

# Directional Total Variation

İlker Bayram and Mustafa E. Kamasak

**Abstract**—This paper introduces a “directional total variation” (TV) where the gradients are weighted depending on their direction. This directional TV has increased (and tunable) sensitivity to variations at a selected direction. In order to demonstrate the utility of the directional TV, we consider an image denoising formulation. This formulation requires the realization of the ‘proximal map’ of the directional TV. Therefore, it is relevant for more general inverse problem settings as well. We derive an algorithm that solves the problem and use the algorithm to study the effects of the parameters of the directional TV.

**Index Terms**—Total variation, directional total variation, image denoising.

## I. INTRODUCTION

Total variation (TV) is a measure of the variations in an image. It penalizes local changes in the image regardless of their direction. To that effect, TV has proved its utility as a simple prior for piecewise smooth images with no globally dominant direction. However, TV is not very suitable for images with a dominant direction (such as the one in Fig. 1(a)). In this paper, we consider a variation of TV that provides a similarly simple prior for such images. We also describe an algorithm that solves a related denoising problem. The denoising problem is equivalent to realizing the ‘proximal operator’ [1] associated with the proposed directional total variation, therefore it can be useful in more general inverse problems formulations [1]–[3].

In brief, the total variation of an image is computed by summing the norms of the gradient. Essentially, our idea is to first weight the gradients depending on their directions and then sum their norms. Such a modification leads to increased sensitivity to variations at a certain direction.

### Previous Work

TV was first proposed to be used in image denoising/restoration problems by Rudin, Osher and Fatemi [4]. The authors also provided a denoising algorithm that is based on a time-evolution equation. Later, Chambolle [5] considered a cost function composed of a quadratic and a TV term. He showed that the problem is essentially equivalent to a projection and proposed an algorithm to obtain the minimizer (see also [6] for an updated algorithm). Beck and Teboulle [7] derived a similar projection algorithm (equivalent to the one in [6]) and showed how to accelerate it by using information from different iterations. Also of relevance is the work of Esedoğlu and Osher [8] where they described a general TV and studied its properties. We note that the directional TV described in this letter does fit into the context described in [8]. However, [8] does not provide a denoising algorithm similar

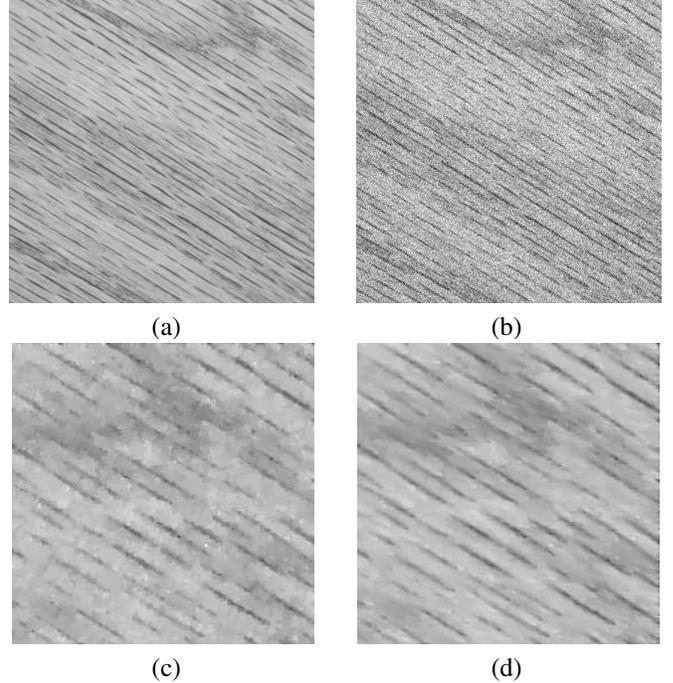


Fig. 1. (a) Texture image with a dominant direction. (b) Noisy texture image used in the experiments (Gaussian, iid, SNR = 12.72 dB) (c) Details of the TV denoised image (SNR = 22.52 dB). (d) Details of the Directional TV denoised image (SNR = 23.86 dB).

to the one in this letter. Two recent studies were brought to our attention by the reviewers [9], [10]. In [9], the authors introduce a one-parameter family of regularizers through the use of ‘support functions’ and present algorithms for typical inverse problems. Reference [10] proposes a generalization of TV so as to incorporate higher order derivatives. Although we also view the directional TV as a support function, neither the proposed directional TV nor the regularizers in [9], [10] are special cases of each other.

### Outline

In Section II, we discuss the isotropic (regular) TV and provide the details on how to render it direction-sensitive. Section III describes the denoising formulation and derives an algorithm that solves the formulation. Experiments that evaluate the proposed algorithm are provided in Section IV. Section V concludes the letter.

## II. DIRECTIONAL TOTAL VARIATION

### A. Isotropic Total Variation

TV of a discrete-space image  $f(i, j)$  is defined as,

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

where  $\Delta_1$  and  $\Delta_2$  denote horizontal and vertical difference operators:<sup>1</sup>

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

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

Note that TV in (1) can be rewritten as:

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

where  $\Delta$  is the linear operator defined as,

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

and  $B_2$  is the unit ball of the  $\ell_2$  norm.

The regular TV is isotropic because it is invariant to rotations of the components of  $\Delta f$ . This is a consequence of the  $\ell_2$  norm (or  $B_2$ ) appearing in (4).

### B. Directional Total Variation

It is possible to obtain a directional total variation by replacing  $B_2$  with some other set. In particular, an ellipse can be used to increase the sensitivity to changes towards a certain direction. If  $B_2$  is replaced by the ellipse,  $E_{\alpha, \theta}$  that is oriented along the angle  $\theta$ , with a unit length minor axis and a major axis of length  $\alpha > 1$  (see Fig. 2(a)), the pseudo norm defined by

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

is more sensitive to variations along the major axis.

The directional TV of the image shown in Fig. 1(a), is computed for different values of  $\theta$ , while keeping  $\alpha = 3$ . The resulting function of  $\theta$  is depicted in Fig. 2(b). Note that the minimum value of the directional TV is obtained around  $\theta = -\pi/6$ , which coincides with the dominant direction of the texture within the image. Also, observe that the curve in Fig. 2(b) is smooth – slight deviations of  $\theta$  from the minimum value do not dramatically increase the value of the directional TV. This can be a desired property especially when faithful direction information is not available.

### III. THE DIRECTIONAL TV DENOISING PROBLEM

Often, inverse problem formulations/algorithms require the realization of the ‘proximal operator’ associated with the signal prior [1]. Realizing the proximal operator is also of interest per se, since it is equivalent to solving a denoising problem. Specifically, in our case, we need to consider an objective function with a directional TV:

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

where  $y$  is the given noisy image (observation) and  $f^*$  is the denoised estimate.

In order to find a solution to (7), let us define a vector field with two components as  $v(i, j) = [v_1(i, j), v_2(i, j)]^T$ . Using

<sup>1</sup>We assume that the image is zero outside its boundaries.

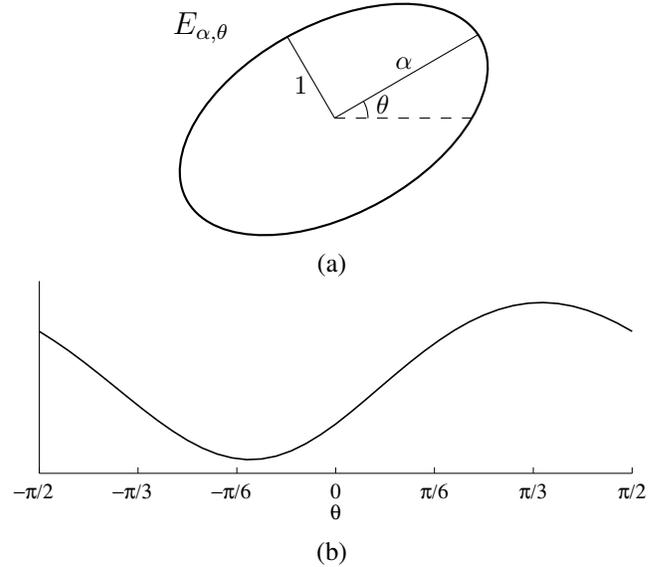


Fig. 2. (a) The ellipse  $E_{\alpha, \theta}$  used to define the directional TV norm. (b) The value of the directional  $\text{TV}_{3, \theta}$  of the image shown in Fig. 1(a) as  $\theta$  varies.

$v(i, j)$ , the directional TV can be written as:

$$\text{TV}_{\alpha, \theta}(f) = \sum_{i,j} \sup_{v(i,j) \in E_{\alpha, \theta}} \langle \Delta f, v \rangle. \quad (8)$$

Results of [5] and [8] indicate that  $y - \Delta^T v^*$  minimizes (7), where

$$v^* = \underset{v(i,j) \in E_{\alpha, \theta}}{\operatorname{argmin}} \|y - \Delta^T v\|_2^2. \quad (9)$$

The problem is essentially equivalent to a projection. Let  $R_\theta$  and  $\Lambda_\alpha$  denote the rotation and scaling matrices respectively as:

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad \Lambda_\alpha = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}. \quad (10)$$

Note that  $E_{\alpha, \theta}$  and  $B_2$  are related as  $E_{\alpha, \theta} = R_\theta \Lambda_\alpha B_2$ . Furthermore, let us now define the operators  $\mathbf{R}_\theta$ ,  $\mathbf{\Lambda}_\alpha$ , that act on the vector fields as  $(\mathbf{R}_\theta v)(i, j) = R_\theta (v(i, j))$  and  $(\mathbf{\Lambda}_\alpha v)(i, j) = \Lambda_\alpha (v(i, j))$ . Using the fact that  $\mathbf{R}_\theta^T = \mathbf{R}_{-\theta}$  and  $\mathbf{\Lambda}_\alpha^T = \mathbf{\Lambda}_\alpha$ ,  $\text{TV}_{\alpha, \theta}$  can be expressed in the following forms:

$$\begin{aligned} \text{TV}_{\alpha, \theta}(f) &= \sum_{i,j} \sup_{v(i,j) \in R_\theta \Lambda_\alpha B_2} \langle \Delta f, v \rangle \\ &= \sum_{i,j} \sup_{v(i,j) \in B_2} \langle \Delta f, \mathbf{R}_\theta \mathbf{\Lambda}_\alpha v \rangle \\ &= \sum_{i,j} \sup_{v(i,j) \in B_2} \langle \mathbf{\Lambda}_\alpha \mathbf{R}_{-\theta} \Delta f, v \rangle. \end{aligned} \quad (11)$$

Then, for

$$v^* = \underset{v(i,j) \in B_2}{\operatorname{argmin}} \|y - \lambda \Delta^T \mathbf{R}_\theta \mathbf{\Lambda}_\alpha v\|_2^2, \quad (12)$$

$y - \lambda \Delta^T \mathbf{R}_\theta \mathbf{\Lambda}_\alpha v^*$  minimizes (7). Eq. (12) 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 [11]), we think this approach is simpler.

Henceforth, we take  $\lambda = 1$  for simplicity. Generalization to arbitrary  $\lambda$  is straightforward. Our goal is to minimize the

function

$$C(v) = \|y - Av\|_2^2 \quad (13)$$

where  $A = \Delta^T \mathbf{R}_\theta \mathbf{A}_\alpha$ , subject to  $v(i, j) \in B_2$  for all  $(i, j)$ . Suppose now that we have  $v^{(k)}$  at the  $k^{\text{th}}$  iteration. We would like to find  $v^{(k+1)}$  where each component  $v^{(k+1)}(i, j)$  belongs to  $B_2$ , such that  $C(v^{(k+1)}) \leq C(v^{(k)})$ . Suppose that  $\rho$  is a constant for which  $\rho I - \Delta \Delta^T$  is positive semi-definite (psd). Then, it follows that,  $(\alpha^2 \rho) I - A^T A$  will also be psd. It can be shown, by a majorization argument [3], that for

$$C^{(k)}(v) = \left\| \underbrace{\left[ v^{(k)} + \frac{1}{2\alpha^2 \rho} A^T (y - A v^{(k)}) \right]}_{\tilde{v}^{(k)}} - v \right\|_2^2, \quad (14)$$

if  $C^{(k)}(v) < C^{(k)}(v^{(k)})$  for some  $v$ , then  $C(v) < C(v^{(k)})$ . Therefore, we might as well consider reducing the function  $C^{(k)}$  at the  $k^{\text{th}}$  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,

$$v^*(i, j) = \tilde{v}^{(k)}(i, j) \frac{1}{\max\{\|\tilde{v}^{(k)}(i, j)\|_2, 1\}}. \quad (15)$$

Algorithm 1 provides the pseudo-code for the general case, with arbitrary  $\lambda$ ,  $\alpha$ ,  $\theta$  parameters to obtain  $f^*$  in (7).

---

#### Algorithm 1 Image Denoising Using Directional TV

---

**Input:**  $\lambda, y, \alpha, \theta$  from (7)

**Output:**  $f^*$  as in (7)

**Require:**  $\rho$ , a constant such that  $\rho I - \Delta \Delta^T$  is psd.

$v_n(i, j) \leftarrow 0$  for  $n = 1, 2$

$\kappa \leftarrow 1/(2\rho\alpha^2\lambda^2)$

$A \leftarrow \lambda \Delta^T \mathbf{R}_\theta \mathbf{A}_\alpha$

**for**  $iter = 1$  to  $MaxIter$  **do**

$\tilde{v} \leftarrow v^{(k)} + \kappa A^T (y - A v^{(k)})$

**for all**  $(i, j)$  pairs **do**

$v(i, j) \leftarrow \tilde{v}^{(k)}(i, j) \frac{1}{\max\{\|\tilde{v}^{(k)}(i, j)\|_2, 1\}}$

**end for**

**end for**

$f^* \leftarrow y - A v$

---

#### Accelerating The Algorithm

Algorithm 1 monotonically decreases the cost function in (12). To that end, FISTA [7], and TwIST [12] can be adopted to accelerate the convergence. An alternative is to choose  $\rho$  such that  $2\rho I - \Delta \Delta^T$  is psd (i.e. reduce  $\rho$  by half – please see the requirements in Algorithm 1). With such a choice of  $\rho$ , if  $\rho I - \Delta \Delta^T$  is not psd, the algorithm is no longer a descent algorithm on (12) but it determines a sequence that converges to the minimizer of (12) (see the conditions of convergence for forward-backward algorithms [2]). In our experience, reducing  $\rho$  by half accelerates the algorithm almost as much as TwIST or FISTA.

## IV. EXPERIMENTS

In order to demonstrate the utility of the proposed directional TV, experiments are performed with two images: <sup>2</sup> ‘wood’ (Fig. 1(a)) and ‘bamboo’ (Fig. 3(a)).

**Experiment 1.** We obtain the noisy observation in Fig. 1(b) by adding iid Gaussian noise to the ‘wood’ image (SNR = 12.72 dB). To denoise the observation, we make use of the formulation in (7) for a TV and a directional TV regularizer. In each case, the weight of the prior ( $\lambda$ ) is selected so as to maximize the SNR. For TV, we obtain an SNR of 22.52dB. For the directional TV with  $\alpha = 5$  and  $\theta = -0.15\pi$ , the SNR is 23.86 dB. We note that this value of  $\theta$  minimizes the function in Fig. 2(b). Details for both regular TV and the introduced directional TV are shown in Fig. 1(c,d). We observe that the artifacts which are clearly visible in the TV denoised image are much weaker in the directional TV denoised image.  $\square$

**Experiment 2.** We added iid Gaussian noise to the clean bamboo image in Fig. 3(a) to obtain the noisy observation in Fig. 3(b) (SNR = 13.55 dB). Once again we use the formulation in (7), but this time we use the anisotropic TV<sup>3</sup> and the directional TV as regularizers. The  $\lambda$  parameters are chosen so as to maximize the SNR of the denoised image. In this case, anisotropic TV denoising yields an SNR of 17.02 dB. Choosing the parameters of the directional TV as  $\alpha = 5$ ,  $\theta = 0$ , we obtain a denoised image with SNR equal to 20.65 dB. Detail images are shown in Fig. 3(c,d). Observe that the horizontal sticks look very smooth in the directional TV denoised image. On the other hand, the string that ties the bamboos together (appearing as a vertical feature) is washed out in the directional TV denoised image in comparison to the anisotropic TV denoised image. This is also evident from the absolute error images shown in Fig. 3(e,f), where high errors appear as black (the images are contrast enhanced). We observe that the directional TV denoised image deviates from the original around the vertical feature, which does not agree with the provided  $\theta$  direction. On the other hand, the error distribution for anisotropic TV is more uniform.  $\square$

#### Effect of Parameter Selection in Directional TV

We also investigated the effect of  $\{\alpha, \theta\}$  parameters of the directional TV in terms of denoising performance. Fig. 4 illustrates the SNR values of the denoised images with different  $\alpha$  values as a function of  $\theta$  for the setup in Experiment 2 (for each fixed  $\{\alpha, \theta\}$ , we search for the  $\lambda$  value that yields the best SNR). We can see that for all values of  $\alpha$ , the highest SNR value is obtained around  $\theta = 0$ , which gives the dominant direction in the bamboo image. Small values of  $\alpha$  lead to poorer denoising performance, whereas higher values of  $\alpha$  lead to low SNR if the TV prior direction ( $\theta$ ) is not chosen correctly. Therefore, one can talk about a trade-off

<sup>2</sup>Matlab code used in the experiments can be found at <http://web.itu.edu.tr/ibayram/DTV/>.

<sup>3</sup>In terms of the terminology of Section II, the anisotropic TV of  $f$  is given by  $\|\Delta_1 f\|_1 + \|\Delta_2 f\|_1$ . This corresponds to taking a square as the ‘Wulf shape’ in [8].

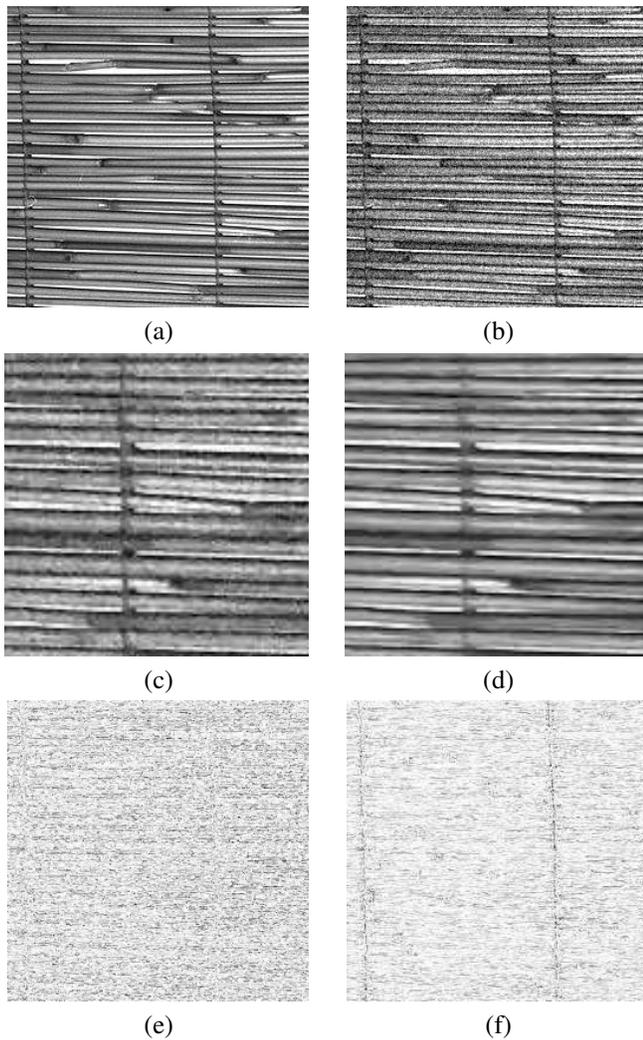


Fig. 3. (a) Clean ‘Bamboo’ image, (b) Noisy bamboo image (SNR = 13.55 dB), (c) Detail from the image denoised using the anisotropic TV as a regularizer (SNR = 17.02 dB), (d) Detail from the image denoised using the proposed directional TV as a regularizer (SNR = 20.65 dB), (e) Absolute differences of the original image and the anisotropic TV denoised image, (f) Absolute differences of the original image and the directional TV denoised image, The images in (e,f) are contrast enhanced for visibility and also inverted so that black indicates high values.

between performance and robustness with respect to the prior information.

## V. CONCLUSION

In this paper, a directional TV and its use for image denoising is described. For images with a dominant direction, denoising with directional TV is shown to outperform isotropic/anisotropic TV in terms of SNR and visual quality. We also observed that the parameters of the directional TV, namely  $\{\alpha, \theta\}$ , have to be chosen carefully. Specifically, if the prior information regarding the dominant direction (i.e.  $\theta$ ) is not reliable, then a relatively small  $\alpha$  should be chosen.

Note that, the directional TV described in this letter can not be effective for denoising arbitrary images with no single dominant direction. However, we can regard an arbitrary image as a collection of patches where each patch has a single dominant direction. Therefore, it is possible to extend the

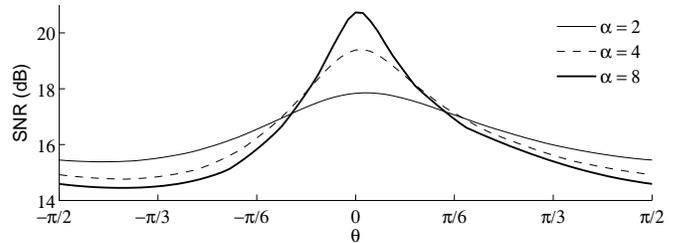


Fig. 4. SNR values for denoised bamboo images obtained with directional TV with different  $\alpha$  and  $\theta$  parameters.

concept of directional TV by locally changing the directions of the ellipses depending on the gradients. This approach could improve/complement the work of Berkels et al. [13] where rectangular objects are assumed to be present in the given image. We plan to pursue this idea in future work.

## REFERENCES

- [1] P. L. Combettes and J. C. Pesquet, “Proximal methods in signal processing,” in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. B. et al., Ed. Springer, 2011.
- [2] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Nov. 2005.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] —, “Total variation minimization and a class of binary MRF models,” *Lecture notes in computer science*, vol. 3757, pp. 136–152, 2005.
- [7] 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.
- [8] 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.
- [9] B. Goldlucke and D. Cremers, “Introducing total curvature for image processing,” in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, 2011.
- [10] K. Bredies, K. Kunisch, and T. Pock, “Total generalized variation,” *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 492–526, 2010.
- [11] A. Lin and S.-P. Han, “On the distance between two ellipsoids,” *SIAM J. on Optimization*, vol. 13, no. 1, pp. 298–308, 2002.
- [12] J. M. Bioucas-Dias and M. A. T. Figueiredo, “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration,” *IEEE Trans. Image Processing*, vol. 16, no. 12, pp. 2992–3003, Dec. 2007.
- [13] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf, “Cartoon extraction based on anisotropic image classification,” in *Proc. Vision, modeling and visualization*, 2006.