A DIRECTIONAL TOTAL VARIATION

İlker Bayram

Electronics and Telecommunication Eng. Dept.
Istanbul Technical University
ibayram@itu.edu.tr

Mustafa E. Kamasak

Computer Eng. Dept.
Istanbul Technical University
kamasak@itu.edu.tr

ABSTRACT

Total variation (TV) is an isotropic image prior that penalizes the abrupt changes in the images in all directions. In this paper, we modify TV so as to make it more suitable for images with a dominant direction. Specifically, we describe the implementation of a directional TV, and we demonstrate its utility for image denoising. We show that image denoising with the directional TV prior can be more effective compared to the regular (isotropic) TV for images with a dominant direction.

Index Terms— total variation, directional total variation, image denoising

1. INTRODUCTION

Total variation (TV) penalizes abrupt changes in images. It is a very effective signal prior for piecewise smooth images as in Fig. 1a. However, TV is an isotropic functional and is not very suitable for images with a dominant direction, like the one in Fig. 1b. For such images, one could, in principle, scale the image in order to reduce the dominance of the direction. However, for discrete-space images, scaling requires interpolation and therefore it is not very feasible. In this paper, we describe a different approach to define a directional TV. We also study a related denoising problem for discrete-space images and provide an algorithm for its solution.

The total variation (TV) of a discrete-space image f is defined as,

$$\mathrm{TV}(f) = \sum_{i,j} \sqrt{\big(\Delta_1 f(i,j)\big)^2 + \big(\Delta_2 f(i,j)\big)^2} \qquad (1)$$

where Δ_1 and Δ_2 denote horizontal and vertical difference operators, (possibly) defined as,

$$\Delta_1 f(i,j) = f(i,j) - f(i-1,j), \qquad (2)$$

$$\Delta_2 f(i,j) = f(i,j) - f(i,j-1). \qquad (3)$$

We can rewrite this as,

$$\mathrm{TV}(f) = \sum_{i,j} \|\Delta f(i,j)\|_2 = \sum_{i,j} \sup_{t \in B_2} \langle \Delta f(i,j), t \rangle \qquad (4)$$

(a) (b)

Fig. 1. Total variation is a simple and effective prior for piecewise smooth images as in (a). We describe a directional total variation for images with a dominant direction as in (b).

where Δ is the linear operator defined as,

$$\Delta f(i,j) = \begin{pmatrix} \Delta_1 f(i,j) \\ \Delta_2 f(i,j) \end{pmatrix} \qquad (5)$$

and B_2 is the unit ball of the ℓ_2 norm. Henceforth, we use Δ to denote the matrix that represents the linear mapping defined in (5).

Total variation is isotropic because it is invariant under a rotation of the image (or, equivalently, the components of Δf). This is a consequence of the ℓ_2 norm (or B_2) appearing in (4). We can obtain a directional total variation by replacing B_2 with some other set. In particular, if we use an ellipse, $E_{\alpha,\theta}$ oriented along the angle θ, with a unit length minor axis and a major axis of length $\alpha > 1$, (see Fig.2), the resulting norm

$$\mathrm{TV}_{\alpha,\theta}(f) = \sum_{i,j} \sup_{t \in E_{\alpha,\theta}} \langle \Delta f(i,j), t \rangle \qquad (6)$$

is more sensitive to variations along θ.

Given this new total variation, we would like to have algorithms that use this pseudo-norm as a regularizer. In this paper, we study the denoising problem,

$$f^* = \operatorname*{argmin}_f \frac{1}{2} \|y - f\|_2^2 + \lambda \, \mathrm{TV}_{\alpha,\theta}(f), \qquad (7)$$

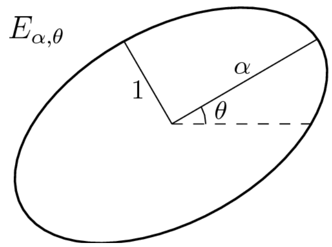


Fig. 2. The ellipse $E_{\alpha,\theta}$ used to define the directional TV norm.

where y is the given noisy image. Specifically, we derive an algorithm that solves this problem and apply it to images to demonstrate its utility.

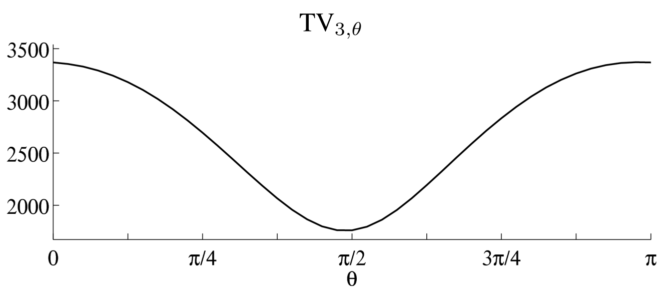


Fig. 3. The directional TV norm of the image in Fig. 1b as a function of θ. Here, α is set to 3. The function has a minimum around $\pi/2$, indicating a dominant vertical appearance.

1.1. Previous Work

Total variation was proposed to be used as an objective function for denoising by Rudin, Osher and Fatemi in [1]. They also describe a method, based on a PDE that solves the proposed optimization problem. Chambolle, in [2], characterizes the solution of the isotropic version of (7) using a certain projection (essentially deriving Prop. 1) and devises an algorithm to realize the projection. Beck and Teboulle [3] derive a similar algorithm and show how it can be further accelerated by using information from different iterates. Esedoğlu and Osher [4] define other general directional total variations by using shapes other than the B_2 ball in (4). They study the properties of the minimizers of the resulting denoising problem (7), however, they do not propose algorithms similar to the one described in this paper.

2. THE DIRECTIONAL TV DENOISING PROBLEM

2.1. Characterizing the Solution of the Problem

In order to characterize the solution to (7), let us define a vector field with two components as

$$v(i,j) = \begin{pmatrix} v_1(i,j) \\ v_2(i,j) \end{pmatrix}. \tag{8}$$

Using $v(i,j)$, we can now write

$$\mathrm{TV}_{\alpha,\theta}(f) = \sup_{v(i,j)\in E_{\alpha,\theta}} \langle \Delta\, f, v\rangle \tag{9}$$

The following characterization can be found in [2, 4].

Proposition 1. For

$$v^* = \operatorname*{argmin}_{v(i,j)\in E_{\alpha,\theta}} \left\| y - \lambda\, \Delta^T\, v \right\|_2^2, \tag{10}$$

set $P_y = \lambda \Delta^T\, v^*$. Then, $y - P_y$ minimizes (7).

The problem is essentially equivalent to a projection. Let us now make a few definitions and transform the problem. Define the rotation and scaling matrices R_θ, Λ_α as,

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \Lambda_\alpha = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}. \tag{11}$$

Using these, $E_{\alpha,\theta}$ and B_2 are related as $E_{\alpha,\theta} = R_\theta\, \Lambda_\alpha\, B_2$. Let us now define the operators $\mathbf{R}_\theta$, $\mathbf{\Lambda}_\alpha$, that act on the vector fields as,

$$\big(\mathbf{R}_\theta\, v\big)(i,j) = R_\theta\, \big(v(i,j)\big), \quad \big(\mathbf{\Lambda}_\alpha\, v\big)(i,j) = \Lambda_\alpha\, \big(v(i,j)\big). \tag{12}$$

We remark that $\mathbf{R}_\theta^T = \mathbf{R}_{-\theta}$, $\mathbf{\Lambda}_\alpha^T = \mathbf{\Lambda}_\alpha$. These facilitate the computation of $\mathrm{TV}_{\alpha,\theta}$ as,

$$\mathrm{TV}_{\alpha,\theta}(f) = \sup_{v(i,j)\in R_\theta\, \Lambda_\alpha\, B_2} \langle \Delta\, f, v\rangle \tag{13}$$

$$= \sup_{v(i,j)\in B_2} \langle \Delta\, f, \mathbf{R}_\theta\, \mathbf{\Lambda}_\alpha\, v\rangle \tag{14}$$

$$= \sup_{v(i,j)\in B_2} \langle \mathbf{\Lambda}_\alpha\, \mathbf{R}_{-\theta} \Delta\, f, v\rangle. \tag{15}$$

Hence, we have

Corollary 1. For

$$v^* = \operatorname*{argmin}_{v(i,j)\in B_2} \left\| y - \lambda\, \Delta^T\, \mathbf{R}_\theta\, \mathbf{\Lambda}_\alpha\, v \right\|_2^2, \tag{16}$$

set $P_y = \lambda\, \Delta^T\, \mathbf{R}_\theta\, \mathbf{\Lambda}_\alpha\, v^*$. Then, $y - P_y$ minimizes (7).

This problem is also equivalent to a projection. However, this time the projections involve disks rather than ellipses. Although one can devise an algorithm based on projection onto ellipses (see for instance [5, 6]), we think that the algorithm described below, which makes use of projections onto disks, is simpler.

Description of the Algorithm

For simplicity, we take $\lambda = 1$ without loss of generality.

Our goal is to minimize the function

$$C(v) = \|y - A\,v\|_2^2 \tag{17}$$

where $A = \Delta^T\,\mathbf{R}_\theta\,\mathbf{\Lambda}_\alpha$, subject to $v(i,j) \in B_2$ for all (i,j). Suppose now that we have $v^{(k)}$ at the k^{th} iteration. We would like to find $v^{(k+1)}$ where each component $v^{(k+1)}(i,j)$ belongs to B_2, such that $C\big(v^{(k+1)}\big) \le C\big(v^{(k)}\big)$. Suppose that ρ is a constant for which $\rho\,I - \Delta\,\Delta^T$ is positive semi-definite. Then, it follows that, $(\alpha^2\rho)\,I - A^T\,A$ will also be positive semi-definite. It can be shown, by a majorization argument [7], that for

$$C^{(k)}\big(v\big) = \Bigg\| \underbrace{\Bigg[v^{(k)} + \frac{1}{2\alpha^2\,\rho}\,A^T\big(y - A\,v^{(k)}\big)\Bigg]}_{\tilde{v}^{(k)}} - v \Bigg\|_2^2, \tag{18}$$

if $C^{(k)}\big(v\big) < C^{(k)}\big(v^{(k)}\big)$ for some v, then we also have $C\big(v\big) < C\big(v^{(k)}\big)$. Therefore, we might as well consider reducing the function $C^{(k)}$ at the k^{th} iteration, subject to $v(i,j) \in B_2$. Notice that $C^{(k)}(v)$ is separable with respect to the indices. The minimizer, subject to $v(i,j) \in B_2$, is given by,

$$v^*(i,j) = \tilde{v}^{(k)}(i,j)\,\frac{1}{\max\big\{\|\tilde{v}^{(k)}(i,j)\|_2, 1\big\}}. \tag{19}$$

Algorithm 1 provides the pseudo-code for the general case, with arbitrary λ, α, θ parameters to obtain f^* in (7).

Algorithm 1 Directional TV Denoising

Input: λ, y, α, θ from (7)
Output: f^* as in (7)
Require: ρ, a constant s.t. $\rho\,I - \Delta\,\Delta^T$ is psd.

$v_n(i,j) \leftarrow 0$ for $n = 1, 2$
$\kappa \leftarrow 1/(2\,\rho\,\alpha^2\,\lambda^2)$
$A \leftarrow \lambda\,\Delta^T\,\mathbf{R}_\theta\,\mathbf{\Lambda}_\alpha$

for $iter = 1$ to $MaxIter$ **do**
  $\tilde{v} \leftarrow v^{(k)} + \kappa\,A^T\big(y - A\,v^{(k)}\big)$
  for all (i,j) pairs **do**
    $v(i,j) \leftarrow \tilde{v}^{(k)}(i,j)\,\frac{1}{\max\big\{\|\tilde{v}^{(k)}(i,j)\|_2, 1\big\}}$
  end for
end for
$f^* \leftarrow y - A\,v$

Remark 1. A similar algorithm may be derived using the frameworks presented in [8, 9].

3. EXPERIMENTS

To demonstrate the utility of the proposed directional total variation, we present three experiments. All images used in the experiments are grayscale and their intensity value are normalized so that their pixel intensities have values between [0,1]. Independent identically distributed (iid) Gaussian noise is added to the images to obtain the noisy observations. The noise is zero mean and spatially uncorrelated. The noise has different standard deviations (σ) for the experiments, which are mentioned below. For each experiment, the amount of TV prior (λ) for both regular and directional TV is optimized for minimum RMSE.

Experiment 1. Gaussian noise with σ=0.1 is added to the directional texture image shown in Fig. 4a to obtain the noisy 'observation' shown in Fig. 4b (RMSE = 0.1009). For $\alpha = 1$ (this corresponds to regular TV), the denoised image is shown in Fig. 4c, where RMSE = 0.0489. For $\alpha = 5$, $\theta = \pi/2$, (directional TV), the denoised image is shown in Fig. 4d, where the RMSE = 0.0429. We observe that directional TV performs better for this image, both in terms of RMSE and visual appearance. □

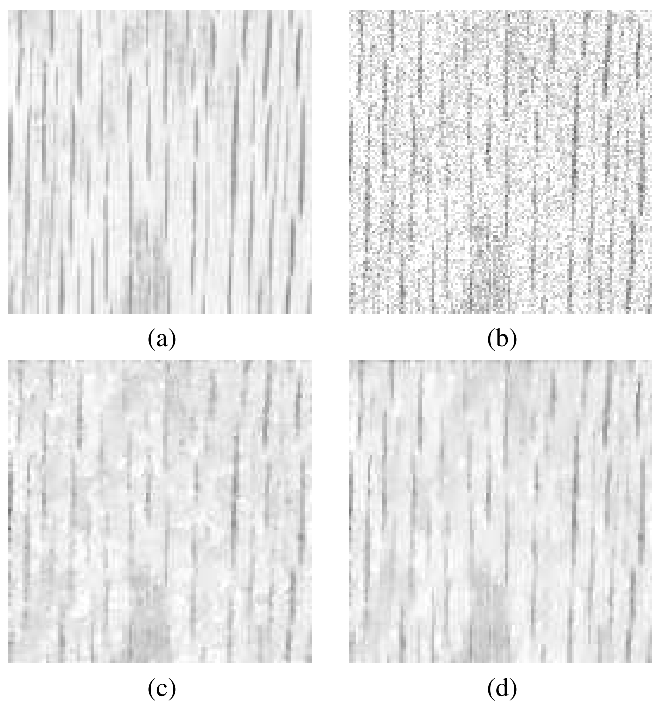


Fig. 4. Texture image (a) Clean Image (b) Noisy Observation (RMSE=0.1009) (c) Denoised with a TV prior (RMSE=0.0489) (d) Denoised with a directional TV prior (RMSE=0.0429).

Experiment 2. Gaussian noise with $\sigma = 0.1$ is added to the spaghetti image shown in Fig. 5a. The obtained noisy image is shown in Fig. 5b (RMSE = 0.1009). The noisy image is denoised using both regular and directional TV priors. The result of the image denoising using TV prior is shown in Fig. 5(c) with RMSE=0.0431. Denoised image with directional TV prior is shown in shown in Fig. 5(d) with RMSE=0.0269. The parameters used in the directional TV prior are $\alpha = 5$ and $\theta = \pi/4$.

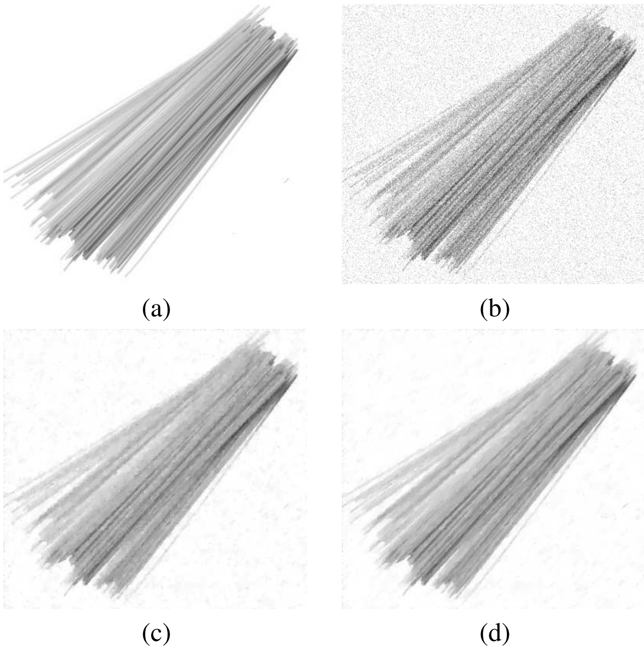


Fig. 5. Spaghetti image (a) Clean Image (b) Noisy Observation (RMSE=0.1002) (c) Denoised with a TV prior (RMSE=0.0431) (d) Denoised with a directional TV prior (RMSE=0.0269).

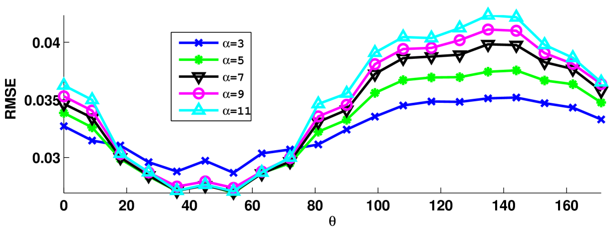


Fig. 6. RMSE values for denoised spaghetti images obtained with directional TV with different α and θ parameters.

We also investigated the effects of the $\{\alpha, \theta\}$ parameters of the directional TV in terms of their performance for denoising. Fig. 6 depicts the RMSE curves for different α values as a function of θ. For the direction parameter θ, we see that for all values of α, the smallest RMSE value is obtained around $\theta = \pi/4$, which is roughly the dominant direction in the spaghetti image. For the parameter α, we observe that small values cannot achieve good denoising. On the other hand, high values of α lead to high RMSE when the TV prior direction (θ) is deviates from the correct value. Therefore, for α, one can talk about a trade-off between denoising performance and dependency on the prior direction information. To that end, for the spaghetti image shown in Fig. 5(a), the value $\alpha = 5$ appears to be a good choice, since it can effectively remove noise is sufficiently prone to deviations from the true θ. □

Experiment 3. A final experiment is performed in order to compare the denoising performance of directional TV prior with regular TV. For this purpose, zero mean Gaussian noise with different σ values is added to the pipe image that is shown in the first row of Fig. 7. Four different noise levels at $\sigma = \{0.1, 0.15, 0.2, 0.25\}$ are used. Each column in Fig. 7 corresponds to one of the noise levels, whose σ values are given at the bottom of each column in the figure. The noisy observations of the different noise levels are shown in the second row of Fig. 7. In this figure, the denoised images using TV prior are shown in the third row, and the denoised images using directional TV prior are shown in the last row.

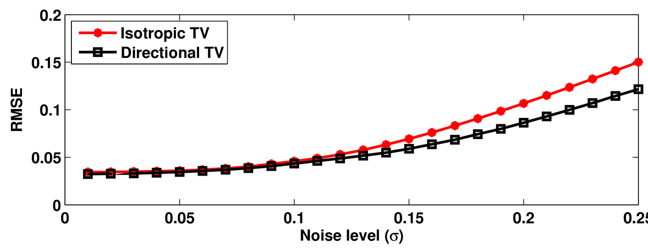


Fig. 8. RMSE values for the denoised pipe image for different noise levels.

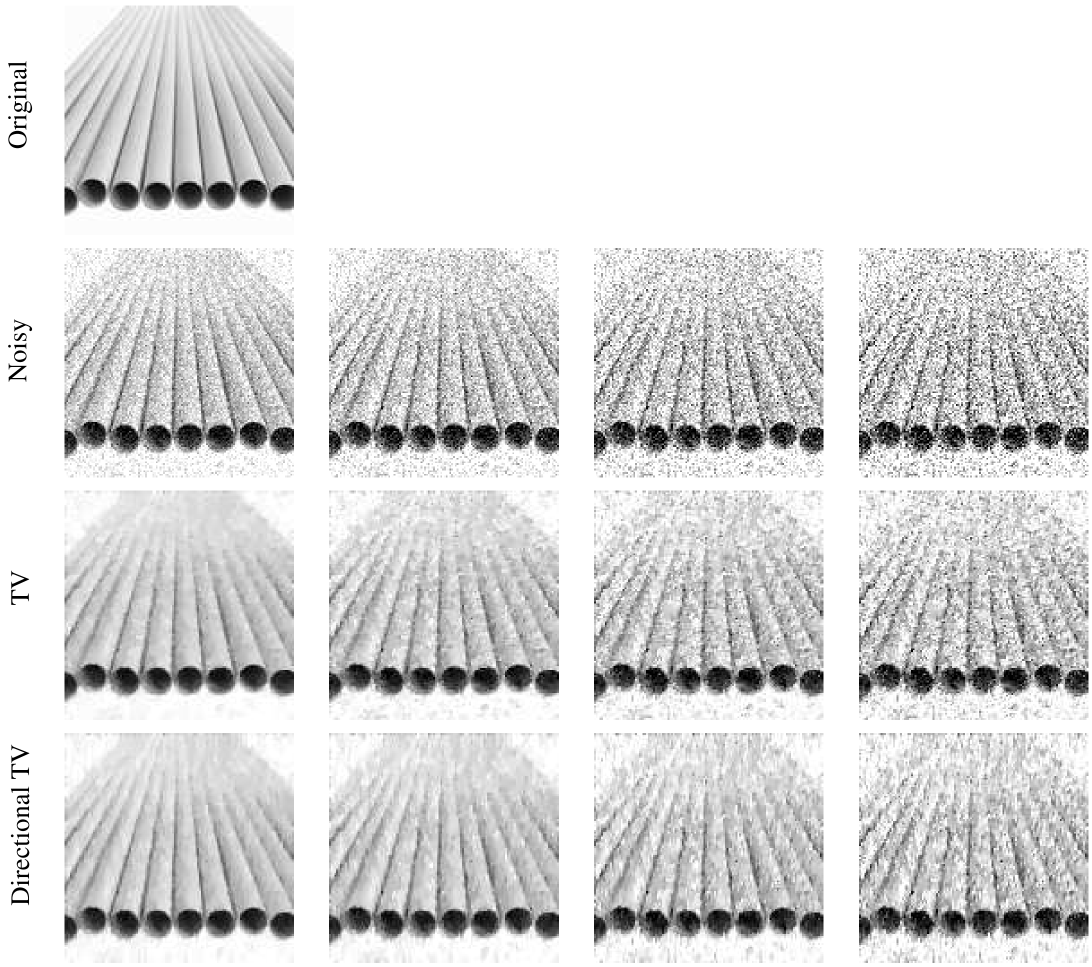
The (best) RMSE values for the denoised images using TV and directional TV priors are shown in Fig. 8 at different noise levels. We see that, as the noise level increases, denoising with directional TV prior outperforms the regular TV. Since denoising in the presence of high noise depends more heavily on the prior, this implies that the directional TV is a more suitable prior for this image, despite the existence of structures whose directions differ from the dominant direction. □

4. CONCLUSION

We described a modification to the regular TV so as to make it more sensitive to a certain direction. As a sample application, we used the directional TV as a prior in an image denoising problem. We demonstrated that for images with a dominant direction, the directional TV is a more suitable prior than regular TV. We also discussed the effects of the parameters in the introduced directional TV.

5. REFERENCES

[1] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.



(a) $\sigma = 0.1$ (b) $\sigma = 0.15$ (c) $\sigma = 0.2$ (d) $\sigma = 0.25$

Fig. 7. Pipe image denoised using TV and directional TV for different values of the noise level (σ). First and second rows show the original and noisy images respectively. Images denoised by TV and directional TV are shown in the second and third row respectively. Columns illustrate the results at different noise levels whose corresponding σ values are given at the bottom of each column.

[2] A. Chambolle, “An algorithm for total variation minimization and applications,” *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, January-March 2004.

[3] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE Trans. Image Processing*, vol. 18, no. 11, pp. 2419–2434, 2009.

[4] S. Esedoğlu and S. Osher, “Decomposition of images by the anisotropic Rudin - Osher - Fatemi model,” *Comm. Pure and Appl. Math*, vol. 57, pp. 1609–1626, 2004.

[5] A. Lin and S.-P. Han, “On the distance between two ellipsoids,” *SIAM J. on Optimization*, vol. 13, no. 1, pp. 298–308, 2002.

[6] İ. Bayram and M. Kamasak, “Directional total variation minimization,” in *Proc. Conf. of IEEE Signal Processing and Communications Applications (SIU)*, 2012.

[7] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, “Majorization-minimization algorithms for wavelet-based image restoration,” *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[8] E. Esser, X. Zhang, and T. Chan, “A general framework for a class of first order primal-dual algorithms for TV minimization,” CAM Report 09-67, UCLA Center for Applied Math., 2009.

[9] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.