NIH Public Access
Author Manuscript
Proc IEEE Int Symp Biomed Imaging. Author manuscript; available in PMC 2014 March 19.

Published in final edited form as:
Proc IEEE Int Symp Biomed Imaging. 2013 ; : 326–329. doi:10.1109/ISBI.2013.6556478.

A FAST MAJORIZE MINIMIZE ALGORITHM FOR HIGHER DEGREE TOTAL VARIATION REGULARIZATION

Yue Hu[*], **Sathish Ramani**[‡], and **Mathews Jacob**[†]

[*]Department of Electrical and Computer Engineering, University of Rochester, NY, USA

[‡]Department of Electrical Engineering and Computer Science, University of Michigan, MI, USA

[†]Department of Electrical and Computer Engineering, University of Iowa, IA, USA

Abstract

The main focus of this paper is to introduce a computationally efficient algorithm for solving image recovery problems, regularized by the recently introduced higher degree total variation (HDTV) penalties. The anisotropic HDTV penalty is the fully separable L_1 semi-norm of the directional image derivatives; the use of this penalty is seen to considerably improve image quality in biomedical inverse problems. We introduce a novel majorize minimize algorithm to solve the HDTV optimization problem, thus considerably speeding it over the previous implementation. Specifically, comparisons with previous iterative reweighted algorithm show an approximate ten fold speedup. The new algorithm enables us to obtain reconstructions that are free of patchy artifacts exhibited by classical TV schemes, while being comparable to state of the art total variation regularization schemes in run time.

Index Terms

Higher degree total variation; majorize minimize; compressed sensing

---

1. INTRODUCTION

The recovery of image data from ill-conditioned measurements is a common problem in many biomedical applications, including MRI, PET, and microscopy. Our main motivation is the application of this framework to accelerated MRI, where we consider the recovery of images from under sampled Fourier data. Many regularization schemes have been proposed to make the recovery well-posed. One of the popular regularization penalty is the total variation (TV) functional, which is the ℓ_1 semi-norm of the image gradient. The TV penalty has many desirable properties such as convexity, rotation invariance, and most importantly the ability to preserve image edges. However, since TV enforces reconstructions with sparse gradients, the reconstructed images often suffer from patchy or staircase artifacts. To overcome these problems, there has been a growing interest in using higher order image regularization penalties [1, 2, 3, 4, 5, 6, 7].

We have introduced a family of image regularization penalties termed as higher degree TV (HDTV) functionals [1, 2]. These penalties generalize several of the classical higher degree regularization schemes mentioned above and provide improved reconstructions. The anisotropic HDTV penalty is the fully separable L_1 norm of the nth degree directional image derivatives. Our experiments show that HDTV regularization inherits the desirable properties of standard TV, while minimizing the patchy artifacts and enhancing the ridge-like features in the image. The main challenge associated with the current HDTV framework is its high computational complexity compared to the state of the art TV schemes.

Specifically, the iterative reweighted algorithm in the previous implementation involves subproblems with large condition numbers, resulting in slow convergence.

The main focus of this paper is to introduce a computationally efficient HDTV algorithm. We approximate the L_1 penalty by a Huber function, parametrized by a single parameter β; the quality of the approximation improves as $\beta \to \infty$. We then majorize the corresponding approximate penalty term by a quadratic function. This majorization enables us to decouple the reconstruction into two simple sub-problems, each of which have analytical solutions. Since the choice of β involves a compromise between accuracy and convergence rate, we rely on a continuation strategy to obtain a fast and accurate algorithm. Specifically, we start with a small value of β and gradually increase it to a high value. We observe that the proposed algorithm offers approximately a ten fold speedup compared to the iteratively reweighted majorize minimize (IRMM) scheme. This development enables us to obtain higher quality reconstructions that are free of patchy artifacts exhibited by classical TV schemes, while being comparable to state of the art total variation schemes in run time. The MATLAB implementation of the fast HDTV algorithm is available for download at http://www.engineering.uiowa.edu/~jcb/software.html.

2. BACKGROUND

We consider the recovery of a continuously differentiable complex image *f* from its noisy and undersampled measurement **b**, specified by $\mathbf{b} = \mathscr{A}(f) + \mathbf{n}$. The problem can be formulated as an optimization problem:

$$\hat{f}=\underset{f}{\operatorname{argmin}}\underbrace{\| \mathscr{A}(f)-\mathbf{b} \|^2+\lambda \mathscr{J}(f)}_{\mathscr{C}(f)}. \quad (1)$$

The standard TV regularization is the L_1 norm of the image gradient, specified as $\mathscr{J}_1(f) = \int_\Omega |\nabla f| d\mathbf{r}$. The anisotropic HDTV regularization penalty [1] is defined as:

$$\mathscr{J}_{n\mathrm{HDTV}}=\int_\Omega \left(\frac{1}{2\pi}\int_0^{2\pi} |f_{\theta,n}(\mathbf{r})|\, d\theta\right) d\mathbf{r} \quad (2)$$

where $f_{\theta,n}(\mathbf{r})$ is the rotation steerable *n*th degree directional derivative along the unit vector $\mathbf{u}_\theta = (\cos\theta, \sin\theta)$. Note that the directional derivatives are rotation steerable:

$$f_{\theta,n}(\mathbf{r})=\mathbf{s}_n^H(\theta)\mathscr{D}_n f(\mathbf{r}). \quad (3)$$

Here, $s_n(\theta)$ is the vector of trigonometric polynomials and $\mathscr{D}_n$ is the differential operator that provides the partial derivatives of *f*. For example, in the 1st degree case ($n = 1$):

$$\mathbf{s}_1(\theta)=[\cos\theta \sin\theta]^T; \mathscr{D}_1 f(\mathbf{r})=[f_x f_y]^T \quad (4)$$

The HDTV penalty is convex and rotation invariant. Moreover, since the L_1 norm is fully separable, a strong directional derivative at a specified orientation will not affect the attenuation of the directional derivatives along other directions. This is not the case with most of the classical higher order TV schemes [3, 4, 5, 6, 7].

Since sparsity promoting non-quadratic penalties are non-differentiable, it is difficult to solve the optimization problem using gradient-based algorithms (e.g. conjugate gradient method). In our previous work, we use an IRMM algorithm. Specifically, the optimization problem (1) is majorized by a weighted quadratic expression, which is solved using a conjugate gradient (CG) algorithm. The algorithm alternates between CG optimization and the recomputation of the weights from the current iterate. Since the spatially varying weights often tend to very high values, the quadratic subproblems have high condition numbers. Hence, the resulting algorithm is computationally expensive.

3. FAST MM ALGORITHM FOR HDTV

The HDTV recovery of a function $f : \mathbb{R}^2 \to \mathbb{C}$ from its measurements $\mathbf{b} = \mathscr{A}(f) + \mathbf{n}$ is posed as:

$$\mathscr{C}(f) = \| \mathscr{A} f - \mathbf{b} \|^2 + \lambda \int_\Omega \int_0^{2\pi} |f_{\theta,n}(\mathbf{r})| \, d\theta d\mathbf{r}. \quad (5)$$

Since the absolute function in the L_1 norm is not continuously differentiable, we approximate it by the Huber function:

$$\varphi_\beta(x) = \begin{cases} |x| - 1/2\beta & \text{if} \quad x \geq \frac{1}{\beta} \\ \beta |x|^2/2 & \text{else.} \end{cases} \quad (6)$$

The approximate cost function is thus specified by :

$$\mathscr{C}_\beta(f) = \| \mathscr{A} f - \mathbf{b} \|^2 + \lambda \int_\Omega \int_0^{2\pi} \varphi_\beta(|f_{\theta,n}(\mathbf{r})|) d\theta d\mathbf{r}. \quad (7)$$

Note that this approximation tends to the original HDTV penalty when $\beta \to \infty$. To realize a computationally efficient algorithm, we majorize the Huber function in the above expression by the quadratic function [8]:

$$\varphi_\beta(\left| f_{\theta,n}(\mathbf{r}) \right|) = \min_{g(\theta,\mathbf{r})} \left\{ \frac{\beta}{2} \| f_{\theta,n}(\mathbf{r}) - g(\theta,\mathbf{r}) \|^2 + \psi(\left| g(\theta,\mathbf{r}) \right|) \right\} \quad (8)$$

where $g(\theta, \mathbf{r}) : \mathbb{R}^2 \times [0, 2\pi] \to \mathbb{C}$ is an auxiliary function. The quadratic function majorizes the original penalty when $\psi(|g(\theta, \mathbf{r})|) = |g(\theta, \mathbf{r})|$ [9]. With this majorization, the cost function can be expressed as

$$\mathscr{C}_\beta(f) = \min_{g(\theta,\mathbf{r})} \| \mathscr{A} f - \mathbf{b} \|^2 + \lambda \int_\Omega \left\{ \frac{\beta}{2} \| f_{\theta,n} - g(\theta) \|^2 + |g(\theta)| \right\} d\mathbf{r} \quad (9)$$

Note that the optimization algorithm now involves the minimization of the right hand side of the above expression with respect to both functions $f(\mathbf{r}) : \mathbb{R}^2 \to \mathbb{C}$ and $g(\theta, \mathbf{r}) : \mathbb{R}^2 \times [0, 2\pi] \to \mathbb{C}$. We rely on an alternating minimization algorithm to solve the two functions. Specifically, we alternate between the minimization with respect to $f_{\theta,n}(\mathbf{r})$ and $g(\theta, \mathbf{r})$ as shown below.

3.1. Step one: minimization with respect to $g(\theta, \mathbf{r})$, assuming $f_{\theta,n}(\mathbf{r})$ to be fixed

Assuming $f_{\theta,n}(\mathbf{r})$ to be fixed, we solve the optimization problem with respect to $g(\theta, \mathbf{r})$. We term this step as the ”g-subproblem”. The minimization at each θ can be de-coupled to obtain

$$\mathscr{C}_f(g) = \frac{\beta}{2} \| f_{\theta,n}(\mathbf{r}) - g(\theta,\mathbf{r}) \|^2 + |g(\theta,\mathbf{r})| \, ; \forall \theta \in [0, 2\pi] \quad (10)$$

Minimizing $\mathscr{C}_f(g)$ with respect to g, we obtain:

$$g(\theta) = \max\left(|f_{\theta,n}| - \frac{1}{\beta}, 0\right)\frac{f_{\theta,n}}{|f_{\theta,n}|}; \forall \mathbf{r} \in \mathbb{R}^2; \theta \in [0, 2\pi]. \quad (11)$$

To obtain good reconstructions that are consistent with the continuous domain formulation, $g(\theta, \mathbf{r})$ needs to be evaluated for all angles; the storage of $g_{\theta,n}$ for large number of angles will result in an algorithm with high memory demand. Fortunately, we will see in the next subsection that the rest of the algorithm does not need $g(\theta, \mathbf{r})$; $\forall\theta$, but its projection to the space spanned by s(θ), specified by $\mathbf{q}(\mathbf{r}) = \int_0^{2\pi} \mathbf{s}(\theta) g(\theta, \mathbf{r}) d\theta$. This simplification results in an algorithm with considerably less memory demand.

3.2. Step two: Minimization with respect to $f_{\theta n}$(r), assuming g(θ, r) to be fixed

Assuming that $g(\theta, \mathbf{r})$ is fixed, we now minimize (9) with respect to $f_{\theta,n}(\mathbf{r})$. This can be reformulated as

$$f = \underset{f}{\mathrm{argmin}} \| \mathscr{A} f - \mathbf{b} \|^2 + \mathscr{C}_g(f), \quad (12)$$

where

$$\mathscr{C}_g(f) = \lambda \int_\Omega \left\{ \frac{\beta}{2} (\| f_\theta \|^2 - 2 \langle f_\theta, g(\theta) \rangle + \| g(\theta) \|^2) \right\} d\theta d\mathbf{r}. \quad (13)$$

Here, the norm and the inner product are defined on $L_2[0, 2\pi]$. Ignoring the constant term $\int_\Omega \| g(\theta) \|^2 d\mathbf{r}$ in the above expression, we obtain:

$$\mathscr{C}_g = \frac{\lambda\beta}{2} \int_\Omega \left\{ \| f_\theta(\mathbf{r}) \|^2 - 2 \langle f_\theta, g(\theta) \rangle \right\} d\mathbf{r} = \frac{\lambda\beta}{2} \int_\Omega \left\{ \mathscr{D} f(\mathbf{r})^H \mathbf{C}_n \mathscr{D} f(\mathbf{r}) - 2 (\mathscr{D} f(\mathbf{r}))^H \mathbf{q}(\mathbf{r}) \right\} d\mathbf{r},$$

where $\mathbf{C}_n = \int_0^{2\pi} \mathbf{s}_n(\theta) \mathbf{s}_n^H(\theta) d\theta \text{and} \mathbf{q}_n(\mathbf{r}) = \int_0^{2\pi} \mathbf{s}_n(\theta) g(\theta) d\theta$. Here, we used the steerability relationship of the directional derivatives from (3) to simplify the expression. Note that the criterion $\mathscr{C}_g(f)$ does not depend on $g(\theta, \mathbf{r})$, but only its projections to the space of trigonometric functions $s_n(\theta)$, specified by $\mathbf{q}(\mathbf{r})$. Since we do not have to store the variable $g(\theta, \mathbf{r})$ for all values of θ, this simplification considerably reduces the memory demand of the algorithm. In addition, the above expression is independent of the directional derivatives $f_{\theta,n}(\mathbf{r})$; it is only dependent on the partial derivatives of f, thanks to the steerability of the directional derivatives in terms of the partial derivatives. Setting the variational derivation of $\mathscr{C}_g(f)$ to be zero, we obtain:

$$(2\mathscr{A}^T \mathscr{A} + \lambda\beta \mathscr{D}_n^T \mathbf{C}_n \mathscr{D}_n) f = 2\mathscr{A}^T \mathbf{b} + \lambda\beta \mathscr{D}_n^T \mathbf{q}_n \quad (14)$$

The operator $\mathscr{D}_n^T \mathbf{C}_n \mathscr{D}_n$ has a simple expression in the discrete Fourier domain. The following are the discrete Fourier domain expressions in 1st and 2nd degree case:

$$\mathscr{F}[\mathscr{D}_1^T \mathbf{C}_1 \mathscr{D}_1] = \omega_x^2 + \omega_y^2 \quad (15)$$

$$\mathscr{F}[\mathscr{D}_2^T \mathbf{C}_2 \mathscr{D}_2] = \frac{1}{8} \left\{ 3\omega_{xx}^2 + 3\omega_{yy}^2 + 4\omega_{xy} + 2\omega_{xx}\omega_{yy} \right\} \quad (16)$$

Here, $\mathscr{F}$ denotes the discrete Fourier transform operator. These equations will get modified by the discrete approximation of the derivatives (e.g. finite differences). Thus, if the measurements are Fourier samples on a Cartesian grid (i.e. $\mathscr{A} f = \mathbf{S}\mathscr{F}$), (14) can be simplified by evaluating the discrete Fourier transform of both sides. Here, **S** is the sampling operator that picks the appropriate Fourier samples. Computing the discrete Fourier transform of both sides of (14), we get

$$(2\mathbf{S}^T\mathbf{S}+\lambda\beta\mathscr{F}[\mathscr{D}_n^T\mathbf{C}\mathscr{D}_n])f=2\mathbf{S}^T\mathbf{b}+\lambda\beta\mathscr{F}[\mathscr{D}_n^T\mathbf{q}_n]. \quad (17)$$

Thus, we obtain the analytical expression for f as:

$$f=\mathscr{F}^{-1}\left\{\frac{2\mathbf{S}^T\mathbf{S}\mathbf{b}+\lambda\beta\mathscr{F}[\mathscr{D}_n^T\mathbf{q}_n]}{2\mathbf{S}+\lambda\beta\mathscr{F}[\mathscr{D}_n^T\mathbf{C}_n\mathscr{D}_n]}\right\}. \quad (18)$$

When the Fourier samples are not on the Cartesian grid (for example, in parallel imaging), where the one step solution is not applicable, we could still solve the minimization problem using CG iterations.

4. RESULTS

We study the performance of the fast HDTV algorithm in the context of compressed sensing. To quantitatively compare methods, we rely on the signal to noise ratio (SNR) measure, defined as:

$$\mathrm{SNR}=-10\log_{10}\left(\frac{\|f_{\mathrm{orig}}-\hat{f}\|_F^2}{\|f_{\mathrm{orig}}\|_F^2}\right), \quad (19)$$

where $\hat{f}$ is the reconstructed image, f_{orig} is the original image, and $\|\cdot\|_F^2$ is the Frobenius norm.

We first study the dependence of the convergence rate on the choice of the parameter β and the specific continuation strategy to increment it. All the methods were implemented in MATLAB on an Intel Dual Core 2.66 GHz PC. In this work, we start with an initial value and increment it by a constant factor (i.e., $\beta_{new} = \beta \cdot \beta_{inc}$). We consider the reconstruction of a brain image with the accelerator of 1.65 using the fast HDTV algorithm. The plot of the cost as a function of number of iterations is shown in Fig. 1(a). It is observed that using the continuation scheme, where the parameters are initialized with a relatively small value and increased at each iteration, results in fast convergence.

We compare the proposed fast HDTV algorithm with the IRMM algorithm in the context of the recovery of brain MR image with acceleration factor of four in Fig. 1(b). We plot the SNR as a function of the CPU time using TV and 2nd degree HDTV with the IRMM algorithm and the proposed algorithm, respectively. We observe that fast HDTV provides an approximate speedup of 10-fold, compared with IRMM. We also implement the algorithm in MATLAB on a Linux workstation with two Core 2 quad-core processors. The proposed algorithm runs in around 6-7 seconds for a 256×256 image. The difference between the reconstructed images of both alorithms is minor. Thus, the speedup by fast HDTV algorithm does not impact the quality of the reconstruction.

10] We finally demonstrate the improvement in image quality offered by HDTV algorithm over standard TV. The reconstructions of a MR wrist image at acceleration rate of 3 with 35dB noise added are shown Fig. 2. We observe that there is a 0.3dB improvement in HDTV2 over standard TV. We also see that HDTV2 preserves subtle details. Besides, the image reconstructed by HDTV2 is more natural than TV reconstruction. The reconstructions of a brain MR image at acceleration of 1.65 and 25dB noise added is shown in Fig. 3. It is observed that HDTV2 provides more accurate reconstructions, while TV reconstruction shows patchy artifacts, which blur some of the details in the image.

5. CONCLUSION

We introduced a computationally efficient HDTV algorithm for MR image recovery. Specifically, we use a fast majorize minimize algorithm to solve the optimization problem. Our experiments show that the proposed algorithm reduces the computation time by 10-fold, compared to the IRMM method used previously,. The numerical results show that compared with TV, the fast HDTV algorithm overcomes the patchy/staircasing artifacts in the reconstruction.

Acknowledgments

This work is supported by NSF CCF-0844812, NSF CCF-1116067, NIH 1R21HL109710-01A1 and ACS RSG-11-267-01-CCE.

REFERENCES

1. Hu Y, Jacob M. Higher degree total variation (HDTV) regularization for image recovery. IEEE Transactions on Image Processing. 2012 May; vol. 21(no. 5):2559–2571. [PubMed: 22249711]
2. Hu Y, Jacob M. Improved higher degree total variation regularization. IEEE International Symposium on Biomedical Imaging. 2012 May.
3. Chan T, Marqina A, Mulet P. Higher-order total variation-based image restoration. SIAM J. Sci. Computing. 2000 Jul; vol. 22(no. 2):503–516.
4. Steidl G, Didas S, Neumann J. Splines in higher order TV regularization. International Journal of Computer Vision. 2006; vol. 70(no. 3):241–255.
5. You Y, Kaveh M. Fourth-order partial differential equations for noise removal. IEEE Transactions on Image Processing. 2000 Oct; vol. 9(no. 10):1723–1730. [PubMed: 18262911]
6. Lysaker M, Lundervold A, Tai X-C. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. IEEE Transactions on Image Processing. 2003 Dec; vol. 12(no. 12):1579–1590. [PubMed: 18244712]
7. Knoll F, Bredies K, Pock T, Stollberger R. Second order total generalized variation (TGV) for MRI. Magnetic Resonance in Medicine. 2010; vol. 65(no. 2):480–491. [PubMed: 21264937]
8. Cai J, Candes E, Shen Z. A singular value thresholding algorithm for matrix completion. SIAM J. Optim. 2010; vol. 20:1956–1982.
9. Lewis AS. The convex analysis of unitarily invariant matrix functions. Journal of Convex Analysis. 1995; vol. 2(no. 1):173–183.
10. Block KT, Uecker M, Frahm J. Suppression of mri truncation artifacts using total variation constrained data extrapolation suppression of mri truncation artifacts using total variation constrained data extrapolation. Int. Journal Biomedical Imaging. 2008

NIH-PA Author Manuscript

NIH-PA Author Manuscript

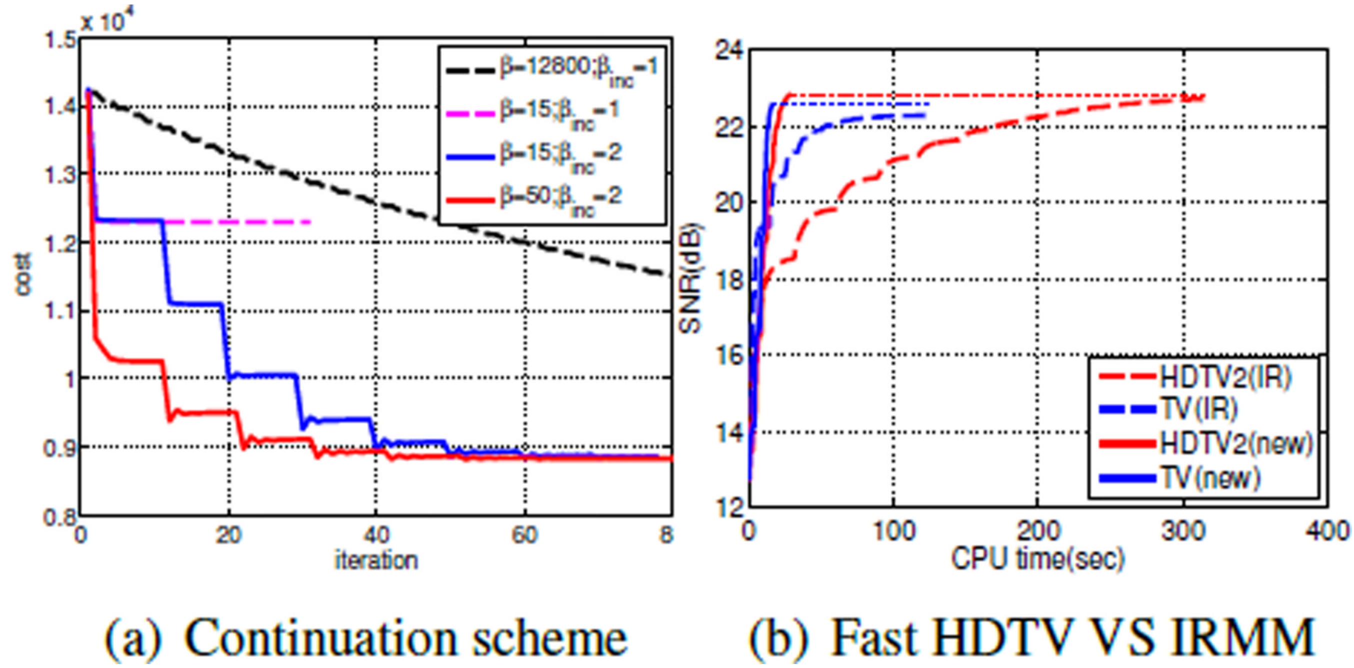

Figure 1.

(a) Comparison of different continuation schemes show that using continuation scheme results in fast convergence. (b) Comparison of IRMM with fast HDTV algorithm. The blue, blue dotted, red, red dotted curves correspond to TV by fast HDTV, TV by IRMM, HDTV2 by fast HDTV, HDTV2 by IRMM algorithm, respectively. We extend the original plot by dotted lines for easier comparisons of the final SNR. We see that fast HDTV takes 1/6 of the time taken by IRMM for TV, and 1/10 of the time taken by IRMM for HDTV2.

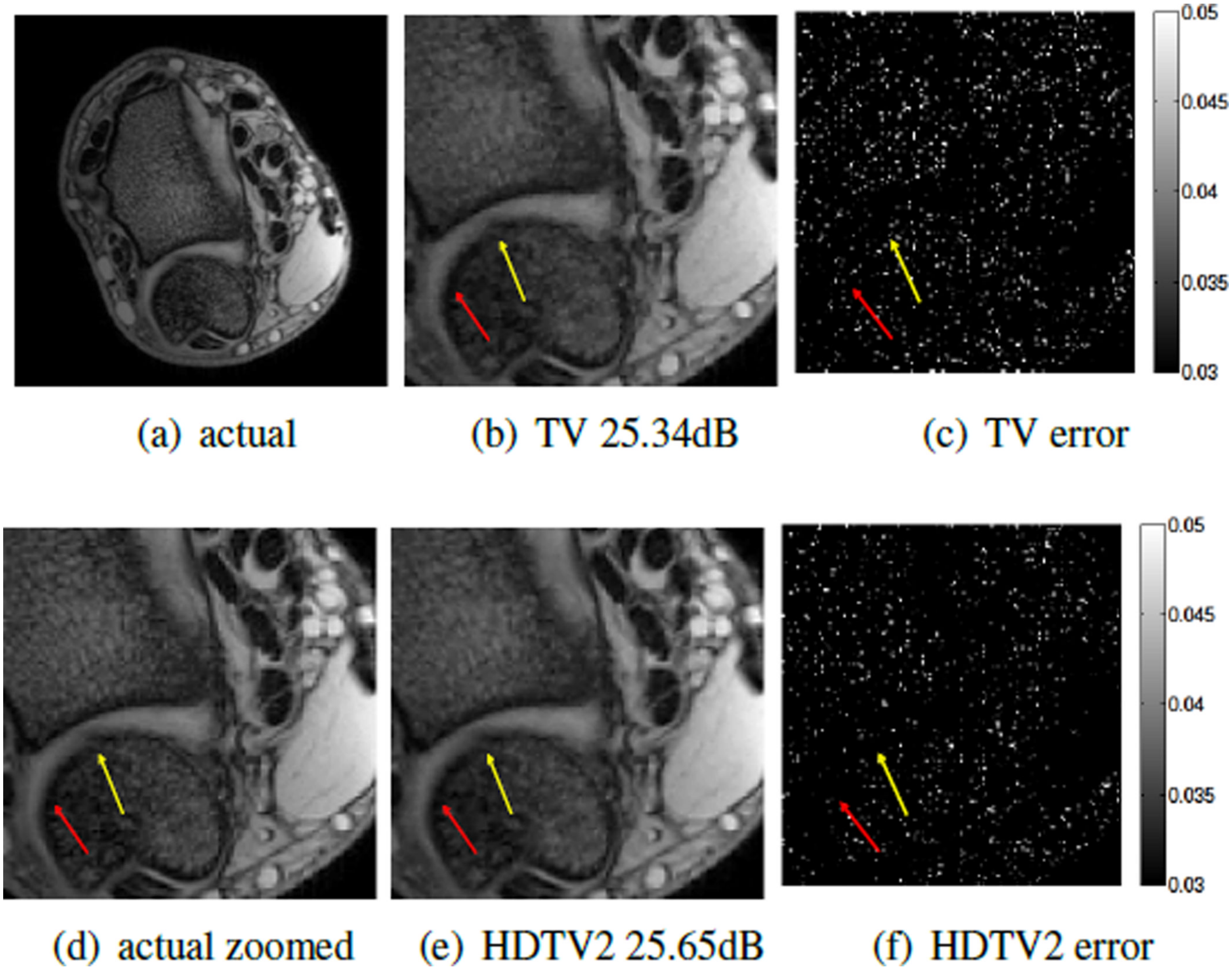

Figure 2.
Compressed sensing reconstruction of a wrist MR image from noisy and undersampled data. The acceleration factor is 3 with 35dB additive noise. (a) and (d) Original image. (b) and (c) TV reconstruction and error image. (e) and (f) HDTV2 reconstruction and error image.

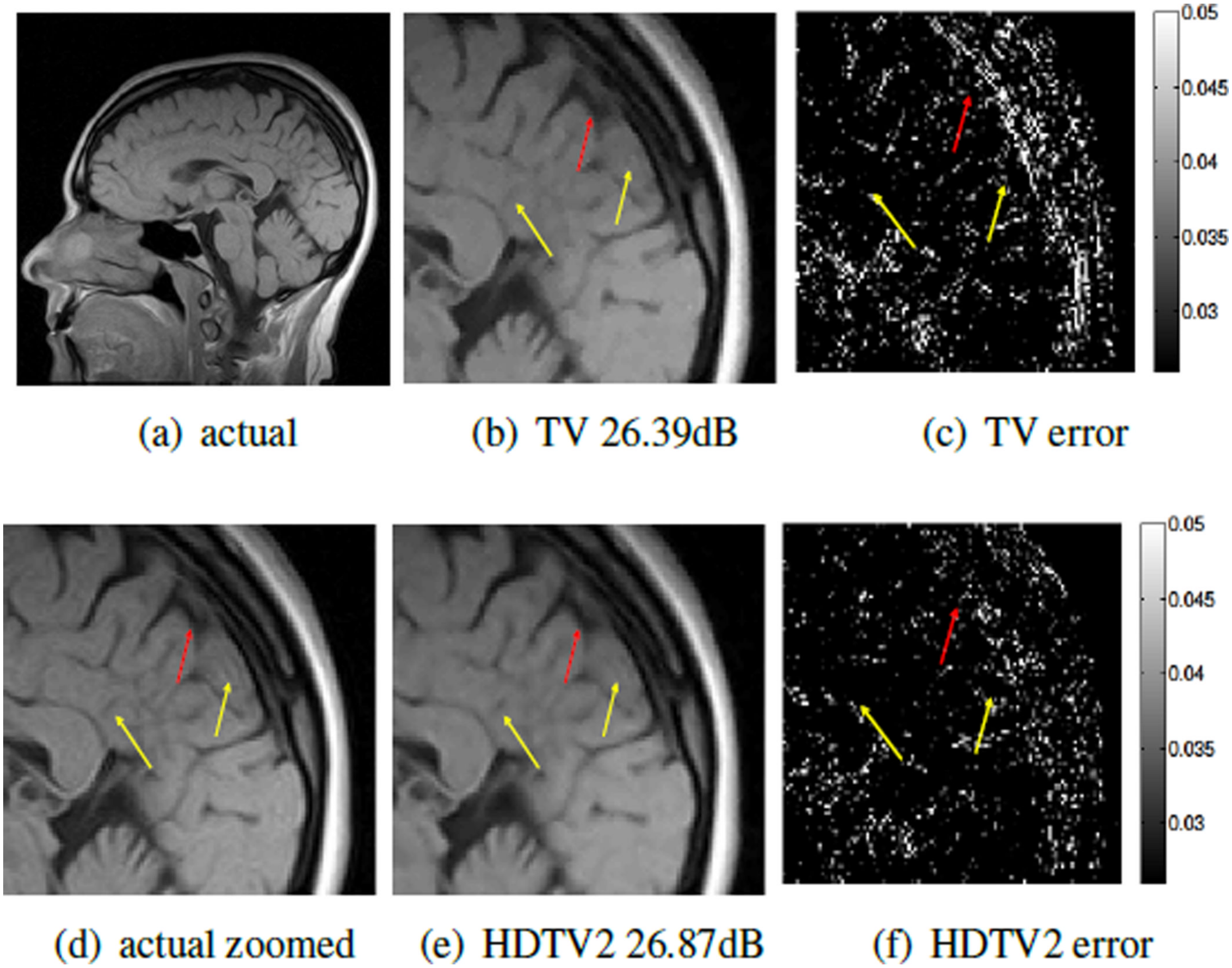

Figure 3.

Compressed sensing recovery of a brain MR image from undersampled and noisy data. The acceleration factor is 1.65 with 25dB additive Gaussian noise. (a) and (d) Original image. (b) and (c) TV reconstruction and error image. (e) and (f) HDTV2 reconstruction and error image.