivations of some interesting properties of the multiscale weighted TV flow. We work with weak solutions of problem 10, as in Andreu et al. (2002). In this section, we use D instead of ∇ to indicate derivative in the distributional sense. First, we compute the $\alpha*$ norm (12) of the speed of the weighted TV flow. To this effect we prove the following proposition.

Proposition 1. *For the multiscale weighted TV flow* (10) *the $\alpha*$ norm of the rate of change of the flow $u(\cdot\,,t)$ is:*

$$\left\| \frac{\partial u}{\partial t}(\cdot\,,t) \right\|_{\alpha*} = \mu(t). \tag{17}$$

Sketch of proof. Since u is a weak solution of (10) there exists a vector valued function $z(x,t) \in L^\infty((0,\infty)\times\Omega)$ with $\|z\|_{L^\infty} \leq 1$, (formally, $z = \frac{Du}{|Du|}$), such that

$$\int_\Omega \phi(x,t)\frac{\partial u}{\partial t}(x,t)\,dx = -\mu(t)\int_\Omega \alpha(x) z(x,t)\cdot D\phi(x)\,dx, \tag{18}$$

for all $\phi \in L^2(\Omega)\cap BV_\alpha(\Omega)$, and

$$\int_\Omega \alpha(x) z(x,t)\cdot Du(x,t)\,dx = |u|_{TV_\alpha(\Omega)}. \tag{19}$$

From (18) we get

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

If $\varphi = u(\cdot\,,t)$, then in view of (19) we get an equality, and the conclusion follows. □

The equation (17) asserts that the speed of the multiscale flow (10), 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 1 that

$$\left\| \frac{\partial v}{\partial t}(\cdot, t) \right\|_{\alpha*} = \mu(t). \tag{20}$$

Since u is a weak solution of (10) we also have

$$\int_\Omega (u(x,t) - \phi(x)) \frac{\partial u}{\partial t}\, dx = \mu(t) \int_\Omega \alpha(x) z(x,t) \cdot D\phi\, dx \tag{21}$$

$$= -\mu(t) |\, u(\cdot, t)\,|_{TV_\alpha(\Omega)} \tag{22}$$

Letting $\phi = 0$, we find

$$\int_\Omega u \frac{\partial u}{\partial t}\, dx \equiv \frac{1}{2} \frac{d}{dt} \| u(\cdot, t) \|^2_{L^2(\Omega)} \equiv -\mu(t)\, |\, u(\cdot, t)\,|_{TV_\alpha(\Omega)}.$$

Thus, we have

$$\frac{d}{dt} \| u(\cdot, t) \|^2_{L^2(\Omega)} = -2\, \mu(t)\, |\, u(\cdot, t)\,|_{TV_\alpha(\Omega)}. \tag{23}$$

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

Proposition 2. *For the multiscale TV flow* (10) *the L^2 energy of the function u is non-increasing for all times t.*

Using (23) we can derive the following energy decomposition of the energy of f.

Proposition 3. *For the multiscale TV flow* (10) *we have*

$$\| f \|^2_{L^2(\Omega)} = \| u(\cdot, t) \|^2_{L^2(\Omega)} + 2 \int_0^t \mu(s)\, |\, u(\cdot, s)\,|_{TV_\alpha(\Omega)}\, ds, \tag{24}$$

for any $t \geq 0$.

This follows from integrating (23) and using the initial condition, $u(\cdot, 0) = f$.

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 4. *For the multiscale weighted TV flow* (10) *the weighted BV seminorm of $u(\cdot, t)$ is non-increasing for all times t.*

Sketch of proof. Substituting $\phi = \frac{\partial u}{\partial t}$ in (18) we have

$$\int_\Omega \Big(\frac{\partial u}{\partial t} \Big)^2 dx = -\mu(t) \int_\Omega \alpha(x) z(x,t) \cdot D\big(\frac{\partial u}{\partial t}\big)\, dx = -\mu(t) \int_\Omega \frac{\partial}{\partial t} \alpha(x) Du\, dx = -\mu(t) \frac{d}{dt} |\, u(\cdot, t)\,|_{TV_\alpha(\Omega)}.$$

Since, $\mu(t) > 0$ for all t it follows that $|\, u(\cdot, t)\,|_{TV_\alpha(\Omega)}$ is nonincreasing. □

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

Proposition 5. *For the weighted TV flow* (10) *we define the residual $v(\cdot, t) := f - u(\cdot, t)$. Then we have the following inequality for its star-norm*

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

Sketch of proof. We have $v(\cdot, t) = f - u(\cdot, t)$ and $v(x, 0) = 0$. Integrating (18) we have for any $\phi \in BV_\alpha(\Omega) \cap L^2(\Omega)$

$$\begin{aligned}
&\Big| \int_\Omega \phi(x) v(x,t)\, dx \Big| \\
&= \Big| \int_\Omega \int_0^t \mu(s)\alpha(x) z(x,s) \cdot D\phi(x)\, dx \Big| \\
&\leq |\, \phi \,|_{TV_\alpha(\Omega)} \int_0^t \mu(s)\, dt
\end{aligned}$$

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

8. Appendix B: Numerical Scheme

In this section we provide the details of a semi-implicit numerical scheme for solving the weighted TV flow(10) for $\Omega \subset \mathbb{R}^2$. The scheme can be easily extended for $\Omega \subset \mathbb{R}^3$. See Breuß et al. (2006) for numerical analysis of the regularized implicit scheme for one dimensional case. Let h and τ be the space and time discretization steps respectively. We denote the value of the function u at the location $(x, y) \equiv (ih, jh)$ at time $t = n\tau$ by $u^n_{i,j}$. We set $u^0_{i,j} = f_{i,j}$. With this notation we have the following discretization for the right hand side of (10)

$$\begin{aligned}
&\mu\left(\triangle_{-x}\Big[\frac{\alpha_{i,j}\triangle_{+x} u_{i,j}}{\sqrt{\epsilon^2 + (\triangle_{+x}u_{i,j})^2 + (\triangle_{0y}u_{i,j})^2}} \Big] + \triangle_{-y}\Big[\frac{\alpha_{i,j}\triangle_{+y} u_{i,j}}{\sqrt{\epsilon^2 + (\triangle_{0x}u_{i,j})^2 + (\triangle_{+y}u_{i,j})^2}} \Big] \right) \\
&= \frac{\mu}{h^2}\Bigg[\frac{\alpha_{i,j}(u_{i+1,j} - u_{i,j})}{\sqrt{\epsilon^2 + (\triangle_{+x}u_{i,j})^2 + (\triangle_{0y}u_{i,j})^2}} - \frac{\alpha_{i-1,j}(u_{i,j} - u_{i-1,j})}{\sqrt{\epsilon^2 + (\triangle_{-x}u_{ij})^2 + (\triangle_{0y}u_{i-1,j})^2}} \\
&\quad + \frac{\alpha_{i,j}(u_{i,j+1} - u_{i,j})}{\sqrt{\epsilon^2 + (\triangle_{0x}u_{i,j})^2 + (\triangle_{+y}u_{i,j})^2}} - \frac{\alpha_{i,j-1}(u_{i,j} - u_{i,j-1})}{\sqrt{\epsilon^2 + (\triangle_{0x}u_{i,j-1})^2 + (\triangle_{-y}u_{i,j})^2}} \Bigg],
\end{aligned}$$

where, $\triangle_+, \triangle_-, \triangle_0$ denote the forward, backward and central difference schemes, respectively. We use the following notation

$$\begin{aligned}
C_E &= \frac{1}{\sqrt{\epsilon^2 + (\triangle_{+x}u^n_{i,j})^2 + (\triangle_{0y}u^n_{i,j})^2}}, \quad C_W = \frac{1}{\sqrt{\epsilon^2 + (\triangle_{-x}u^n_{i,j})^2 + (\triangle_{0y}u^n_{i-1,j})^2}}, \\
C_S &= \frac{1}{\sqrt{\epsilon^2 + (\triangle_{0x}u^n_{i,j})^2 + (\triangle_{+y}u^n_{i,j})^2}}, \quad C_N = \frac{1}{\sqrt{\epsilon^2 + (\triangle_{0x}u^n_{i,j-1})^2 + (\triangle_{-y}u^n_{i,j})^2}}.
\end{aligned}$$

Thus, we get the following discretization for (10)

$$\begin{aligned}
&\frac{u^{n+1}_{i,j} - u^n_{i,j}}{\tau} \\
&= \frac{\mu^{n+1}}{h^2}\Big[C_E\alpha_{i,j}u^n_{i+1,j} + C_W\alpha_{i-1,j}u^n_{i-1,j} + C_S\alpha_{i,j}u^n_{i,j+1} + C_N\alpha_{i,j-1}u^n_{i,j-1} - (C_E\alpha_{i,j} + C_W\alpha_{i-1,j} + C_S\alpha_{i,j} + C_N\alpha_{i,j-1})u^{n+1}_{i,j} \Big].
\end{aligned}$$

Rearranging the terms we get,

$$u^{n+1}_{i,j} = \frac{u^n_{i,j}h^2 + \mu^{n+1}\tau(C_E\alpha_{i,j}u^n_{i+1,j} + C_W\alpha_{i-1,j}u^n_{i-1,j} + C_S\alpha_{i,j}u^n_{i,j+1} + C_N\alpha_{i,j-1}u^n_{i,j-1})}{h^2 + \mu^{n+1}\tau(C_E\alpha_{i,j} + C_W\alpha_{i-1,j} + C_S\alpha_{i,j} + C_N\alpha_{i,j-1})}.$$

9. Acknowledgements

The corresponding author is supported by the Fields-Ontario Postdoctoral Fellowship. The research is funded by an NSERC Discovery Grant, the Federal Development Agency of Canada, and the Canadian Institutes of Health Research grant number MOP-93531.

10. References

Acar, R., Vogel, C., 1994. Analysis of bounded variation penalty methods of ill-posed problems. Inverse Problems 10, 1217–1229.

Andreu, F., Ballester, C., Caselles, V., Mazón, J., 2001. Minimizing total variational flow. Differential and integral equations 14, 321–360.

Andreu, F., Caselles, V., Diaz, J., Mazón, J., 2002. Some qualitative properties for the total variational flow. Journal of functional analysis 188, 516–547.

Andreu, F., Caselles, V., Mazón, J., 2004. Parabolic quasilinear equations minimizing linear growth Functionals. volume 223 of *Progress in mathematics*. Birkhäuser, Switzerland.

Anzellotti, G., 1983. Pairings between measures and bounded functions and compensated compactness. Ann. di Mathematica Pura ed Appl. 135, 293–318.

Athavale, P., 2009. Novel integro-differential schemes for multiscale image representation. Ph.D. thesis. University of Maryland, College Park.

Athavale, P., Xu, R., Radau, P., Nachman, A., Wright, G., 2013. Multiscale TV flow with applications to fast denoising and registration, in: SPIE Medical Imaging, International Society for Optics and Photonics. pp. 86692K–86692K.

Aubert, G., Kornprobst, P., 2006. Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (second edition). volume 147 of *Applied Mathematical Sciences*. Springer-Verlag.

Borsic, A., Graham, B., Adler, A., Lionheart, W., 2010. In vivo impedance imaging with total variation regularization. IEEE Transactions on Medical Imaging 29, 44–54.

Breuß, M., Brox, T., Bürgel, A., Sonar, T., Weickert, J., 2006. Numerical aspects of tv flow. Numerical Algorithms 41, 79–101.

Brezis, H., 1983. Analyse fonctionelle. volume 5. Masson.

Burger, M., Frick, K., Osher, S., Scherzer, O., 2007. Inverse total variation flow. Multiscale Modeling & Simulation 6, 366–395.

Carrillo, A., Duerk, J., Lewin, J., Wilson, D., 2000. Semiautomatic 3-D image registration as applied to interventional MRI liver cancer treatment. Medical Imaging, IEEE Transactions on 19, 175–185.

Catté, F., Lions, P.L., Morel, J.M., Coll, T., 1992. Image selective smoothing and edge detection by nonlinear diffusion. SIAM J. Numer. Anal. 29, 182–193.

Chambolle, A., Lions, P.L., 1997. Image recovery via total variation minimization and related problems. Numer. Math. 76, 167–188.

Chen, Y., Rao, M., 2003. Minimization problems and associated flows related to weighted p energy and total variation. SIAM journal on mathematical analysis 34, 1084–1104.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory, in: Information processing in medical imaging, pp. 263–274.

Coupé, P., Manjón, J., Gedamu, E., Arnold, D., Robles, M., Collins, D., 2010. Robust rician noise estimation for MR images. Medical image analysis 14, 483–493.

Crandall, M., Liggett, T., 1971. Generation of semi-groups of nonlinear transformations on general banach spaces. American Journal of Mathematics 93, 265–298.

Detsky, J., Paul, G., Alexander, J., Graham, A., 2009. Reproducible classification of infarct heterogeneity using fuzzy clustering on multicontrast delayed enhancement magnetic resonance images. Medical Imaging, IEEE Transactions on 28, 1606–1614.

Evans, L., Gariepy, R., 1992. Measure theory and fine properties of functions. CRC press, Boca Raton, FL.

Fei, B., Wheaton, A., Lee, Z., Duerk, J., Wilson, D., 2002. Automatic MR volume registration and its evaluation for the pelvis and prostate. Physics in Medicine and Biology 47, 823.

Gudbjartsson, H., Patz, S., 1995. The rician distribution of noisy MRI data. Magnetic Resonance in Medicine 34, 910–914.

Haber, E., Modersitzki, J., 2006. Intensity gradient based registration and fusion of multi-modal images, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006. Springer, pp. 726–733.

Heindel, W., Jernigan, T., Archibald, S., Achim, C., Masliah, E., Wiley, C., 1994. The relationship of quantitative brain magnetic resonance imaging measures to neuropathologic indexes of human immunodeficiency virus infection. Archives of neurology 51, 1129.

Holden, M., Hill, D., Denton, E., Jarosz, J., Cox, T., Rohlfing, T., Goodey, J., Hawkes, D., 2000. Voxel similarity measures for 3-D serial MR brain image registration. Medical Imaging, IEEE Transactions on 19, 94–102.

Iijima, T., 1959. Basic theory of pattern observation. Papers on technical group on Automata and Automatic Control, IECE, Dec. (in Japanese) .

Jonasson, L., Bresson, X., Thiran, J.P., Wedeen, V., Hagmann, P., 2007. Representing diffusion MRI in 5-D simplifies regularization and segmentation of white matter tracts. Medical Imaging, IEEE Transactions on 26, 1547–1554.

Koenderink, J., 1984. Structure of images. Biological Cybernetics 50, 363–370.

Li, D., Li, H., Wan, H., Chen, J., Gong, G., Wang, H., Wang, L., Yin, Y., 2012. Multiscale registration of medical images based on edge preserving scale space with application in image-guided radiation therapy. Physics in Medicine and Biology 57, 5187.

Meyer, Y., 2001. Oscillating patterns in image processing and nonlinear evolution equations. volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

Michálek, J., Čapek, M., Kubínová, L., 2011. Nonrigid registration of CLSM images of physical sections with discontinuous deformations. Microscopy and Microanalysis 17, 923.

Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Thirion, B., 2011. Total variation regularization for fMRI-based prediction of behavior. Medical Imaging, IEEE Transactions on 30, 1328–1340.

Modersitzki, J., 2004. Numerical Methods for Image Registration (Numerical Mathematics and Scientific Computation). Oxford university press USA.

Montillo, A., Udupa, J., Axel, L., Metaxas, D., 2003. Interaction between noise suppression and inhomogeneity correction in MRI, in: Medical Imaging 2003, International Society for Optics and Photonics. pp. 1025–1036.

Pennell, D.J., Sechtem, U.P., Higgins, C.B., Manning, W.J., Pohost, G.M., Rademakers, F.E., van Rossum, A.C., Shaw, L.J., Yucel, E.K., 2004. Clinical indications for cardiovascular magnetic resonance (CMR): Consensus panel report. European heart journal 25, 1940–1965.

Perona, P., Malik, J., 1988. Scale-space and edge detection using anisotropic diffusion. Technical Report. University of California at Berkeley. Berkeley, CA, USA.

Pluim, J., Maintz, J., Viergever, M., 2003. Mutual-information-based registration of medical images: a survey. Medical Imaging, IEEE Transactions on 22, 986–1004.
Ramani, S., Fessler, J., 2012. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. Medical Imaging, IEEE Transactions on 31, 677–688.
Rudin, L., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Phys. D 60, 259–268.
Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: application to breast MR images. Medical Imaging, IEEE Transactions on 18, 712–721.
Scherzer, O., Weickert, J., 2000. Relations between regularization and diffusion filtering. Journal of Mathematical Imaging and Vision 12, 43–63.
Song, X., Pogue, B., Jiang, S., Doyley, M., Dehghani, H., Tosteson, T., Paulsen, K., 2004. Automated region detection based on the contrast-to-noise ratio in near-infrared tomography. Applied optics 43, 1053–1062.
Strong, D., Chan, T., 1996. Spatially and Scale Adaptive Total Variation Based Regularization and Anisotropic Diffusion in Image Processing. Technical Report. UCLA Math Department CAM Report.
Strong, D., Chan, T., 2003. Edge-preserving and scale-dependent properties of total variation regularization. Inverse problems 19, S165.
Sylvain, P., Kornprobst, P., Tumblin, J., Durand, F., 2008. Bilateral filtering: theory and applications. volume 4 of *Foundations and trends in computer graphics and vision* ®. Now Publishers Inc., Hanover, MA.
Tadmor, E., Athavale, P., 2009. Multiscale image representation using integro-differential equations. Inverse problems and imaging 3, 693–710.
Tadmor, E., Nezzar, S., Vese, L., 2004. A multiscale image representation using hierarchical (BV, L^2) decompositions. Multiscale Model. Simul. 2, 554–579.
Tomasi, C., Manduchi, R., 1998. Bilateral filtering for gray and color images, in: Proc. IEEE ICCV, Bombay, India. pp. 839–846.
Vese, L., 2001. A study in the BV space of a denoising-deblurring variational problem. Appl. Math. Optim. 44, 131–161.
Viola, P., Wells III, W.M., 1995. Alignment by maximization of mutual information, in: Computer Vision, 1995. Proceedings., Fifth International Conference on, IEEE. pp. 16–23.
Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: From error visibility to structural similarity. Image Processing, IEEE Transactions on 13, 600–612.
Weickert, J., Ishikawa, S., Imiya, A., 1997. On the history of gaussian scale-space axiomatics, in: Gaussian Scale-space theory. Springer, pp. 45–59.
Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R., 1996. Multi-modal volume registration by maximization of mutual information. Medical image analysis 1, 35–51.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics bulletin 1, 80–83.
Witkin, A.P., 1984. Scale-space filtering: A new approach to multi-scale description, in: Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84., IEEE. pp. 150–153.
Xu, R., Athavale, P., Lu, Y., Radau, P., Wright, G., 2013. Myocardial segmentation in late-enhancement MR images via registration and propagation of cine contours, in: IEEE Int. Symp. Biomed. Imag., pp. 844–847.

Prashant Athavale received his B.E. in electrical engineering in 1999 from VJTI, Mumbai University. He obtained his M.S. in Electrical engineering, and in Applied Mathematics in 2001& 2003, respectively from University of Toledo. He received his Ph.D. in Applied Mathematics in 2009 from University of Maryland, College Park. He worked as a postdoctoral scholar at UCLA from 2009 to 2011, where he worked on placental image processing. Currently, he is working in the University of Toronto as a Fields-Ontario postdoctoral fellow. His research interests include multiscale analysis, image processing, numerical analysis, and image registration.

Robert Xu received both of his BASc. And MASc. degrees in Electrical and Computer Engineering from the University of Waterloo, Waterloo, Ontario, Canada in 2008 and 2010 respectively. He is currently a doctoral candidate in the Department of Medical Biophysics at the University of Toronto, Toronto, Ontario, Canada. His research interests include medical image processing, image registration, and motion modeling in the context of image-guided therapy.

Perry Radau received his B.Sc. in Astrophysics (University of Calgary), M.Sc. in Neuroscience and a PhD in Medical Biophysics (Western University). He developed automated analysis techniques for brain medical images, and in cooperation with Hermes Medical Solutions he developed the internationally successful clinical software BRASS for brain images. Perry joined Sunnybrook Hospital (Toronto), and as Technical Director of the Imaging Research Centre for Cardiac Intervention (IRCCI) he developed visualization software and licensed segmentation techniques for cardiac MRI. Perry has initiated Straylight Health Solutions Ltd. and is translating academic research for use in commercial diagnostic software, focusing on dementia and parkinsonism.

Adrian Nachman received his B.Sc. degree from McGill University in 1974, and his PhD in Mathematics from Princeton University in 1980. From 1979–1981 he taught at Yale University, and from 1981–2001 at the University of Rochester. He joined the University of Toronto in 2001 as Professor in the Department of Mathematics and the Edward S. Rogers Sr. Department of Electrical & Computer Engineering. His research interests are mainly in Inverse Problems and Image Processing in Medical Imaging. He has made fundamental contributions to the field of

Inverse Problems, solving several well-known longstanding open problems, and introducing powerful methods which are widely used.

Graham A. Wright is the Director of the Schulich Heart Research Program at Sunnybrook Health Sciences Centre, a senior scientist at Sunnybrook Research Institute, and a professor in the Department of Medical Biophysics at the University of Toronto. Dr. Wright was appointed the Canada Research Chair in Imaging for Cardiovascular Therapeutics in 2010. The research focus of Dr. Wrights group is cardiovascular imaging, with an emphasis on MRI. Together with trainees and collaborators, he has published over 144 peer-reviewed papers and 400 conference abstracts, which have garnered numerous awards and peer-reviewed operating grant funding, and resulted in 20 patents.