IMAGE RECOVERY USING IMPROVED TOTAL VARIATION REGULARIZATION

Yue Hu* and Mathews Jacob†

*Department of Electrical and Computer Engineering, University of Rochester, NY, USA
†Department of Biomedical Engineering, University of Rochester, NY, USA

ABSTRACT

We introduce generalized regularization functionals to overcome the practical problems associated with current total variation (TV) penalty. Specifically, we extend the TV scheme to higher order derivatives to improve the representation of smoothly varying image regions. In addition, we introduce a rotation invariant anisotropic TV penalty to improve the regularity of the edge contours. The validation of the scheme demonstrates the significantly improved performance of the proposed methods in the context of compressed sensing and denoising.

1. INTRODUCTION

Total variation (TV) regularization is widely used in denoising and inverse problems. Since its performance is comparable to more sophisticated schemes such as x-lets, it is emerging as a very popular tool in practical applications. Inspite of its desirable properties, the TV scheme has some limitations, which restricts its performance in practical applications. The main challenge is its poor approximation property. Steidl et. al has shown that the use of TV for denoising a 1-D signal is equivalent to approximating the original signal as a non-uniform spline of degree zero [1]. Clearly, zeroth order representations are inefficient in representing smoothly varying regions, resulting in staircase artifacts and visually unappealing patchy reconstructions. The use of TV schemes has been shown to prevent smoothing across the edges, while enhancing the contours by smoothing along the edges [2]. However, we observe that the smoothing along strong edge contours are highly attenuated, resulting in reconstructions with irregular contours. While anisotropic TV methods have been demonstrated to provide improved results [3], these schemes are not rotation invariant, thus resulting in irregular contours.

The main focus of this paper is to introduce novel penalty functions that can overcome the above limitations, while retaining the simplicity and desirable properties of standard TV schemes. We start by reinterpreting the standard TV penalty as a group separable L_1 norm of the oriented first order derivatives. Using this novel interpretation, we extend the regularization functional to higher degree derivatives, thus improving the ability to accurately recover smoothly varying image regions. Thanks to the steerability of higher degree

derivatives, the higher degree schemes have closed form expressions and efficient algorithms, similar to the standard TV schemes. The use of these penalties significantly minimizes the staircasing artifacts and considerably improves the visual quality in practical applications. To improve the regularity of the contours, we introduce a novel rotation invariant and anisotropic TV scheme by formulating the penalty as the separable L_1 norm of the oriented derivatives. Eventhough the new penalties do not have an analytical expression similar to standard TV schemes, we introduce efficient majorize-minimize algorithms that exploit the steerability of the derivatives to solve for the minimizer. We verified the utility of the new algorithms in practical applications, which indicate a significant improvement in performance over state of the art methods. We observe that the improved approximation order and fidelity of contours translate to better reconstructions.

2. THEORY

We focus on the recovery of an image $f : \mathbb{C}^2 \to \mathbb{C}$ from its finite number of noisy linear measurements $\mathbf{b} = \mathcal{A}(f) + \mathbf{n}$, where $\mathcal{A}$ is a linear operator and $\mathbf{n}$ is Gaussian white noise of standard deviation σ. The general practice is to pose the inversion as an optimization problem:

$$\hat{f}(\mathbf{r}) = \arg\min_f \ \|\mathcal{A}(f) - b\|_2 + \lambda \ \mathcal{J}(f). \qquad (1)$$

The optimal λ is chosen such that $\|A(\hat{f}) - b\|_2 = \sigma^2$ and the regularizer $\mathcal{J}(f)$ is a convex functional of $f(\mathbf{r})$. A popular choice is

$$\mathcal{J}(f) = \int_{\mathbf{r}} |\nabla f(\mathbf{r})| \ d\mathbf{r},$$

where $\nabla f(\mathbf{r}) = (f_x(\mathbf{r}), f_y(\mathbf{r}))$ is the gradient vector. We will now reinterpret the TV functional as a group separable L_1 norm of the oriented derivatives of f.

Proposition 2.1. *The gradient based regularizer, specified by*

$$\mathcal{J}(f) = \int_{\mathbf{r}} |\nabla f(\mathbf{r})| \ d\mathbf{r}, \qquad (2)$$

is a group separable penalty of the oriented derivatives of f:

$$\int_{\mathbf{r}} |\nabla f(\mathbf{r})| \ d\mathbf{r} = \int_{\mathbb{R}} \sqrt{\frac{1}{\pi} \int_0^{2\pi} |f_{\theta,1}(\mathbf{r})|^2 d\theta} \ d\mathbf{r}, \qquad (3)$$

where $f_{\theta,1}$ is the oriented derivative along $\mathbf{u}_\theta = (\cos(\theta), \sin(\theta))$; (i.e., $f_{\theta,1} = \frac{\partial}{\partial\gamma} f(r + \gamma\mathbf{u}_\theta)$).

Proof. The first order oriented derivative is rotation steerable:

$$f_{\theta,1}(\mathbf{x}) = f_x \cos(\theta) + f_y \sin(\theta). \qquad (4)$$

Substituting for $f_{\theta,1}$ in (3) and integrating, we get

$$\frac{1}{\pi} \int_0^{2\pi} |f_{\theta,1}|^2 \, d\theta = |f_x|^2 + |f_y|^2 = |\nabla f|^2 \qquad (5)$$

$\square$

Note from (3) that the TV penalty involves the L_1 norm of $\|f_{\theta,1}\|_{L_2[0,2\pi]}$. Such $L_1 - L_2$ penalties are used to exploit the group structure in the coefficients [4]; this approach promotes the joint sparsity of the coefficients, combined using the L_2 norm. Since the TV penalty promotes the joint sparsity of all the oriented derivatives, we term this penalty as the isotropic TV penalty.

2.1. Isotropic higher degree TV (HDTV) penalty

Based on the above re-interpretation, we introduce the isotropic higher degree TV (HDTV) regularizer as

$$\mathcal{J}_n(f) = \int_{\mathbf{r}} \|f_{\theta,n}(\mathbf{r})\|_{L_2[0,2\pi]} \, d\mathbf{r}, \qquad (6)$$

where $f_{\theta,n}(\mathbf{r})$ is the nth degree oriented derivative of f:

$$f_{\theta,n} = \partial^n f(r + \gamma \, \mathbf{u}_\theta) / \partial \gamma^n. \qquad (7)$$

Since the only functions for which all angular derivatives vanish are the polynomials of degree $n-1$, the above functionals have small kernels. Hence, the use of these operators will results in well-posed reconstructions. The use of higher degree derivatives will enable the representation of the signal as piecewise polynomials, thus providing representations with improved approximation properties. We will now exploit the rotation steerability of derivative operators to obtain simple expressions for the regularizer.

The nth order derivatives are steerable [5]:

$$f_{\theta,n}(\mathbf{r}) = \sum_{i=0}^{n} \binom{n}{i} \frac{\partial^n f(\mathbf{r})}{\partial x^i \partial y^{n-i}} \cos(\theta)^i \sin(\theta)^{n-i} \qquad (8)$$

This property enables us to obtain analytical expressions for the new penalties as in (5). For example, the second order derivative in a specified orientation can be expressed as $f_{\theta,2}(\mathbf{r}) = \mathbf{r}(\theta)^H \mathbf{f}_2$, where $\mathbf{f}_2 = [f_{xx}, f_{xy}, f_{yy}]^T$ and $\mathbf{r}_\theta = [\cos(\theta)^2, 2\sin(\theta)\cos(\theta), \sin(\theta)^2]^H$. Thus, we obtain $\mathcal{J}_2(f) = \int \|f(\mathbf{r})\|_{L_2[0,2\pi]} d\mathbf{r}$, where

$$\|f\|_{L_2[0,2\pi]} = \frac{1}{2}\sqrt{3f_{xx}^2 + 3f_{yy}^2 + 4f_{xy}^2 + 2\Re\left(f_{xx}f_{yy}\right)} \quad (9)$$

Thus, the HD-TV criterion can be simplified to an analytical expression involving the nth order partial derivatives, thanks to the steerability of the derivatives. This enables the development of a simple algorithm, discussed in Section 3. The above expression can also be written as

$$\|f\|_{L_2[0,2\pi]} = \sqrt{\mathbf{f}_2(\mathbf{r})^H \, \mathbf{C}_2 \, \mathbf{f}_2(\mathbf{r})}, \qquad (10)$$

where

$$\mathbf{C}_2 = \frac{1}{\pi} \int_0^{2\pi} \mathbf{r}_2(\theta) \mathbf{r}_2(\theta)^H d\theta = \frac{1}{4} \begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix} \qquad (11)$$

2.2. Anisotropic rotation invariant HDTV penalty

The TV regularization can be considered as a 1-D heat flow at each spatial point/pixel, orthogonal to the direction of the gradient [2]. This approach enhances the edges and improves the regularity of the contours by smoothing orthogonal to the edges. However, the strength of the smoothing along the contours is proportional to $1/\|\nabla f\|$, resulting in little or no smoothing along the contours of strong edges. An anisotropic version of the TV penalty was introduced in [3] to improve the contour regularity

$$\mathcal{J}(f) = \int_{\mathbb{R}^2} |f_{xx}(\mathbf{r})| \, d\mathbf{r} + \int_{\mathbb{R}^2} |f_{yy}(\mathbf{r})| \, d\mathbf{r}. \qquad (12)$$

The separable nature of the criterion ensures that the penalty will continue to smooth along vertical direction, even on very strong horizontal edges and vice-versa [3]. However, the main drawback of this scheme is the lack of rotation invariance; rotating the image by an arbitrary angle will alter the smoothing properties.

We introduce a novel rotation invariant anisotropic penalty:

$$\mathcal{G}_n(f) = \frac{1}{\pi} \int_{\mathbb{R}^2} \int_0^{2\pi} |f_{\theta,n}(\mathbf{r})| \, d\theta \, d\mathbf{r}, \qquad (13)$$

to improve the contour regularity. Note that the new criterion is the fully separable L_1 norm of the oriented derivatives. Since this functional is not group separable similar to the standard TV penalty, a strong singularity along a specific orientation will not reduce the smoothing along other directions. Unfortunately, the above penalty does not have analytical expressions, similar to isotropic HDTV regularizer. However, we will introduce an efficient majorize minimize algorithm to solve the regularized reconstruction problem.

3. ALGORITHM

We focus on the group separable penalty, before developing the algorithm for the anisotropic TV.

3.1. Isotropic HDTV

Assuming the mth iteration as f_m, we majorize (10) as

$$J(f) \le J\left(f_m\right) + \int_{\mathbf{r}} \mathbf{f}_n(\mathbf{r})^H \mathbf{D}_{m,n}(\mathbf{r}) \, \mathbf{f}_n(\mathbf{r}) \, d\mathbf{r}, \qquad (14)$$

where the spatially varying weighting matrix $D_{m,n}(\mathbf{r}) = \phi_{m,n}(\mathbf{r})\mathbf{C}_n$. Here, the spatially varying modulating term $\phi_{m,n}(\mathbf{r})$

$$\phi_{m,n}(\mathbf{r}) = 1/2\sqrt{\mathbf{f}_n(\mathbf{r})^H \mathbf{C}_n \, \mathbf{f}_n(\mathbf{r})}. \qquad (15)$$

is inversely proportional to the oriented energy at that specified location ($\|f_{\theta,n}(\mathbf{r})\|_{L_2[0,2\pi]}$): This term is recomputed at each iteration, denoted by m. The spatial modulation

by $\phi_{m,n}(\mathbf{r})$ suppresses the regularization in regions with strong n^{th} order singularities, thus enabling the preservation of edges/ridges in the image. Assuming f_m to be the mth iteration, we minimize the right hand side of (14)

$$f_{m+1}(r) = \arg\min_{\mathbf{f}} \int_{\mathbf{r}} \mathbf{f}_n(\mathbf{r})^H \mathbf{D}_{m,n}(\mathbf{r}) \mathbf{f}_n(\mathbf{r})\, d\mathbf{r} + \lambda \|A(f) - b\|^2,$$
(16)

where $\mathbf{D}_{m,n}(r) = \phi_{m,n}(\mathbf{r})\mathbf{C}_n$. We solve (16) using a simple conjugate gradients algorithm. The gradient of the above expression is given by

$$\nabla C_m = \mathbf{h}_n(\mathbf{r})^H * \mathbf{D}_{\phi_m}(\mathbf{r})\, \mathbf{f}_n(\mathbf{r}) + \lambda A^* \left(A(f) - b\right), \quad (17)$$

where $\mathbf{h}_n(\mathbf{r})$ is the $M \times 1$ vector of differential operators. For example, $\mathbf{h}_2 = [\frac{\partial^2}{\partial xx}, \frac{\partial^2}{\partial yy}, \frac{\partial}{\partial x \partial y}]^T$. Thus the gradient of the first term is given by

$$\underbrace{\left[\frac{\partial^2}{\partial xx}, \frac{\partial^2}{\partial xy}, \frac{\partial^2}{\partial yy}\right]^H}_{h(\mathbf{r})^H} * \phi_m(\mathbf{r}) \frac{1}{4} \underbrace{\begin{bmatrix} 3 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}}_{\mathbf{C}} \underbrace{\begin{bmatrix} f_{xx}(\mathbf{r}) \\ f_{xy}(\mathbf{r}) \\ f_{yy}(\mathbf{r}) \end{bmatrix}}_{\mathbf{f}_n(\mathbf{r})}.$$
(18)

The application of this algorithm to the first order TV results in the standard iterative reweighted TV implementation.

3.2. Anisotropic HDTV

We majorize the anisotropic HDTV penalty as

$$\mathcal{G}_n(f) \le \mathcal{G}_n(f_m) + \frac{1}{\pi} \int_{\mathbb{R}^2} \int_0^{2\pi} \psi_m(r, \theta)\, |f_{\theta,n}(\mathbf{r})|^2\, d\theta\, d\mathbf{r},$$
(19)

where the modulation function is specified by

$$\psi_m(\mathbf{r}, \theta) = \frac{1}{2\,|f_{\theta,n}(\mathbf{r})|}. \quad (20)$$

The angular integral in the second term can be simplified using the steerability of $f_{\theta,n}(\mathbf{r})$ as

$$\psi_m(r, \theta)\, |f_{\theta,n}(\mathbf{r})|^2\, d\theta = \mathbf{f}_2(\mathbf{r})^H\, \mathbf{B}_{m,n}(\mathbf{r})\, \mathbf{f}_2(\mathbf{r}), \quad (21)$$

where

$$\mathbf{B}_{m,n}(\mathbf{r}) = \frac{1}{\pi} \int_0^{2\pi} \mathbf{r}_n(\theta)\, \psi_m(\mathbf{r}, \theta)\, \mathbf{r}_n(\theta)^H d\theta \quad (22)$$

The entries of the matrix $\mathbf{B}_{m,n}(r)$ are modulated depending on the orientation of the singularity, resulting in anisotropic smoothing. This approach is very different from the uniform scaling the entries of $\mathbf{C}$ to obtain $\mathbf{D}_{m,n}(\mathbf{r}) = \phi_{m,n}(\mathbf{r})\mathbf{C}$. The anisotropic scheme computes the entries of $\mathbf{B}_{m,n}(r)$ using (22); we discretize $\psi_{m,n}(\mathbf{r}, \theta)$ on a uniform angular grid and evaluate the angular integral as a summation. Once $\mathbf{B}_{m,n}(r)$ is known, we determine the mth iterate as

$$f_{m+1}(r) = \arg\min_{\mathbf{f}} \int_{\mathbf{r}} \mathbf{f}_n(\mathbf{r})^H \mathbf{B}_{m,n}(\mathbf{r}) \mathbf{f}_n(\mathbf{r})\, d\mathbf{r} + \lambda \|A(f) - b\|^2,$$
(23)

We solve for the minimum of the above expression using conjugate gradients algorithm.

	lena	peppers	MRI	MRA
isotropic TV1	21.14	28.86	29.02	30.20
anistropic TV1	21.95	29.51	**29.99**	30.49
isotropic TV2	20.08	27.78	22.29	26.58
anisotropic TV2	**23.02**	**29.59**	26.60	**31.30**

Table 1. SNRs of the compressed sensing reconstructed images using different methods. It is observed that the anisotropic TV2 provides the best reconstructions in most cases, except for the MRI image which is almost piecewise constant.

4. RESULTS

We compare the performance of the novel TV schemes in the context of two challenging applications: compressed sensing and denoising. The compressed sensing schemes focus on the recovery of the images from their Fourier samples, acquired on random locations. We set the acceleration as 6.25 and the SNR of the measurements to be 60 dB. Table 1 indicates the SNRs of several reconstructed images. It is seen that the anisotropic second order TV provided the best SNR for all images, except for the brain MRI image; the brain image is almost piecewise constant, which is efficiently captured by anisotropic first order TV. It is interesting to see that the anisotropic method provided almost a 1dB improvement in SNR over the isotropic version in this case. Figs 1 and 2 shows the reconstructed brain MRA and Lena images. Note that the anisotropic second order TV provides smoother and more accurate reconstructions, resulting in higher SNR.

We compare the novel TV schemes with several state-of-

(a) original   (b) isotropic TV1: 30.20dB

(c) isotropic TV2: 26.58dB   (d) anisotropic TV2: 31.30dB
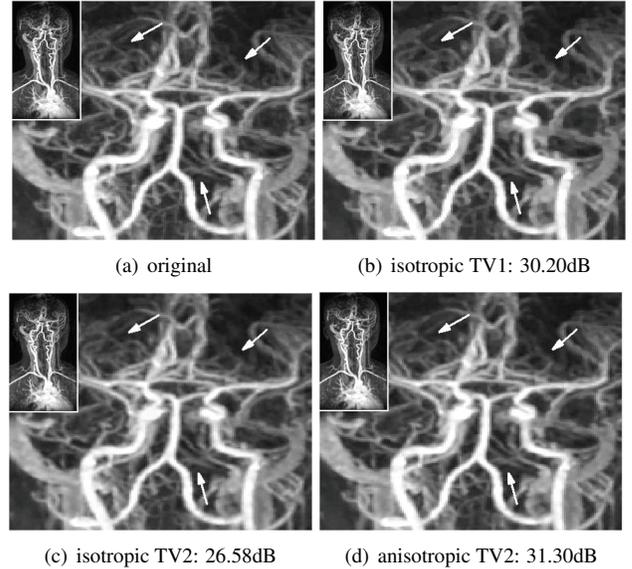
Fig. 1. Comparison of the compressed sensing reconstructions of an MRA image using several TV schemes. The original reconstructions are shown in the inset, while zoomed versions are shown in the main figures. It is seen that the standard TV (isotropic TV1) distorts the small vessels, while the anisotropic TV2 preserves these vessels more accurately. The anisotropic version provides a 4.7 dB improvement over isotropic TV2.

(a) original   (b) isotropic TV: 21.14dB

(c) isotropic TV2: 20.08dB   (d) anisotropic TV2: 23.02dB
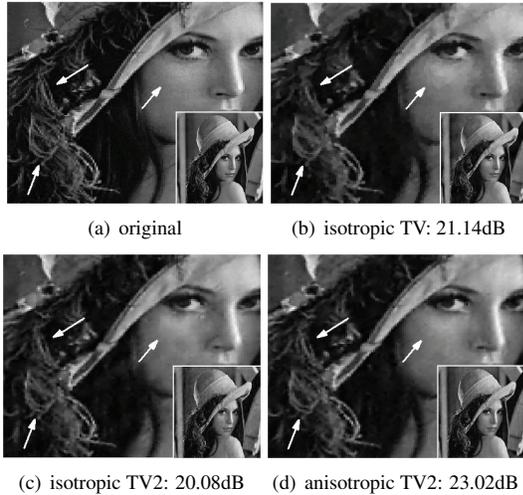
Fig. 2. Comparison of the compressed sensing reconstructions of Lena image. The full reconstructions are shown in the inset. The second order TV schemes preserves the smooth regions and the hairs, resulting in improved SNR.

the-art denoising methods on six images in Table 2. The parameters are chosen so that the reconstructed SNR matches the input SNR. It is seen that the second order anisotropic TV scheme provides the best reconstructions in almost all the cases. The images recovered using the different algorithms are shown in Fig 3 and Fig 4.

(a) noisy: SNR=15dB   (b) isotropic TV1: SNR=21.66dB

(c) isotropic TV2: SNR=23.18dB   (d) aniso. TV2: SNR=23.59dB
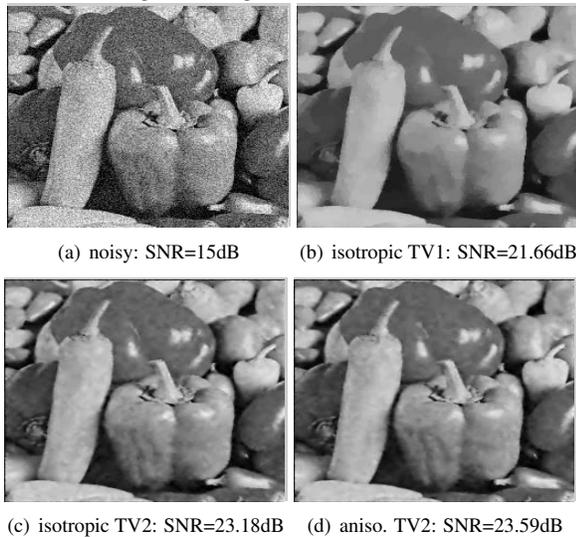
Fig. 3. Comparison of the denoising performance of several TV schemes on peppers image, corrupted with white Gaussian noise (SNR=15 dB). It is observed that the standard TV scheme (isotropic TV1) results in patchy reconstructions. The anisotropic TV2 gives the best SNR; it is around 2dB better than standard TV.

5. CONCLUSION

We introduced novel families of image regularization functions that provide significantly improved performance over classical total variation regularized reconstruction schemes.

SNR(dB)	5dB	15dB	30dB	5dB	15dB	30dB	5dB	15dB	30dB
	Lena			House			microtubule		
Iso. TV1	15.72	20.47	30.01	18.40	23.83	32.21	13.03	18.50	29.92
Aniso. TV1	15.39	20.28	29.98	17.97	23.36	31.89	12.86	18.10	29.64
Iso. TV2	16.46	22.35	33.67	18.85	24.11	34.29	14.35	20.85	33.51
Aniso. TV2	**16.53**	**22.48**	**33.77**	**18.97**	24.46	**34.49**	**14.41**	**20.91**	**33.61**
Curvelet	16.20	21.43	30.80	18.47	**24.55**	33.26	14.28	20.44	31.15
Surelet	16.29	21.12	31.43	18.72	24.19	33.11	14.37	20.44	31.76
	Pepper			Cameraman			microtubule1		
Iso. TV1	16.34	21.66	31.37	**16.68**	21.55	31.47	14.01	18.89	29.17
Aniso. TV1	15.88	21.20	31.14	16.38	21.36	31.30	13.64	18.52	28.85
Iso. TV2	16.46	23.18	34.41	16.46	22.67	34.40	15.41	21.29	32.83
Aniso. TV2	**17.13**	**23.59**	**34.57**	16.59	**22.91**	**34.56**	**15.50**	**21.44**	**32.89**
Curvelet	16.99	22.98	31.93	16.38	21.80	32.19	15.06	20.59	30.53
Surelet	16.85	22.72	32.58	16.59	21.92	32.47	10.81	19.59	31.30

Table 2. SNR of the denoised images, recovered using different methods.

(a) original   (b) noisy: SNR=15dB   (c) iso. TV1:18.89dB

(d) anis. TV1:18.52dB   (e) iso. TV2:21.29dB   (f) anis. TV2:21.44dB
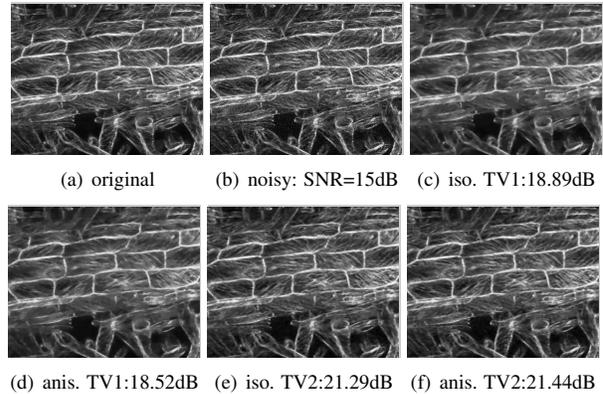
Fig. 4. Comparison of the denoising performance of several TV schemes on the microtubules image, corrupted with white Gaussian noise (SNR=15 dB). The anisotropic TV2 gives the best SNR; it is around 2dB better than standard TV.

We also developed efficient iterative algorithms to solve the optimization problem, using the steerability of the differential operators. Experimental results demonstrate the significant improvement in performance over standard TV schemes.

6. REFERENCES

[1] G. Steidl, S. Didas, and J. Neumann, "Relations between higher order tv regularization and support vector regression," *Scale-Space and PDE Methods in Computer Vision*, Jan 2005.

[2] P. Kornprobst, R. Deriche, and G. Aubert, "Non-linear operators in image restoration," in *cvpr*. Published by the IEEE Computer Society, 1997, p. 325.

[3] S. Osher and S. Esedoglu, "Decomposition of images by the anisotropic rudin-osher-fatemi model," *Comm. Pure Appl. Math*, vol. 57, no. 12, pp. 1609–1626, 2004.

[4] S. Wright and R. Nowak, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Processing*, vol. 57, pp. 2479 – 2493, 2009.

[5] M. Jacob and M. Unser, "Design of steerable filters for feature detection using canny-like criteria," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 8, pp. 1007–19, Aug 2004.