

Total Directional Variation for Video Denoising [★]

Simone Parisotto¹ and Carola-Bibiane Schönlieb²

¹ CCA, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK
sp751@cam.ac.uk

² DAMTP, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK
cbs31@cam.ac.uk

Abstract. In this paper we propose a variational approach for video denoising, based on a total directional variation (TDV) regulariser proposed in [21,20] for image denoising and interpolation. In the TDV regulariser, the underlying image structure is encoded by means of weighted derivatives so as to enhance the anisotropic structures in images, e.g. stripes or curves with a dominant local directionality. For the extension of TDV to video denoising, the space-time structure is captured by the volumetric structure tensor guiding the smoothing process. We discuss this and present our whole video denoising workflow. The numerical results are compared with some state-of-the-art video denoising methods.

Keywords: Total directional variation · Video denoising · Anisotropy · Structure tensor · Variational methods.

1 Introduction

Video denoising refers to the task of removing noise in digital videos. Compared to image denoising, video denoising is usually a more challenging task due to the computational cost in processing large data and the redundancy of information, i.e. the expected similarity between two consecutive frames that should be inherited by the denoised video. A straightforward approach to video denoising is to denoise each frame of the video independently, by using the broad literature on image denoising methods, see e.g. [26,24,6,23,4,11,15,3,19,12,21]. Computational cost is then stratified across image frames by sequentially processing them, which is seen as an advantage. However, a significant disadvantage of this frame-by-frame processing is the appearance of flickering artefacts and post-processing motion compensation step may be required [18,2].

In recent years different approaches have been proposed for solving the video denoising problem: we refer to the introduction of [1] for an extensive survey. Notably, patch-based approaches are usually considered among the most promising video denoising methods in that they are able to achieve qualitatively good denoising results. For example, V-BM3D is the 3D extension of the BM3D collaborative filters [10]: without inspecting the motion time-consistency, V-BM3D

[★] SP acknowledges UK EPSRC grant EP/L016516/1 for the CCA DTC. CBS acknowledges support from Leverhulme Trust project on Breaking the non-convexity barrier, EPSRC grant Nr. EP/M00483X/1, the EPSRC Centre EP/N014588/1, the RISE projects CHiPS and NoMADS, the CCIMI and the Alan Turing Institute.

independently filters 2D patches resulting similar in the 3D spatio-temporal neighbourhood domain. As mentioned in [1], while generally receiving good denoising results, the problem of flickering still occurs in V-BM3D. For this reason, authors of V-BM3D developed an extension, called V-BM4D, where the patch-similarity is explored along space-temporal blocks defined by a motion vector, see [17]. Similarly, in [5] the authors propose to group patches via an optical flow equation based on [28] and implemented in [25]. In these approaches, while the incorporation of motion helps to provide consistency in time, denoising results also suffer from the lack of accuracy in the estimated motion. A possible way to avoid the motion estimation is to consider 3D rectangular patches so as to inherently model the 3D structure and motion in the spatio-temporal video dimensions, based on the fact that rectangular 3D patches are less repeatable than motion-compensated patches. However, such approach is not efficient for uniform motion or homogeneous spatial patterns, cf. the discussion on this topic in [1]. Motivated by this reasoning the authors of [1] introduce a Bayesian patch-based video denoising approach with rectangular 3D patches modelled as independent and identically distributed samples from an unknown *a priori* distribution: then each patch is denoised by minimising the expected mean square error. Other approaches in video denoising are the straightforward extension of the Rudin-Osher-Fatemi (ROF) model [24] to 3D data, by using a spatio-temporal total variation (referred in the next as ROF 2D+t), the joint video denoising with the computation of the flow [7] and CNN approaches [13].

Scope of the paper. In this paper we propose an extension of the recently introduced *total directional variation* (TDV) regulariser [21,20] for video denoising, via the following variational regularisation model:

$$u^* \in \arg \min_u \left(\text{TDV}(u, \mathbf{M}) + \frac{\eta}{2} \|u - u^\diamond\|_2^2 \right), \quad (1)$$

where u^* is the denoised video, \mathbf{M} is a weighting field that encodes directional features in two spatial and one temporal dimension, $\eta > 0$ is the regularisation parameter and u^\diamond is a given noisy video. The model (1) will be made more precise in the next sections where we mainly focus on its discrete and numerical aspects. In order to accommodate for spatial-temporal data, we consider here a modification of the TDV regulariser given in [21,20] that derives directionality in the temporal dimension. Differently from the patch-based approach, we compute for each voxel the vector field of the motion, to be encoded as a weight in the TDV regulariser. With this voxel-based approach we will reduce the flickering artefact which appears in patch-based approaches due to the patch selection, especially in regions of smooth motion. Results are presented for a variety of videos corrupted with Gaussian white noise.

Organisation of the paper. This paper is organised as follows: in Section 2 we describe the estimation of the vector fields, the TDV regulariser and the variational model to be minimised; in Section 3 we describe the optimisation method for solving the TDV video denoising problem and comment on the selection

of parameters; in Section 4 we show denoising results on a selection of videos corrupted with Gaussian noise of varying strength.

2 Total directional variation for video denoising

Let $\bar{u} : \Omega \times [1, \dots, T] \rightarrow \mathbb{R}_+^C$ be a clean video and Ω a spatial, rectangular domain indexed by $\mathbf{x} = (x, y)$, with number of T frames and C colours. Let u^\diamond be a corrupted version of \bar{u} in each space-time voxel $(\mathbf{x}, t) \in \Omega \times [1, \dots, T]$ by i.i.d. Gaussian noise n of zero mean and (possibly known) variance $\zeta^2 > 0$:

$$u^\diamond(\mathbf{x}, t) = \bar{u}(\mathbf{x}, t) + n(\mathbf{x}, t), \quad \forall (\mathbf{x}, t) \in \Omega \times [1, \dots, T]. \quad (2)$$

In what follows, we propose to compute a denoised video $u^* \approx \bar{u}$ by solving

$$u^* \in \arg \min_u \left(\text{TDV}(u, \mathbf{M}) + \frac{\eta}{2} \|u - u^\diamond\|_2^2 \right), \quad (3)$$

where $\text{TDV}(u, \mathbf{M})$ is the proposed total direction regulariser w.r.t. a weighting field \mathbf{M} , both specified in the next sections, and $\eta > 0$ a regularisation parameter.

2.1 The directional information

In order to capture directional information of u in (3), we eigen-decompose the two-dimensional structure tensor [27] in each coordinate plane.

To do so, we first construct the 3D structure tensor: let $\rho \geq \sigma > 0$ be two smoothing parameters, K_σ, K_ρ be the Gaussian kernels of standard deviation σ and ρ , respectively, and let $u_\sigma = K_\sigma * u$. Then the 3D structure tensor reads as

$$\mathbf{S} := K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma) = \begin{pmatrix} u_{\sigma,\rho}^{x,x} & u_{\sigma,\rho}^{x,y} & u_{\sigma,\rho}^{x,t} \\ u_{\sigma,\rho}^{y,x} & u_{\sigma,\rho}^{y,y} & u_{\sigma,\rho}^{y,t} \\ u_{\sigma,\rho}^{t,x} & u_{\sigma,\rho}^{t,y} & u_{\sigma,\rho}^{t,t} \end{pmatrix}, \quad (4)$$

where $\nabla u_\sigma \otimes \nabla u_\sigma = \nabla u_\sigma \nabla u_\sigma^T$, $u_{\sigma,\rho}^{p,q} := K_\rho * (\partial_p u_\sigma \otimes \partial_q u_\sigma)$ for each $p, q \in \{x, y, t\}$.

For a straightforward application to the TDV regulariser in [21], we extract the 2D sub-tensors of (4), whose eigen-decomposition encodes structural information in each of the coordinate frames spanned by $\{x, y\}$, $\{x, t\}$ and $\{y, t\}$:

$$\begin{aligned} \text{on coordinates } \{x, y\}: \quad \mathbf{S}^{x,y} &= \begin{pmatrix} u_{\sigma,\rho}^{x,x} & u_{\sigma,\rho}^{x,y} \\ u_{\sigma,\rho}^{y,x} & u_{\sigma,\rho}^{y,y} \end{pmatrix} = \lambda_1(\mathbf{e}_1 \otimes \mathbf{e}_1) + \lambda_2(\mathbf{e}_2 \otimes \mathbf{e}_2); \\ \text{on coordinates } \{x, t\}: \quad \mathbf{S}^{x,t} &= \begin{pmatrix} u_{\sigma,\rho}^{x,x} & u_{\sigma,\rho}^{x,t} \\ u_{\sigma,\rho}^{t,x} & u_{\sigma,\rho}^{t,t} \end{pmatrix} = \lambda_3(\mathbf{e}_3 \otimes \mathbf{e}_3) + \lambda_4(\mathbf{e}_4 \otimes \mathbf{e}_4); \\ \text{on coordinates } \{y, t\}: \quad \mathbf{S}^{y,t} &= \begin{pmatrix} u_{\sigma,\rho}^{y,y} & u_{\sigma,\rho}^{y,t} \\ u_{\sigma,\rho}^{t,y} & u_{\sigma,\rho}^{t,t} \end{pmatrix} = \lambda_5(\mathbf{e}_5 \otimes \mathbf{e}_5) + \lambda_6(\mathbf{e}_6 \otimes \mathbf{e}_6). \end{aligned} \quad (5)$$

For each $s \in \{1, \dots, 6\}$, the eigenvector $\mathbf{e}_s = (e_{s,1}, e_{s,2})$ has eigenvalue λ_s . The tangential directions in the 2D planes $\{x, y\}$, $\{x, t\}$ and $\{y, t\}$ are $\mathbf{e}_2, \mathbf{e}_4, \mathbf{e}_6$, respectively, with $\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_5$ the gradient directions, see Fig. 1.

From (5), the ratios between the eigenvalues, called *confidence*, measure the local anisotropy of the gradient on the slices within a certain neighbourhood:

$$a^{x,y} = \frac{\lambda_2}{\lambda_1 + \varepsilon}, \quad a^{x,t} = \frac{\lambda_4}{\lambda_3 + \varepsilon}, \quad a^{y,t} = \frac{\lambda_6}{\lambda_5 + \varepsilon}, \quad \text{with } \varepsilon > 0. \quad (6)$$

Here, $a^{x,y}, a^{x,t}, a^{y,t} \in [0, 1]$ and the closer to 0, the higher is the local anisotropy.

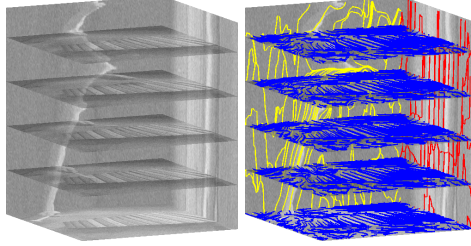


Fig. 1. Left: grey-scale video `xylophone.mp4`, corrupted by Gaussian noise ($\zeta = 20$); right: streamlines of the weighting field with e_2 (blue), e_4 (red) and e_6 (yellow).

2.2 The regulariser

The TDV regulariser is composed of a gradient operator weighted by a tensor \mathbf{M} , whose purpose is to smooth along selected directions. In view of the spatial-temporal data, we extend the natural gradient operator to the Cartesian planes $\{x, y\}$, $\{x, t\}$ and $\{y, t\}$. We will denote with $\tilde{\nabla}$ the concatenation of resulting 2-dimensional gradients. Further, we encode (5) and (6) in \mathbf{M} , leading to the *weighted gradient* $\mathbf{M}\tilde{\nabla}$ for the video function $u = u(x, y, t)$:

$$\mathbf{M}\tilde{\nabla} \otimes u = \underbrace{\begin{pmatrix} a^{x,y} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & a^{x,t} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & a^{y,t} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{M}} \underbrace{\begin{pmatrix} e_{1,1} & e_{1,2} & 0 & 0 & 0 & 0 \\ e_{2,1} & e_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & e_{3,1} & e_{3,2} & 0 & 0 \\ 0 & 0 & e_{4,1} & e_{4,2} & 0 & 0 \\ 0 & 0 & 0 & 0 & e_{5,1} & e_{5,2} \\ 0 & 0 & 0 & 0 & e_{6,1} & e_{6,2} \end{pmatrix}}_{\tilde{\nabla}} \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_x \\ \partial_t \\ \partial_y \\ \partial_t \end{pmatrix} \otimes u \quad (7)$$

$$= (a^{x,y} \nabla_{e_1}^{x,y} u, \nabla_{e_2}^{x,y} u, a^{x,t} \nabla_{e_3}^{x,t} u, \nabla_{e_4}^{x,t} u, a^{y,t} \nabla_{e_5}^{y,t} u, \nabla_{e_6}^{y,t} u)^\top. \quad (8)$$

Note that \mathbf{M} is computed once from the noisy input u^\diamond . For a fixed frame $\{p, q\}$ with $p, q \in \{x, y, t\}$ and direction $\mathbf{z} = (z_1, z_2)$ the gradient $\nabla_{\mathbf{z}}^{p,q} u = \partial_p u \cdot z_1 + \partial_q u \cdot z_2$ is the directional derivative of u along \mathbf{z} w.r.t. the frame $\{p, q\}$. See [22, Fig. 3.12] for more details about this choice. With this notation in place, we consider the *total directional variation* (TDV) regulariser,

$$\text{TDV}(u, \mathbf{M}) = \sup_{\Psi} \left\{ \int_{\Omega} (\mathbf{M}\tilde{\nabla} \otimes u) \cdot \Psi \, d\mathbf{x} \mid \text{for all suitable test functions } \Psi \right\}. \quad (9)$$

By plugging (8) into (9) we reinterpret (9) as a penalisation of the rate of change along e_2, e_4, e_6 , with coefficients $a^{x,y}, a^{x,t}, a^{y,t}$ as bias in the gradient estimation. Note, that while in [21] the TDV regulariser has been proposed for a general order of derivatives, we consider here only a TDV regulariser of first differential order.

2.3 Connections to optical flow

Let $(x, y, t) \in \Omega \times [1, \dots, T]$ be a voxel and $u(x, y, t)$ its intensity in the grey-scale video sequence u . If $u(x, y, t)$ is moved by a small increment $(\delta_x, \delta_y, \delta_t)$ between two frames, then the *brightness constancy* constraint reads

$$u(x, y, t) = u(x + \delta_x, y + \delta_y, t + \delta_t). \quad (10)$$

If u is sufficiently smooth, then the *optical flow* constraint is derived [16,14] as a linearisation of (10) with respect to a velocity field \mathbf{z} :

$$\nabla u(x, y, t)^T \cdot \mathbf{z} = 0, \quad \text{for all } (x, y, t) \in \Omega \times [1, T]. \quad (11)$$

For a specific field $\mathbf{z} = (\tilde{\mathbf{z}}, 1)$ with $\tilde{\mathbf{z}} = (z_1(x, y), z_2(x, y))$, Equation (11) is equivalent to

$$-\partial_t u = \partial_x u \cdot z_1 + \partial_y u \cdot z_2 = \nabla_{\tilde{\mathbf{z}}}^{x,y} u \quad \text{for all } (x, y, t) \in \Omega \times [1, T]. \quad (12)$$

We can now re-write (11) by means of the following velocity vector fields:

$$\begin{aligned} a^{x,y}(e_{1,1}, e_{1,2}, 1), \quad a^{x,t}(e_{3,1}, 1, e_{3,2}), \quad a^{y,t}(1, e_{5,1}, e_{5,2}), \\ (e_{2,1}, e_{2,2}, 1), \quad (e_{4,1}, 1, e_{4,2}), \quad (1, e_{6,1}, e_{6,2}), \end{aligned} \quad (13)$$

leading to

$$\begin{pmatrix} -a^{x,y} \partial_t u \\ -\partial_t u \\ -a^{x,t} \partial_y u \\ -\partial_y u \\ -a^{y,t} \partial_x u \\ -\partial_x u \end{pmatrix} = \begin{pmatrix} a^{x,y} \partial_x u \cdot e_{1,1} + a^{x,y} \partial_y u \cdot e_{1,2} \\ \partial_x u \cdot e_{2,1} + \partial_y u \cdot e_{2,2} \\ a^{x,t} \partial_x u \cdot e_{3,1} + a^{x,t} \partial_t u \cdot e_{3,2} \\ \partial_x u \cdot e_{4,1} + \partial_t u \cdot e_{4,2} \\ a^{y,t} \partial_y u \cdot e_{5,1} + a^{y,t} \partial_t u \cdot e_{5,2} \\ \partial_y u \cdot e_{6,1} + \partial_t u \cdot e_{6,2} \end{pmatrix} = \begin{pmatrix} a^{x,y} \nabla_{e_1}^{x,y} u \\ \nabla_{e_2}^{x,y} u \\ a^{x,t} \nabla_{e_3}^{x,t} u \\ \nabla_{e_4}^{x,t} u \\ a^{y,t} \nabla_{e_5}^{y,t} u \\ \nabla_{e_6}^{y,t} u \end{pmatrix}. \quad (14)$$

Here, the right-hand side of (14) encodes the components that we aim to penalise in (8). Thus, the penalisation of (8) is equivalent to the penalisation of the left-hand side of (14), assumed (11) holds with velocity fields in (13). Note that the weights $a^{x,y}$, $a^{x,t}$ and $a^{y,t}$ add a contribution in the direction of the gradients e_1 , e_3 , e_5 , respectively.

2.4 The minimisation problem

We aim to find the denoised video u^* from the noisy input video u^\diamond by solving the TDV – L^2 minimisation problem (3). For the numerical optimisation of (3) we use a primal-dual scheme [9]. For this, we rewrite (3) as a saddle point problem for the operator $\mathcal{K} := \mathbf{M} \nabla$, whose adjoint will be denoted by \mathcal{K}^* . In what follows, we denote by u the primal variable, y the dual variable, f^* the Fenchel conjugate of f , by g the fidelity term and by $\sigma, \tau > 0$ the dedicated parameters of the primal-dual algorithm, see [9] for more details on the primal-dual schemes in image processing and [21] for their application to variational problems with TDV regulariser. The resulting saddle-point problem reads

$$u^* \in \arg \min_u \max_y \left(\langle \mathcal{K}u, y \rangle - \underbrace{\delta_{\{\|\cdot\|_{2,\infty} \leq 1\}}(y)}_{f^*(y)} + \underbrace{\frac{\eta}{2} \|u - u^\diamond\|_2^2}_{g(u^\diamond)} \right). \quad (15)$$

In the primal-dual algorithm solving (15) we need the proximal operators:

$$\mathbf{prox}_{\sigma f^*}(y) = \frac{y}{\max\{1, \|y\|_2\}}, \quad \mathbf{prox}_{\tau g}(u) = u + (\mathbf{I} + \tau\eta)^{-1}\tau\eta(u^\diamond - u), \quad (16)$$

where \mathbf{I} is the identity matrix. Note that g is uniformly convex, with convexity parameter η , so the dual problem is smooth. An accelerated version of the primal-dual algorithm can be used in this case, e.g. [8, Alg. 2], starting with $\tau_0, \sigma_0 > 0$ where $\tau_0\sigma_0L^2 \leq 1$ and L^2 is the squared operator norm, $L^2 := \|\mathcal{K}\|^2 \leq 24$ (which holds in connection with the discretisation in (17) and stepsize $h = 1$).

3 The discrete model

In the discrete model, Ω is a rectangular grid of size $M \times N$ and a video \mathbf{u} is a volumetric data of size $M \times N \times T \times C$ (height \times width \times frames \times colours). Here, we consider grey-scale videos ($C = 1$) along the axes $(i, j, k) \in \Omega \times [1, T]$, with $i = 1, \dots, M$, $j = 1, \dots, N$ and $k = 1, \dots, T$. An extension to coloured videos is straightforward by processing each colour channel separately. Here, a fixed $(i, j, k) \in \Omega \times [1, T]$ identifies a voxel in the gridded video domain, i.e. a small cube of size h in each axis direction. Then, $\mathbf{u}_{i,j,k} := \mathbf{u}(i, j, k)$ is the intensity in the voxel (i, j, k) in the grey-scale video sequence \mathbf{u} . The noisy input video is denoted by \mathbf{u}^\diamond as well as the other discrete vectorial quantities, namely $\mathbf{a}^{1,2}, \mathbf{a}^{1,3}, \mathbf{a}^{2,3}$ and $\boldsymbol{\lambda}_s$ for $s = 1, \dots, 6$.

3.1 Discretisation of derivative operators and vector fields

We describe a finite difference scheme on the voxels by introducing the discrete gradient operator $\nabla : \mathbb{R}^{M \times N \times T} \rightarrow \mathbb{R}^{M \times N \times T \times 3}$, with $\nabla = (\partial_1, \partial_2, \partial_3)$ defined via the central finite differences on half step-size and Neumann conditions as

$$\begin{aligned} (\mathbf{u}^1)_{i,j,k} &:= (\partial_1 \mathbf{u})_{i+0.5,j,k} = \begin{cases} \frac{\mathbf{u}_{i+1,j,k} - \mathbf{u}_{i,j,k}}{h}, & \text{if } i = 1, \dots, M-1, \\ 0 & \text{if } i = M; \end{cases} \\ (\mathbf{u}^2)_{i,j,k} &:= (\partial_2 \mathbf{u})_{i,j+0.5,k} = \begin{cases} \frac{\mathbf{u}_{i,j+1,k} - \mathbf{u}_{i,j,k}}{h}, & \text{if } j = 1, \dots, N-1, \\ 0 & \text{if } j = N; \end{cases} \\ (\mathbf{u}^3)_{i,j,k} &:= (\partial_3 \mathbf{u})_{i,j,k+0.5} = \begin{cases} \frac{\mathbf{u}_{i,j,k+1} - \mathbf{u}_{i,j,k}}{h}, & \text{if } k = 1, \dots, T-1, \\ 0 & \text{if } k = T. \end{cases} \end{aligned} \quad (17)$$

Remark 1. While \mathbf{u} lies at the vertices of the discrete grid, $\nabla \mathbf{u}$ lies on its edges. Thus, (17) is advantageous for local anisotropy since it has sub-pixel precision and a more compact stencil radius than the classical forward scheme.

In (7), $\tilde{\nabla} : \mathbb{R}^{M \times N \times T} \rightarrow \mathbb{R}^{M \times N \times T \times 6}$ acts on \mathbf{u} as follows:

$$\left(\tilde{\nabla} \otimes \mathbf{u} \right)_{i,j,k} := (\mathbf{u}^1, \mathbf{u}^2, \mathbf{u}^1, \mathbf{u}^3, \mathbf{u}^2, \mathbf{u}^3)_{i,j,k}^T. \quad (18)$$

Algorithm 1: TDV for video denoising

Input : A video $\mathbf{u}^\diamond \in [0, 255]$ ($M \times N \times 1 \times T$), $\varsigma \in [0, 255]$.
Output : the denoised video \mathbf{u} ;
Parameters: for the primal-dual maxiter, tol; for the variational model: (σ, ρ, η) .

Function TDV_video_denoising:

```

// Compute operators for the weighted derivative
 $[\partial_1, \partial_2, \partial_3] = \text{compute\_derivative\_operator}(M, N, T)$ ;
 $\mathbf{S} = \text{compute\_3D\_structure\_tensor}(\mathbf{u}^\diamond, \sigma, \rho)$ ;
 $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5, \mathbf{e}_6, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6] = \text{eigendecomposition}(\mathbf{S})$ ;
 $[\mathbf{a}^{1,2}, \mathbf{a}^{1,3}, \mathbf{a}^{2,3}] = \text{compute\_anisotropy}(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6)$ ;
 $\mathbf{M} = \text{compute\_weights}(\mathbf{a}^{1,2}, \mathbf{a}^{1,3}, \mathbf{a}^{2,3}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5, \mathbf{e}_6)$ ;

// Proximal operators, adjoints and primal-dual from [8]
 $[\mathcal{K}, \mathcal{K}^*] = \text{compute\_K\_and\_adjoint}(\mathbf{M}, \partial_1, \partial_2, \partial_3)$ ;
 $\text{prox}_{f^*} = \text{@}(\mathbf{y}) \mathbf{y} / \max\{1, \|\mathbf{y}\|_2\}$ ;
 $\text{prox}_g = \text{@}(\mathbf{u}, \tau) \mathbf{u} + (\mathbf{I} + \tau\eta)^{-1} \tau\eta(\mathbf{u}^\diamond - \mathbf{u})$ ;
 $\mathbf{u} = \text{primal\_dual}(\mathbf{u}^\diamond, \mathcal{K}, \mathcal{K}^*, \text{prox}_{f^*}, \text{prox}_g, \text{maxiter}, \text{tol})$ ;

return

```

Any field \mathbf{e}_s with $s = 1, \dots, 6$ and confidence $a^{1,2}, a^{1,3}, a^{2,3}$ will be discretised in the cell centres $(i+0.5, j+0.5, k+0.5)$ of the discrete grid domain. The weighting multiplication in (7) is performed via an intermediate averaging interpolation operator $\mathcal{W} : \mathbb{R}^{M \times N \times T \times 6} \rightarrow \mathbb{R}^{(M-1) \times (N-1) \times (T-1) \times 6}$ that avoids artefacts due to the grid offset: this gives $\mathbf{M}\mathcal{W}\tilde{\nabla} : \mathbb{R}^{M \times N \times T} \rightarrow \mathbb{R}^{(M-1) \times (N-1) \times (T-1) \times 6}$.

3.2 TDV for video denoising

The TDV-based workflow consists of two steps, with pseudo-code in Alg. 1. The first one computes the directions via the eigen-decomposition in (5) while the second one is the primal-dual algorithm [8, Alg. 2], whose stopping criterion is the root mean square difference between two consecutive dual variable iterates.

4 Results

In this section we discuss the numerical results for video denoising obtained with Alg. 1. Considered videos have been taken from a benchmark video dataset³. Each video has values in $[0, 255]$ corrupted with Gaussian noise. We tested different noise levels with standard deviation $\varsigma = [10, 20, 35, 50, 70, 90]$ without clipping the videos so as to conform to the observation model.

³ Videos are freely available: *Salesman* and *Miss America* at www.cs.tut.fi/~foi/GCF-BM3D *Xylophone* in MATLAB; *Water* (re-scaled, grey-scaled and clipped, Jay Miller, CC 3.0) at www.videvo.net/video/water-drop/477; *Franke*'s function (a synthetic surface moving on fixed trajectories: the coloured one changes with the *parula* colormap).

⁴ Results are available at <http://www.simoneparisotto.com/TDV4videodenoising>.

The quality of the denoised result \mathbf{u}^* is evaluated by the peak signal-to-noise ratio (PSNR) value w.r.t. a ground truth video $\bar{\mathbf{u}}$. The model requires the parameters (σ, ρ, η) as input. Once provided, we solve the saddle-point minimisation problem in (15) via the accelerated primal-dual algorithm with $L^2 = \|\mathcal{K}\|^2 = 24$, see [8, Alg. 2]. Here the tolerance for the stopping criterion is fixed to 10^{-4} (on average reached in 300 iterations). However, we experienced faster convergence and similar results with $L^2 \ll 24$ and bigger tolerances, e.g. 10^{-3} .

4.1 Selection of parameters

In the model, \mathbf{u}^* is sensitive to the choice of both (σ, ρ) for the vector fields, and the regularisation parameter η that is chosen according to the noise level. Choosing those parameters by a trial and error approach is computationally expensive and the best parameters may differ, even for videos with the same noise level. In particular, the parameters (σ, ρ) depend on structure in the data, e.g. flat regions versus motions versus small details. Therefore, a strategy for tuning them is needed.

To estimate appropriate values for (σ, ρ, η) that render good results for a variety of videos we compute optimal parameters via line-search for maximising the PSNR for a small selection of video denoising examples for which the ground truth is available. The result of this optimisation is given in Table 1. For the line-search the parameters for the maximal PSNR values are computed iteratively, by applying Alg. 1 for two different choices of (σ, ρ, η) at a time, and subsequently adapt this parameter-set for the next iteration towards the ones in the neighbourhood of the one that returns a larger PSNR. In this search we constrain $\sigma \leq \rho$ [27]. The line-search is stopped when, for the currently best parameters (σ, ρ, η) all the other neighbours in a certain radius of distance report an inferior PSNR value. In Fig. 2 we show the trajectory of the parameters during this line-search for the *Franke* video corrupted with Gaussian noise with $\varsigma = 10$. We observe that there exists a range of parameters in which the PSNR values are almost the same.

By looking at the estimated parameters from the line-search approach in Table 1, we suggest the following rule of thumb for their selection in Alg. 1:

$$\sigma = \rho = 3.2\eta^{-0.5} \quad \text{and} \quad \eta = 255\varsigma^{-1}. \quad (19)$$

4.2 Numerical results

For the so-found optimal parameters we compare in Table 1 the PSNR values achieved for our approach (TDV) with patch-based filters (V-BM3D v2.0 and V-BM4D v1.0, default parameters and *normal-complexity profile*).

In Figs. 3 and 4 the visual comparison is shown for selected frames of the *Franke* and *Water* videos (corrupted by noise with $\varsigma = 70$). The time-consistency achieved by our approach is apparent in the frame-by-frame PSNR comparison.

Video denoising results that use the quasi-optimal parameters computed with (19) are reported in Table 2: selected frames of videos corrupted with a high noise level of $\varsigma = 90$ are shown in Figs. 5 and 6, with frame-by-frame PSNR values.

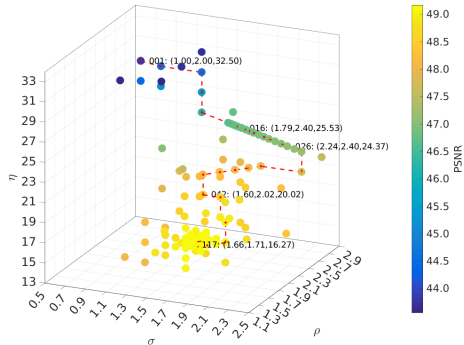


Fig. 2. Line-search (*Franke*, $\varsigma = 10$): optimal trajectory (dashed red line); PSNR values (coloured bullets). Optimal PSNR: 49.16, 117th iteration, $(\sigma, \rho, \eta) = (1.66, 1.71, 16.27)$.

4.3 Discussion of results

We compared our variational TDV denoising approach with patch-based (V-BM3D/V-BM4D) and variational (ROF 2D+t) methods. Patch-based methods are usually computationally faster than the variational approaches (including ours) but they tend to suffer from flickering and staircasing artefacts due to their patch-based nature. We experienced that our MATLAB code (not optimised for speed) is approximately $7\times$ slower than V-BM4D (C++ code with MEX interface) with *normal-complexity profile*. Both quantitative (via PSNR) and qualitative results (visual inspection) are relevant indicators for video denoising.

From the PSNR values in Tables 1 and 2 the TDV approach is comparable with the patch-based ones, with many single frames achieving higher PSNR value than the patch-based methods did. Also, by changing the noise level, the PSNR values are deteriorating less than with the patch-based methods, demonstrating the consistency of our approach.

Visual results confirm that the TDV approach improves upon patch-based methods producing less flickering and stair-casing artefacts, especially when the motion is smooth due to the coherence imposed also along the time dimension.

5 Conclusions

In this paper, we proposed a variational approach with the total directional variation (TDV) regulariser for video denoising. We extended the range of applications of TDV regularisation from image processing as demonstrated in [21] to videos. We compared TDV with some state of the art patch-based algorithms for video denoising and obtained comparable results especially for high level noises while reducing artefacts in regions with smooth large motion, where the patch-based approach shows some weakness. We expect to improve further the results by refining the estimation of the anisotropic fields [5] and by using higher-order derivatives in the TDV definition [21]. This is left for future research.

Table 1. PSNR comparison (best in bold), with TDV parameters from line-search.

Name (M, N, C, T)	ς	input	V-BM3D	V-BM4D	TDV (σ, ρ, η)	ROF 2D+t (η)
<i>Franke grey-scale</i> (120, 120, 1, 120)	10	28.13	45.99	46.90	49.16 (1.66, 1.71, 16.27)	42.56 (16.27)
	20	22.11	41.64	42.67	45.23 (2.00, 2.00, 08.10)	38.18 (08.10)
	35	17.25	38.63	39.34	41.89 (2.40, 2.40, 04.70)	34.59 (04.70)
	50	14.15	36.37	37.17	39.64 (2.70, 2.70, 03.30)	32.30 (03.30)
	70	11.23	30.60	35.03	37.44 (3.00, 3.00, 02.45)	30.45 (02.45)
<i>Franke coloured</i> (120, 120, 3, 120)	10	28.13	47.13	48.21	50.51 (1.89, 1.92, 16.59)	44.10 (16.59)
	20	22.11	42.96	43.97	46.46 (2.35, 2.35, 08.35)	39.93 (08.35)
	35	17.25	40.18	40.47	42.97 (2.79, 2.83, 04.74)	36.36 (04.74)
	50	14.15	38.11	38.15	40.74 (3.13, 3.17, 03.45)	34.41 (03.45)
	70	11.23	31.72	35.90	38.62 (3.50, 3.50, 02.45)	32.29 (02.45)
<i>Salesman</i> (288, 352, 1, 050)	10	28.13	37.30	37.12	35.24 (0.55, 0.68, 29.25)	31.48 (29.25)
	20	22.11	34.13	33.33	31.96 (0.70, 0.75, 13.93)	28.16 (13.93)
	35	17.25	30.79	30.20	29.36 (0.89, 0.89, 07.95)	26.01 (07.95)
	50	14.15	28.32	28.33	27.78 (1.05, 1.06, 05.45)	24.78 (05.45)
	70	11.23	24.55	26.68	26.34 (1.27, 1.32, 03.96)	23.87 (03.96)
<i>Water</i> (180, 320, 1, 120)	10	28.13	43.83	44.68	43.13 (0.93, 1.15, 25.75)	39.18 (25.75)
	20	22.11	40.59	41.02	39.84 (1.18, 1.35, 12.60)	35.94 (12.60)
	35	17.25	37.75	37.90	37.14 (1.40, 1.40, 06.95)	33.36 (06.95)
	50	14.15	35.58	35.85	35.41 (1.61, 1.65, 04.80)	31.83 (04.80)
	70	11.23	30.11	33.87	33.78 (1.80, 1.85, 03.45)	30.51 (03.45)

Table 2. PSNR comparison (best in bold), with quasi-optimal TDV parameters.

Name (M, N, C, T)	ς	input	V-BM3D	V-BM4D	TDV (σ, ρ, η)	ROF 2D+t (η)
<i>Miss America</i> (288, 360, 1, 150)	10	28.13	39.64	39.93	39.25 (0.63, 0.63, 25.50)	36.93 (25.50)
	20	22.11	37.95	37.78	37.28 (0.90, 0.90, 12.75)	34.60 (12.75)
	35	17.25	36.03	35.77	35.44 (1.19, 1.19, 07.29)	32.72 (07.29)
	50	14.15	34.19	34.26	34.14 (1.42, 1.42, 05.10)	31.47 (05.10)
	70	11.23	28.86	32.64	32.85 (1.68, 1.68, 03.64)	30.29 (03.64)
	90	09.05	27.42	31.27	31.87 (1.90, 1.90, 02.83)	29.42 (02.83)
<i>Xylophone coloured</i> (240, 320, 3, 141)	10	28.13	37.82	37.49	35.96 (0.63, 0.63, 25.50)	32.70 (25.50)
	20	22.11	34.70	34.13	33.06 (0.90, 0.90, 12.75)	29.57 (12.75)
	35	17.25	32.06	31.65	30.93 (1.19, 1.19, 07.29)	27.16 (07.29)
	50	14.15	29.98	30.07	29.58 (1.42, 1.42, 05.10)	25.72 (05.10)
	70	11.23	25.89	28.51	28.32 (1.68, 1.68, 03.64)	24.43 (03.64)
	90	09.05	24.50	27.32	27.37 (1.90, 1.90, 02.83)	23.57 (02.83)

References

1. Arias, P., Morel, J.M.: Video Denoising via Empirical Bayesian Estimation of Space-Time Patches. *J Math Imaging Vis* **60**(1), 70–93 (2018)
2. Brailean, J.C., et al.: Noise reduction filters for dynamic image sequences: a review. *Proceedings of the IEEE* **83**(9), 1272–1292 (1995)
3. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J Imaging Sci* **3**(3), 492–526 (2010)
4. Buades, A., Coll, B., Morel, J.: A non-local algorithm for image denoising. In: *IEEE CVPR 2005*. vol. 2, pp. 60–65 (2005)
5. Buades, A., Lisani, J., Miladinović, M.: Patch-based video denoising with optical flow estimation. *IEEE Trans Image Process* **25**(6), 2573–2586 (2016)
6. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. *SIAM J Multiscale Model Simul* **4**(2), 490–530 (2005)

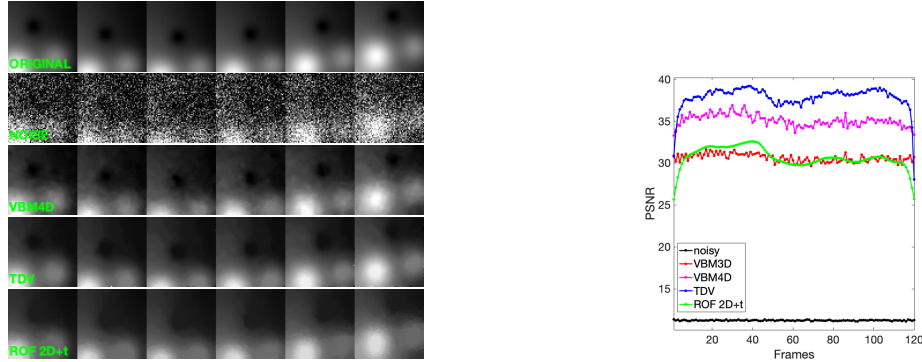


Fig. 3. *Frankie*: frames [10, 20, 30, 40, 50, 60] for $\zeta = 70$ and PSNR comparison.

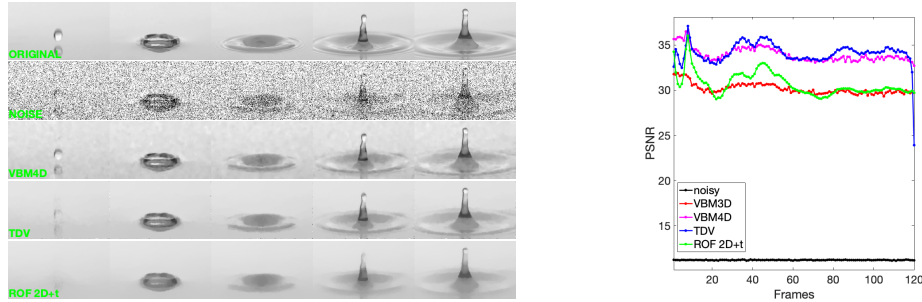


Fig. 4. *Water*: frames [5, 25, 45, 65, 85] for $\zeta = 70$ and PSNR comparison.

7. Burger, M., Dirks, H., Schönlieb, C.: A variational model for joint motion estimation and image reconstruction. *SIAM J Imaging Sci* **11**(1), 94–128 (2018)
8. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Imaging Vis* **40**(1), 120–145 (2011)
9. Chambolle, A., Pock, T.: An introduction to continuous optimization for imaging. *Acta Numerica* **25**, 161–319 (2016)
10. Dabov, K., Foi, A., Egiazarian, K.: Video denoising by sparse 3d transform-domain collaborative filtering. In: *15th Eur Signal Process.* pp. 145–149 (2007)
11. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: BM3D Image Denoising with Shape-Adaptive Principal Component Analysis. In: *SPARS’09* (2009)
12. Dalgas Kongskov, R., Dong, Y., Knudsen, K.: Directional Total Generalized Variation Regularization. *ArXiv e-prints* (2017)
13. Davy, A., Ehret, T., Facciolo, G., Morel, J., Arias, P.: Non-local video denoising by CNN. *ArXiv e-prints* (2018)
14. Horn, B.K., Schunck, B.: “determining optical flow”: a retrospective. *Artificial Intelligence* **59**(1), 81–87 (1993)
15. Lebrun, M., Buades, A., Morel, J.: A nonlocal bayesian image denoising algorithm. *SIAM J Imaging Sci* **6**(3), 1665–1688 (2013)
16. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *IJCAI’81.* pp. 674–679. Morgan Kaufmann (1981)
17. Maggioni, M., Boracchi, G., Foi, A., Egiazarian, K.: Video denoising using separable 4d nonlocal spatiotemporal transforms. *Proc.SPIE* **7870** (2011)



Fig. 5. *Miss America*: frames [5, 10, 15, 20, 25, 30] for $\zeta = 90$ and PSNR comparison.

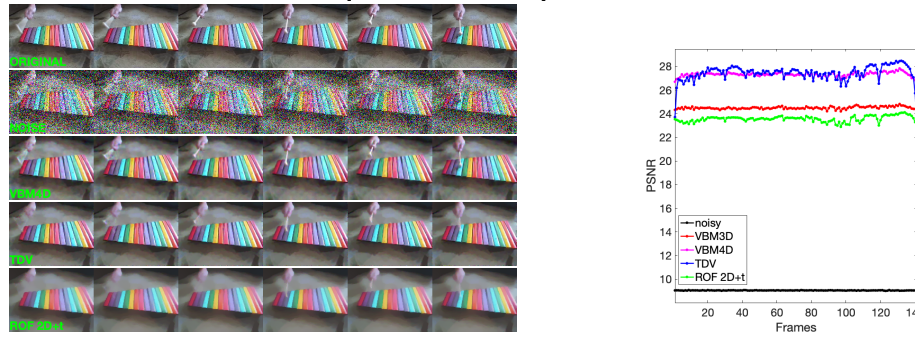


Fig. 6. *Xylophone*: frames [5, 10, 15, 20, 25, 30] for $\zeta = 90$ and PSNR comparison.

18. Ozkan, M.K., Sezan, M.I., Tekalp, A.M.: Adaptive motion-compensated filtering of noisy image sequences. *IEEE Trans Circuits Syst Video Technol* **3**(4) (1993)
19. Papafitsoros, K., Schönlieb, C.B.: A combined first and second order variational approach for image reconstruction. *J Math Imaging Vis* **48**(2), 308–338 (2014)
20. Parisotto, S., Masnou, S., Schönlieb, C.B.: Higher order total directional variation. Part II: Analysis. *ArXiv e-prints* (2018)
21. Parisotto, S., Masnou, S., Schönlieb, C.B., Lellmann, J.: Higher order total directional variation. Part I: Imaging applications. *ArXiv e-prints* (2018)
22. Parisotto, S.: Anisotropic variational models and PDEs for inverse imaging problems. Ph.D. thesis, University of Cambridge (2019)
23. Portilla, J., et al.: Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans Image Process* **12**(11), 1338–1351 (2003)
24. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* **60**(1), 259 – 268 (1992)
25. Sánchez Pérez, J., Meinhardt-Llopis, E., Facciolo, G.: TV-L1 Optical Flow Estimation. *IPOP* **3**, 137–150 (2013)
26. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: *Proc IEEE Int Conf Comput Vis*. pp. 839–846 (1998)
27. Weickert, J.: Anisotropic diffusion in image processing (1998)
28. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV-L1 optical flow. *Pattern Recognition* pp. 214–223 (2007)