Joint Removal of Random and Fixed-Pattern Noise through Spatiotemporal Video Filtering

Matteo Maggioni, Enrique Sánchez-Monge, Alessandro Foi

Abstract—We propose a framework for the denoising of videos jointly corrupted by spatially correlated (i.e. non-white) random noise and spatially correlated fixed-pattern noise. Our approach is based on motion-compensated 3-D spatiotemporal volumes, i.e. a sequence of 2-D square patches extracted along the motion trajectories of the noisy video. First, the spatial and temporal correlations within each volume are leveraged to sparsify the data in 3-D spatiotemporal transform domain, and then the coefficients of the 3-D volume spectrum are shrunk using an adaptive 3-D threshold array. Such array depends on the particular motion trajectory of the volume, the individual power spectral densities of the random and fixed-pattern noise, and also the noise variances which are adaptively estimated in transform domain. Experimental results on both synthetically corrupted data and real infrared videos demonstrate a superior suppression of the random and fixed-pattern noise from both an objective and a subjective point of view.

Index Terms—Video denoising, spatiotemporal filtering, fixed-pattern noise, power spectral density, adaptive transforms, thermal imaging.

I. INTRODUCTION

DIGITAL videos may be degraded by several spatial and temporal corrupting factors which include but are not limited to noise, blurring, ringing, blocking, flickering, and other acquisition, compression or transmission artifacts. In this work we focus on the joint presence of random and fixed-pattern noise (FPN). The FPN typically arises in raw images acquired by focal plane arrays (FPA), such as CMOS sensors or thermal microbolometers, where spatial and temporal nonuniformities in the response of each photodetector generate a pattern superimposed on the image approximately constant in time. The spatial correlation characterizing the noise corrupting the data acquired by such sensors [1], [2], [3] invalidates the classic AWGN assumptions of independent and identically distributed (i.i.d.) –and hence white– noise.

The FPN removal task is prominent in the context of long wave infrared (LWIR) thermography and hyperspectral imaging. Existing denoising methods can be classified into reference-based (also known as calibration-based) or scene-based approaches. Reference-based approaches first calibrate the FPA using (at least) two homogeneous infrared targets, having different and known temperatures, and then linearly estimate the nonuniformities of the data [4], [5]. However, since the FPN slowly drifts in time, the normal operations of the camera need to be periodically interrupted to update the estimate which has become obsolete. Differently, scene-based approaches are able compensate the noise directly from the acquired data, by modeling the statistical nature of the FPN; this is typically achieved by leveraging nonlocal self-similarity and/or the temporal redundancy present along the direction of motion [6], [7], [8], [9], [10], [11].

We propose a scene-based denoising framework for the joint removal of random and fixed-pattern noise based on a novel observation model featuring two spatially correlated (non-white) noise components. Our framework, which we denote as RF3D, is based on motion-compensated 3-D spatiotemporal volumes characterized by local spatial and temporal correlation, and on a filter designed to sparsify such volumes in 3-D spatiotemporal transform domain leveraging the redundancy of the data in a fashion similar to [12], [13], [14], [15]. Particularly, the 3-D spectrum of the volume is filtered through a shrinkage operator based on a threshold array calculated from the motion trajectory of the volume and both from the individual power spectral densities (PSD) and the noise variances of the two noise components. The PSDs are assumed to be known, whereas the noise standard deviations are adaptively estimated from the noisy data. We also propose an enhancement of RF3D, denoted E-RF3D, in which the realization of the FPN is first progressively estimated using the data already filtered, and then subtracted from the subsequent noisy frames.

To demonstrate the effectiveness of our approach, we evaluate the denoising performance of the proposed method and the current state of the art in video and volumetric data denoising [13], [15] using videos corrupted by synthetically generated noise and also real LWIR therm sequences acquired with a FLIR Tau 320 microbolometer camera. We implement RF3D (and E-RF3D) as a two-stage filter: in each stage use the same multi-scale motion estimator to build the 3-D volumes but a different shrinkage operator for the filtering. Specifically, we use a hard-thresholding operator in the first stage and an empirical Wiener filter in the second. Let us remark that the proposed framework can be also generalized to other filtering strategies based on a separable spatiotemporal patch-based model.

The remainder of the paper is organized as follows. In

Matteo Maggioni and Alessandro Foi are with the Department of Signal Processing, Tampere University of Technology, Finland. Enrique Sánchez-Monge is with Noiseless Imaging Ltd, Tampere, Finland.

Section II we formalize the observation model, and in Section III we analyze the class of spatiotemporal transform-domain filters. Section IV gives a description of the proposed denoising framework, whereas Section V discusses the modification required to implement the enhanced fixed-pattern suppression scheme. The experimental evaluation and the conclusions are eventually given in Section VI and Section VII, respectively.

II. OBSERVATION MODEL

We consider an observation model characterized by two spatially correlated noise components having distinctive PSDs defined with respect to the corresponding spatial frequencies. Formally, we denote a noisy video $z : X \times T \to \mathbb{R}$ as

$$z(\mathbf{x}, t) = y(\mathbf{x}, t) + \eta_{\mathrm{RND}}(\mathbf{x}, t) + \eta_{\mathrm{FPN}}(\mathbf{x}, t), \qquad (1)$$

where $(\mathbf{x}, t) \in X \times T$ is a voxel of spatial coordinate $\mathbf{x} \in X \subseteq \mathbb{Z}^2$ and temporal coordinate $t \in T \subseteq \mathbb{Z}$, $y : X \times T \to \mathbb{R}$ is the unknown noise-free video, and $\eta_{\mathrm{FPN}} : X \times T \to \mathbb{R}$ and $\eta_{\mathrm{RND}} : X \times T \to \mathbb{R}$ denote a realization of the FPN and zero-mean random noise, respectively.

In particular, we model η_{RND} and η_{FPN} as colored Gaussian noise whose variance can be defined as

$$\mathrm{var}\left\{ \mathcal{T}_{\mathrm{2D}}\Big(\eta_{\mathrm{RND}}(\cdot, t)\Big)(\boldsymbol{\xi}) \right\} = \sigma_{\mathrm{RND}}^2(\boldsymbol{\xi}, t)$$
$$= \varsigma_{\mathrm{RND}}^2(t)\, \Psi_{\mathrm{RND}}(\boldsymbol{\xi}), \qquad (2)$$

$$\mathrm{var}\left\{ \mathcal{T}_{\mathrm{2D}}\Big(\eta_{\mathrm{FPN}}(\cdot, t)\Big)(\boldsymbol{\xi}) \right\} = \sigma_{\mathrm{FPN}}^2(\boldsymbol{\xi}, t)$$
$$= \varsigma_{\mathrm{FPN}}^2(t)\, \Psi_{\mathrm{FPN}}(\boldsymbol{\xi}), \qquad (3)$$

where $\mathcal{T}_{\mathrm{2D}}$ is a 2-D transform, such as the DCT, operating on $N \times N$ blocks, $\boldsymbol{\xi}$ belongs to the $\mathcal{T}_{\mathrm{2D}}$ domain Ξ, σ_{RND}^2 and σ_{FPN}^2 are the time-variant PSDs of the random and fixed-pattern noise defined with respect to $\mathcal{T}_{\mathrm{2D}}$; the time-variant PSDs can be separated into their normalized time-invariant counterparts $\Psi_{\mathrm{RND}}, \Psi_{\mathrm{FPN}} : \Xi \to \mathbb{R}$ and the corresponding time-variant scaling factors $\varsigma_{\mathrm{RND}}^2, \varsigma_{\mathrm{FPN}}^2 : T \to \mathbb{R}$. We observe that the PSDs Ψ_{RND} and Ψ_{FPN} are known and fixed; moreover the random noise component η_{RND} is independent with respect to t, whereas the fixed-pattern noise component η_{FPN} is roughly constant in time, that is

$$\frac{\partial}{\partial t} \eta_{\mathrm{FPN}}(\mathbf{x}, t) \approx 0. \qquad (4)$$

The model (1) is much more flexible than the standard i.i.d. AWGN model commonly used in image and video denoising. In this paper we successfully use (1) to describe the raw output of a LWIR microbolometer array thermal camera; specifically, Fig. 1 and Fig. 2 show the PSDs of the random and fixed-pattern noise of video acquired by a FLIR Tau 320 camera. The power spectral densities in the figures are defined with respect to the global 2-D Fourier transform and the 8×8 2-D block DCT, respectively. As can be clearly seen from the figures, the two noise components are not white and instead are characterized by individual and nonuniform PSDs.
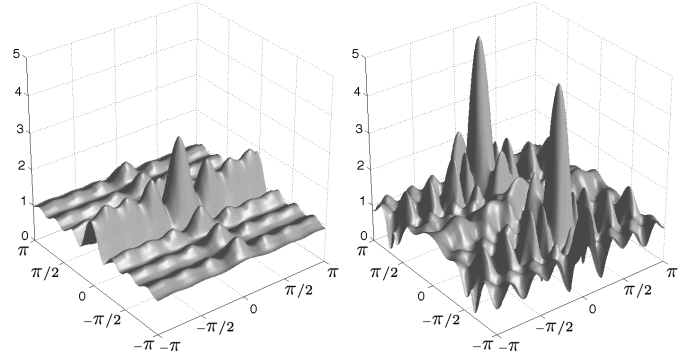

Fig. 1. Normalized root power spectral densities of the random (left) and fixed-pattern (right) noise components computed with respect to the global 2-D Fourier transform. The DC coefficient is located in the center (0,0) of the grid.
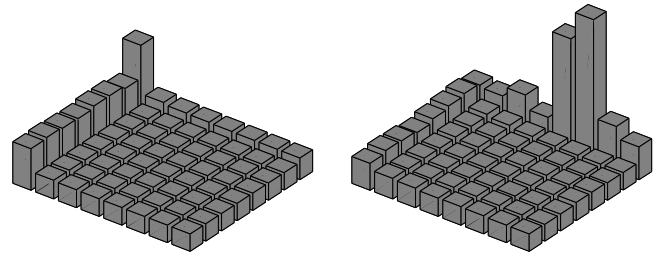

Fig. 2. Power spectral densities of the random (left) and fixed-pattern (right) noise components calculated with respect to the 2-D block DCT of size 8×8. The DC coefficient is located in the top corner.

III. SPATIOTEMPORAL FILTERING

In this section we generally analyze the class of spatiotemporal video filters, and, in particular, those characteristics of spatiotemporal filtering that are essential to the proposed noise removal framework.

A. Related Work

Natural signals tend to exhibit high auto correlation and repeated patterns at different location within the data [16], thus significant interest has been given to image denoising and compression methods which leverage redundancy and self-similarity [17], [18], [19], [20], [21]. For example, in [18] each pixel estimate is obtained by averaging all pixels in the image within an adaptive convex combination, whereas in [12] self-similar patches are first stacked together in a higher dimensional structure called "group", and then jointly filtered in transform domain. Highly correlated data can be sparsely represented with respect to a suitable basis in transform domain [22], [23], where the energy of the noise-free signal can be effectively separated from that of the noise through coefficient shrinkage. Thus, self-similarity and sparsity are the foundations of modern image [18], [12], video [13], [20], [14], and volumetric data [24], [15] denoising filters.

For the case of video processing, self-similarity can be naturally found along the temporal dimension. In [25], [26], [14] it has been shown that natural videos exhibit a strong temporal smoothness, whereas the nonlocal spatial redundancy only provides a marginal contribution to the filtering quality
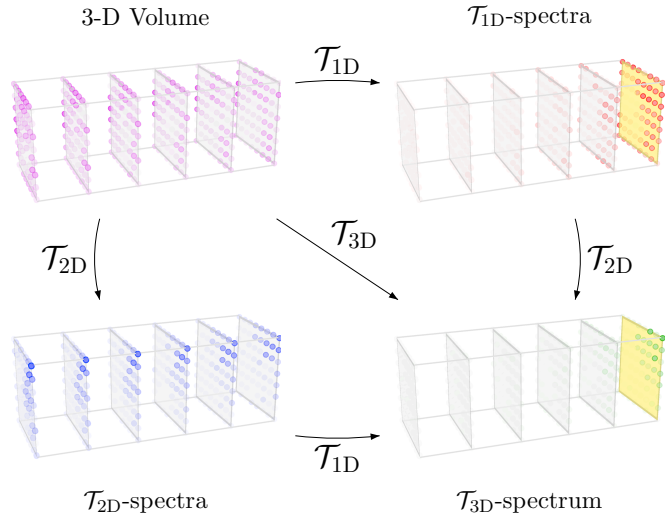
Fig. 3. Separable $\mathcal{T}_{3D}$ DCT transform applied to the 3-D volume illustrated in the top-left position. The magnitude of each 3-D element is proportional to its opacity. Whenever the temporal $\mathcal{T}_{1D}$ transform is applied, we highlight the 2-D temporal DC plane with a yellow background. Note that $\mathcal{T}_{1D}$-spectra is sparse outside the temporal DC plane, and $\mathcal{T}_{2D}$-spectra becomes sparser as we move away from the spatial DC coefficients (top-right corner of each block). Consequently, the energy of $\mathcal{T}_{3D}$-spectrum is concentrated around the spatial DC of the temporal DC plane.

[14]. Methods that do not explicitly account motion information have also been investigated [27], [28], [29], [30], but motion artifacts might occur around the moving features of the sequence if the temporal nonstationarities are not correctly compensated. Typical approaches employ a motion estimation technique to first compensate the data and then apply the filtering along the estimated motion direction [31], [32], [14]. A proper motion estimation technique is required to overcome the imperfections of the motion model, computational constraints, temporal discontinuities (e.g., occlusions in the scene), and the presence of the noise [33].

In this work, we focus on spatiotemporal video filters, so that the peculiar correlations present in the spatial and temporal dimension can be leveraged to minimize filtering artifacts in the estimate [34].

B. Filtering in Transform Domain

The spatiotemporal volume is a sequence of 2-D blocks following a motion trajectory of the video, and thus, in a fashion comparable to the "group" in [12], is characterized by local spatial correlation within each block and temporal correlation along its third dimension. As in [12], [13], [14], [15], the filtering is formalized as a coefficient shrinkage in spatiotemporal transform domain after a separable linear transform is applied on the data to separate the meaningful part of the signal from the noise. We use an orthonormal 3-D transform $\mathcal{T}_{3D}$ composed by a 2-D spatial transform $\mathcal{T}_{2D}$ applied to each patch in the volume followed by a 1-D temporal transform $\mathcal{T}_{1D}$ applied along the third dimension.

The $\mathcal{T}_{1D}$ transform should be comprised of a DC (direct current) coefficient representing the mean of the data, and a number of AC (alternating current) coefficients representing
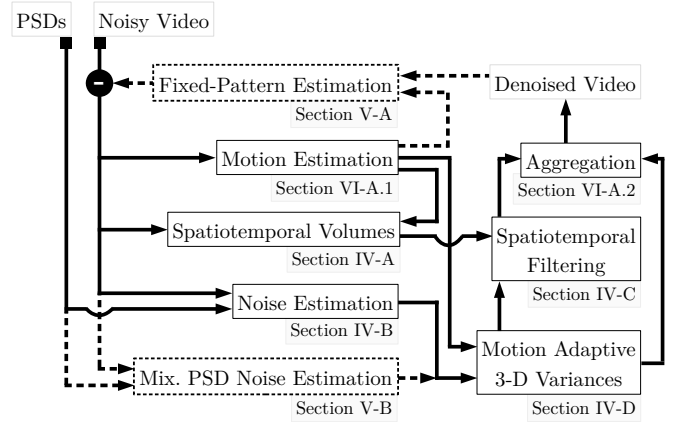

Fig. 4. Flowchart of random and fixed-pattern joint noise removal framework RF3D and the enhanced E-RF3D. A complete overview of RF3D is given in Section IV, while the modifications required to implement E-RF3D, illustrated as dashed lines, are described in Section V.

the local changes within the data. The 2-D temporal DC plane obtained after the application of the 1-D temporal transforms along the third dimension of the volume is of particular interest, as it encodes the features shared among the blocks, and thus can be used to capture the FPN present in the spatiotemporal volume. Fig. 3 provides a schematic representation of the 3-D spectrum obtained after applying a $\mathcal{T}_{1D}$, $\mathcal{T}_{2D}$, and $\mathcal{T}_{3D}$ DCT transforms on a typical spatiotemporal 3-D volume. The magnitude of each spectrum coefficient is directly proportional to its opacity, thus coefficients close to zero are almost transparent. The 2-D temporal DC plane in $\mathcal{T}_{1D}$-spectra and $\mathcal{T}_{3D}$-spectrum is highlighted with a yellow background, whereas the spatial DC coefficients in $\mathcal{T}_{2D}$-spectra are located at the top-right corner of each 2-D spectrum. Thus, the 3-D DC coefficient of the $\mathcal{T}_{3D}$-spectrum is located at the top-right corner of the temporal DC plane. Note how the data is differently sparsified in the different spectrum: in $\mathcal{T}_{1D}$-spectra the energy is concentrated in the temporal DC plane, in $\mathcal{T}_{2D}$-spectra the energy is concentrated around each spatial DC coefficients, and consequently in $\mathcal{T}_{3D}$-spectrum the energy is concentrated around the spatial DC of the temporal DC plane.

The PSDs of the noise in (1) are defined with respect to the 2-D spatial transform $\mathcal{T}_{2D}$. For example, in Fig. 2 we show the root PSDs of the random and fixed-pattern noise, obtained from a 2-D DCT of size 8×8. These PSDs provide the variances of the two noise components within each 2-D block coefficients before the application of the 1-D temporal transform to the spatiotemporal volume. The analogies with the corresponding PSDs defined with respect to the 2-D Fourier transform can be appreciated by referring to Fig. 1.

IV. JOINT NOISE REMOVAL FRAMEWORK

In this section, we describe the proposed RF3D framework for the joint removal of random and fixed-pattern noise. The RF3D works as follows: first a 3-D spatiotemporal volume is built for a specific position in the video (Section IV-A), and then the noise standard deviations are estimated from a set of frames (Section IV-B). Finally, the 3-D volume is filtered in

spatiotemporal transform domain (Section IV-C) using adaptive shrinkage coefficients (Section IV-D). A flowchart of the framework is illustrated in Fig. 4. This generic algorithm and its various applications are the object of a patent application [35].

The model (1) is simplified by (2) and (3), where we assume that the PSDs of η_{RND} and η_{FPN} are fixed modulo normalization with the corresponding scaling factors $\varsigma_{\mathrm{RND}}^2$ and $\varsigma_{\mathrm{FPN}}^2$. As a result, the PSDs do not need to be periodically estimated, but can be treated as known parameters. During the filtering, such parameters are scaled with the scaling factors to obtain the actual PSDs of the noise components corrupting the video. Further, we assume that the time-variant scaling factors of (2) and (3) vary slowly with time, so that they can be treated as constant within the local temporal extent of each spatiotemporal volume. Formally, we define the following conditions on the partial derivatives of ς_{RND} and ς_{FPN} with respect to time:

$$\frac{\partial}{\partial t}\varsigma_{\mathrm{RND}}(t) \approx 0, \qquad \frac{\partial}{\partial t}\varsigma_{\mathrm{FPN}}(t) \approx 0. \qquad (5)$$

A. Spatiotemporal Volumes

The proposed framework is based on motion-compensated 3-D spatiotemporal volumes composed by a sequence of 2-D blocks following a motion trajectory of the video [12], [13], [14]. Let $B(\mathbf{x}, t)$ be a 2-D $N \times N$ block extracted from the noisy video z, whose top-right corner is located at the 3-D coordinate $(\mathbf{x}, t)$. Formally, a motion trajectory corresponding to a (reference) block $B(\mathbf{x}_R, t_R)$ is a sequence of coordinates defined as

$$\Gamma(\mathbf{x}_R, t_R) = \left\{ (\mathbf{x}_i, t_i) \right\}_{i=h^-}^{h^+}, \qquad (6)$$

where $\mathbf{x}_i$ is the spatial location of the block within the frame at time t_i with $i = h^-, \dots, h^+$, and each voxel is consecutive in time with respect to the precedent, i.e. $t_{i+1} - t_i = 1 \; \forall i$. Note that in (6) we do not restrict the reference coordinate $(\mathbf{x}_R, t_R)$ to occupy a predefined position in the sequence, thus the trajectory can be grown backward and/or forward in time, i.e. $t_{h^-} \leq t_R \leq t_{h^+}$. Finally, we call $H = t_{h^+} - t_{h^-}$ the temporal extent of the volume.

Assuming that the trajectory for any given reference block $B(\mathbf{x}_R, t_R)$ is known, we can easily define the corresponding motion-compensated 3-D spatiotemporal volume as

$$\mathcal{V}(\mathbf{x}_R, t_R) = \left\{ B(\mathbf{x}_i, t_i) : (\mathbf{x}_i, t_i) \in \Gamma(\mathbf{x}_R, t_R) \right\}. \qquad (7)$$

The trajectories can be either known *a-priori*, or built in-loop, e.g., by concatenating motion vectors along time. However, let us stress that the motion estimation technique needs to be tolerant to noise [33], [29], [32], [14].

In Fig. 5, we show a schematic illustration of a spatiotemporal volume (7). In the figure, the reference block $B(\mathbf{x}_R, t_R)$ is shown in blue and occupies the middle position, the other blocks of the volume are shown in grey.
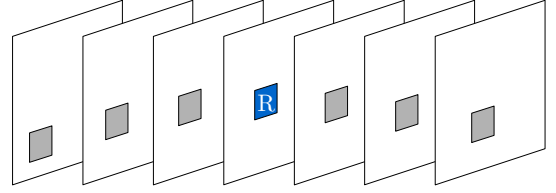

Fig. 5. Schematic illustration of a spatiotemporal volume. The blocks of the volume are grey with the exception of the reference block "R", which is blue.

B. Noise Estimation

The noise can be estimated leveraging the fact that the FPN is roughly constant in time (4): thus a spatial high-pass filtering of the video captures both random and fixed-pattern noise components, whereas a temporal high-pass filter captures only the random one.

The overall PSD σ^2 of the random and fixed-pattern noise is simply defined as the sum of (2) and (3)

$$\sigma^2(\boldsymbol{\xi}, t) = \varsigma_{\mathrm{RND}}^2(t)\,\Psi_{\mathrm{RND}}(\boldsymbol{\xi}) + \varsigma_{\mathrm{FPN}}^2(t)\,\Psi_{\mathrm{FPN}}(\boldsymbol{\xi}), \qquad (8)$$

being Ψ_{FPN} and Ψ_{RND} the only known terms of the equation.

Firstly we estimate σ as the median absolute deviation (MAD) [36], [37] of the $\mathcal{T}_{\mathrm{2D}}$-coefficients of all the blocks having temporal coordinates within $[t_{h^-}, t_{h^+}] \ni t$ as

$$\hat{\sigma}(\boldsymbol{\xi}, t) = \frac{1}{0.6745} \cdot \operatorname*{MAD}_{\substack{\mathbf{x} \in X \\ t_{h^-} \leq \tau \leq t_{h^+}}} \left(\mathcal{T}_{\mathrm{2D}}\Big(B(\mathbf{x}, \tau) \Big)(\boldsymbol{\xi}) \right), \qquad (9)$$

because $\mathcal{T}_{\mathrm{2D}}$ also embeds some high-pass filters and both ς_{RND} and ς_{FPN} are slowly varying in time (5). Then, we estimate σ_{RND} through a similar MAD on a temporal high-pass version of the video, obtained by differentiating consecutive blocks:

$$\hat{\sigma}_{\mathrm{RND}}(\boldsymbol{\xi}, t) = \frac{1}{0.6745} \cdot \operatorname*{MAD}_{\substack{\mathbf{x} \in X \\ t_{h^-} \leq \tau < t_{h^+}}} \left(\mathcal{T}_{\mathrm{2D}}\Big(B(\mathbf{x}, \tau + 1) \Big)(\boldsymbol{\xi}) \right.$$
$$\left. - \mathcal{T}_{\mathrm{2D}}\Big(B(\mathbf{x}, \tau) \Big)(\boldsymbol{\xi}) \right). \qquad (10)$$

We recognize that the MAD scaled by the usual factor 0.6745 (from the inverse cumulative Gaussian distribution at 3/4) is designed for Gaussian data. Even though in the practice the distribution of the noise in (1) may deviate from a Gaussian, the MAD/0.6745 is nevertheless a viable estimator for (9) and (10) because it is not applied directly on the observed data but on the $\mathcal{T}_{\mathrm{2D}}$ transform coefficients. Each transform coefficient is obtained as a linear combination involving many data samples (e.g., 64 samples when using a linear 8×8 $\mathcal{T}_{\mathrm{2D}}$), a "Gaussianization" kicks in, analogous to the central limit theorem. This makes the MAD/0.6745 an unbiased estimator of the standard deviation of each individual subband of transformed coefficients. In other words, we can safely use the MAD to estimate the root-PSD.

According to (2) and (3), σ_{RND}^2 and σ_{FPN}^2 must be respectively equal to Ψ_{RND} and Ψ_{FPN} modulo the non-negative scaling factors $\varsigma_{\mathrm{RND}}^2$ and $\varsigma_{\mathrm{FPN}}^2$, and as can be seen from (8) an analogous condition applies to σ^2. However, up to this point neither $\hat{\sigma}^2$ nor $\hat{\sigma}_{\mathrm{RND}}^2$ are guaranteed to satisfy such

scaling property. To find such scaling factors, we resort to the following non-negative least-squares optimization, whose solutions $\hat{\varsigma}^2_{\text{RND}}(t)$ and $\hat{\varsigma}^2_{\text{FPN}}(t)$ are defined as

$$
\begin{aligned}
\underset{\substack{\varsigma^2_{\text{RND}}(t)\geq 0 \\ \varsigma^2_{\text{FPN}}(t)\geq 0}}{\arg\min} \Bigg\{ &\sum_{\boldsymbol{\xi}\in\Xi} \Big(\Psi_{\text{RND}}(\boldsymbol{\xi})\varsigma^2_{\text{RND}}(t) \\
&+ \Psi_{\text{FPN}}(\boldsymbol{\xi})\varsigma^2_{\text{FPN}}(t) - \hat{\sigma}^2(\boldsymbol{\xi},t) \Big)^2 w_1^2(\boldsymbol{\xi}) \\
&+ \sum_{\boldsymbol{\xi}\in\Xi} \Big(\Psi_{\text{RND}}(\boldsymbol{\xi})\varsigma^2_{\text{RND}}(t) - \hat{\sigma}^2_{\text{RND}}(\boldsymbol{\xi},t) \Big)^2 w_2^2(\boldsymbol{\xi}) \Bigg\},
\end{aligned}
\tag{11}
$$

where $w_1, w_2 : \Xi \to \mathbb{R}$ give different weights to each coefficients fed to (9) and (10), and in practice can be used as logical operators to select linearly independent high-frequency coefficients in the $\mathcal{T}_{\text{2D}}$ domain.

C. Spatiotemporal Filtering

During the spatiotemporal filtering, the volume (7) is first transformed from its voxel representation to a new domain via a separable linear transform $\mathcal{T}_{\text{3D}}$, then a shrinkage operator Υ such as the hard thresholding modifies the magnitude of the spectrum coefficients to attenuate the noise. This strategy leverages the sparsification of the 3-D volume induced by $\mathcal{T}_{\text{3D}}$ as illustrated in Fig. 3. An estimate of the noise-free volume is eventually obtained after inverting the transform $\mathcal{T}_{\text{3D}}$ on the thresholded spectrum. The complete process can be formally defined as

$$
\hat{\mathcal{V}}(\mathbf{x}_R, t_R) = \mathcal{T}_{\text{3D}}^{-1}\Big(\Upsilon\Big(\mathcal{T}_{\text{3D}}\big(\mathcal{V}(\mathbf{x}_R, t_R) \big) \Big) \Big),
\tag{12}
$$

which in turn generates individual estimates of each noise-free patch in the volume. This strategy is referred to as collaborative filtering, and a deeper analysis of its rationale can be found in [12], [13], [19], [14].

D. Motion-Adaptive 3-D Spectrum Variances

The shrinkage operator Υ in (12) modulates the applied filtering strength relying on the variances $s^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, \vartheta)$ of the $\mathcal{T}_{\text{3D}}$-spectrum coefficients, where $\vartheta \in \{1, \ldots, H\} \subset \mathbb{N}$ indicates the coefficient position with respect to the $\mathcal{T}_{\text{1D}}$ spectrum, $\vartheta = 1$ corresponding to the temporal DC. Observe that $s^2_{\mathbf{x}_R, t_R}$ constitutes a 3-D array of variances. Each $s^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, \vartheta)$ depends on the $\mathcal{T}_{\text{2D}}$-PSDs (8) of each block $B(\mathbf{x}_i, t_i)$ in the volume (7) through the $\mathcal{T}_{\text{1D}}$ transform. Since both ς_{RND} and ς_{FPN} are slowly varying in time, we can use their respective estimates at the time t_R for the whole volume $\mathcal{V}(\mathbf{x}_R, t_R)$. However, due to the FPN, the relative spatial alignment of the blocks has an impact on the variance of the $\mathcal{T}_{\text{3D}}$ spectrum coefficients and thus needs to be taken into account for the design of the threshold coefficients.

To understand this phenomenon, let us consider the following two extreme cases. In one case all blocks are perfectly overlapping, i.e. they share the same spatial position $\mathbf{x}_i$ for all t_i in (7), such as when no motion is detected. Thus the FPN component, being the same across all blocks, accumulates through averaging in the 2-D temporal DC plane of the 3-D volume spectrum, shown in yellow in Fig. 3. For this reason the variances of temporal DC plane and AC coefficients are different:

$$
\begin{aligned}
s^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, 1) &= \varsigma^2_{\text{RND}}(t_R)\Psi_{\text{RND}}(\boldsymbol{\xi}) + H\varsigma^2_{\text{FPN}}(t_R)\Psi_{\text{FPN}}(\boldsymbol{\xi}), \\
s^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, \vartheta) &= \varsigma^2_{\text{RND}}(t_R)\Psi_{\text{RND}}(\boldsymbol{\xi}),
\end{aligned}
\tag{13}
$$

with $\vartheta \in \{2, \ldots, H\}$. In the other extreme case all blocks have different spatial positions and their relative displacement is such that the FPN exhibits uncorrelated patterns over different blocks. Thus, restricted to the volume, the FPN behaves just like another random component and the variances of the coefficients can be simply obtained as

$$
s^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, \vartheta) = \varsigma^2_{\text{RND}}(t_R)\Psi_{\text{RND}}(\boldsymbol{\xi}) + \varsigma^2_{\text{FPN}}\Psi_{\text{FPN}}(\boldsymbol{\xi}),
\tag{14}
$$

for all $\vartheta \in \{1, \ldots, H\}$.

We stress that the variances of the 3-D spectrum coefficients depend not only on the two PSDs and on the temporal extent H of the spatiotemporal volume, but also on the relative spatial alignment of the blocks within the volume, on the temporal position of the coefficients within the 3-D spectrum, and on the unknown covariance matrices of the overlapping blocks which however are impracticable to compute. Nevertheless we resort to a formulation that interpolates (13) and (14), approximating all the intermediate cases for which any number of blocks in the volume is aligned or partially aligned with any of the others.

For a spatiotemporal volume of temporal extent H, let $L_h \leq H$, with $1 \leq h \leq H$, be the number of blocks sharing the same spatial coordinates as the h-th block in the volume, and let $L = \max_{1 \leq h \leq H}\{L_h\}$, with $1 \leq L \leq H$, denote the maximum number of perfectly overlapping blocks. With this, we approximate the variances of the 3-D spatiotemporal coefficients by interpolating (13) and (14) with respect to L as

$$
\begin{aligned}
\hat{s}^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, 1) &= \hat{\varsigma}^2_{\text{RND}}(t_R)\Psi_{\text{RND}}(\boldsymbol{\xi}) \\
&\quad + \frac{L^2 + H - L}{H}\hat{\varsigma}^2_{\text{FPN}}(t_R)\Psi_{\text{FPN}}(\boldsymbol{\xi}),
\end{aligned}
\tag{15}
$$

$$
\begin{aligned}
\hat{s}^2_{\mathbf{x}_R, t_R}(\boldsymbol{\xi}, \vartheta) &= \hat{\varsigma}^2_{\text{RND}}(t_R)\Psi_{\text{RND}}(\boldsymbol{\xi}) \\
&\quad + \left[1 - \frac{L(L-1)}{H(H-1)} \right] \hat{\varsigma}^2_{\text{FPN}}(t_R)\Psi_{\text{FPN}}(\boldsymbol{\xi}),
\end{aligned}
\tag{16}
$$

with $\vartheta \in \{2, \ldots, H\}$. By construction, the variances (15) and (16) reduce to the exact formulae (13) for $L = H$ and to (14) for $L = 1$, but observe that (15) is also exact in the configuration where L blocks are perfectly overlapping and the other $H - L$ are completely displaced. In order to attain exact results in every configuration, (15) and (16) should have taken into account the basis coefficients of the $\mathcal{T}_{\text{1D}}$ temporal transform as well as the spatiotemporal position of the volume coefficients. The chosen formula (16) is such that the total $\mathcal{T}_{\text{2D}}$ noise spectrum, given by the sum of (15) with $H - 1$ times (16), is the same for all values of L and is equal to H times (14). Other approximate formulae are possible.

V. ENHANCED FIXED-PATTERN SUPPRESSION

In this section, we discuss the enhanced noise removal framework E-RF3D. Leveraging the fact that the fixed-pattern noise component varies slowly with time (4), it is possible to exploit its actual realization, i.e. the *fixed pattern* (FP), in a progressive fashion. In particular, the FP is first estimated from the noise that has been removed during previous filtering, and then subtracted from the following noisy frames to ease the denoising task (Section V-A). Consequently, the PSDs and the noise standard deviation of the data after the subtraction of the FP are updated (Section V-B). The modifications required to implement E-RF3D are illustrated as dashed lines in Fig. 4.

A. Fixed-Pattern Estimation

According to (1) and assuming that $\hat{y}$ is a good estimate of y, the noise realization at any position $(\mathbf{x}, t) \in X \times T$ can be estimated as

$$\hat{\eta}_{\text{FPN}}(\mathbf{x}, t) + \hat{\eta}_{\text{RND}}(\mathbf{x}, t) = z(\mathbf{x}, t) - \hat{y}(\mathbf{x}, t). \quad (17)$$

Since the FPN component η_{FPN} is assumed to be time-invariant within any short temporal extent (4), an estimate $\hat{\eta}_{\text{FPN}}(\mathbf{x}, t)$ of the FP can be simply obtained by averaging the noise residuals (17) of the previous $M(t) \in \mathbb{N}$ frames as

$$\hat{\eta}_{\text{FPN}}(\mathbf{x}, t) = \frac{1}{M(t)} \sum_{\tau = t - M(t) - 1}^{t-1} \Big(z(\mathbf{x}, \tau) - \hat{y}(\mathbf{x}, \tau) \Big), \quad (18)$$

for every position $\mathbf{x} \in X$ and time $t \in T$. Furthermore, if we assume that our estimate of the video is perfect, i.e. $\hat{y} = y$, then

$$\hat{\eta}_{\text{FPN}}(\mathbf{x}, t) = \eta_{\text{FPN}}(\mathbf{x}, t) + \bar{\eta}_{\text{RND}}(\mathbf{x}, t), \quad (19)$$

where $\bar{\eta}_{\text{RND}}$ is an average random component which has the same distribution and spatial correlation of $\eta_{\text{RND}}/\sqrt{M(t)}$. In this case, (18) is unbiased:

$$\mathbb{E}\Big\{ \hat{\eta}_{\text{FPN}}(\mathbf{x}, t) \Big\} = \eta_{\text{FPN}}(\mathbf{x}, t).$$

The number of frames $M(t)$ in (18) can be adjusted in different manners. In this work, we empirically set $M(t)$ to be approximately proportional to $\hat{\varsigma}_{\text{RND}}^2(t)/\hat{\varsigma}_{\text{FPN}}^2(t)$. Thus, $M(t)$ adapts conveniently to the current noise characteristics by balancing the accuracy of (18) with respect to its variance, which is proportional to $\varsigma_{\text{RND}}^2(t)/M(t)$. Note that the estimation of the FP is performed continuously during denoising in order to adapt to possible changes (drift) in the FP component.

Since $\hat{y}$ is never perfectly identical to y, (17) may contain structures belonging to the noise-free signal. This is particularly problematic whenever the video is stationary, because the static image content may accumulate into the FP (18). To counteract the consequent risks of fading and/or ghosting in the denoised signal, we select only those frames where motion is present. In particular, we use the displacement of the blocks between consecutive frames, since this information is readily available from (6): if the absolute mean displacement exceeds a certain threshold, we reckon that there is enough motion between the frames which can thus be used for the FP estimation.
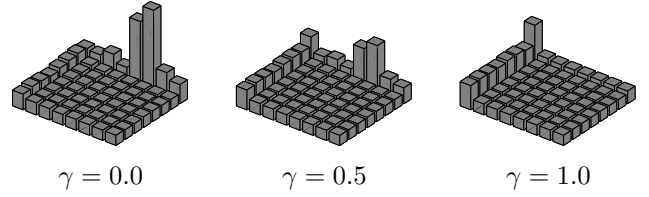

Fig. 6. Mixture of power spectral densities Ψ_{FPNnew} (20) describing the updated FPN component after the fixed-pattern subtraction. The power spectral densities are computed with respect to the 2-D DCT transform $\mathcal{T}_{\text{2D}}$ of size 8×8 and three different values for γ.

B. Noise Estimation with Mixed Power Spectral Density

We observe from (19) that the FP estimate (18) is still corrupted by an average random component distributed as $\eta_{\text{RND}}/\sqrt{M(t)}$. Thus, after subtraction of $\hat{\eta}_{\text{FPN}}(\cdot, t)$ from $z(\cdot, t)$, a new estimation of the standard deviation and the PSD of the updated FPN component becomes necessary.

Firstly, we model the PSD of the updated FPN Ψ_{FPNnew} as a convex combination of the original PSDs Ψ_{RND} and Ψ_{FPN}

$$\Psi_{\text{FPNnew}}(\boldsymbol{\xi}, t) = \gamma(t) \Psi_{\text{RND}}(\boldsymbol{\xi}) + \Big(1 - \gamma(t) \Big) \Psi_{\text{FPN}}(\boldsymbol{\xi}), \quad (20)$$

where the parameter $\gamma \in [0, 1]$ determines the contributions of the original PSDs. In Fig. 6 we present few PSDs combinations obtained with different values of γ: obviously, at the extreme values $\gamma = 1$ and $\gamma = 0$ (20) reduces to the original Ψ_{RND} and Ψ_{FPN}, respectively.

Secondly, similar to (11), we estimate the scaling factors of the mixed PSDs as the solutions $\hat{\varsigma}_{\text{RND}}^2(t)$, $\hat{\varsigma}_{\text{FPNmix}}^2(t)$, and $\hat{\varsigma}_{\text{RNDmix}}^2(t)$ of the non-negative least-squares problem

$$\underset{\substack{\varsigma_{\text{RND}}^2(t) \geq 0 \\ \varsigma_{\text{FPNmix}}^2(t) \geq 0 \\ \varsigma_{\text{RNDmix}}^2(t) \geq 0}}{\arg \min} \Bigg\{ \sum_{\boldsymbol{\xi} \in \Xi} \Big(\Psi_{\text{RND}}(\boldsymbol{\xi}) \varsigma_{\text{RND}}^2(t) + \Psi_{\text{FPN}}(\boldsymbol{\xi}) \varsigma_{\text{FPNmix}}^2(t) $$
$$+ \Psi_{\text{RND}}(\boldsymbol{\xi}) \varsigma_{\text{RNDmix}}^2(t) - \hat{\sigma}^2(\boldsymbol{\xi}, t) \Big)^2 w_1^2(\boldsymbol{\xi}) $$
$$+ \sum_{\boldsymbol{\xi} \in \Xi} \Big(\Psi_{\text{RND}}(\boldsymbol{\xi}) \varsigma_{\text{RND}}^2(t) - \hat{\sigma}_{\text{RND}}^2(\boldsymbol{\xi}, t) \Big)^2 w_2^2(\boldsymbol{\xi}) \Bigg\}, \quad (21)$$

where $\hat{\sigma}$ and $\hat{\sigma}_{\text{RND}}$ are obtained from the MAD of the high-frequency coefficients scaled by the weights $w_1, w_2 : \Xi \to \mathbb{R}$ as in (9) and (11). The optimization (21) aims to find the best non-negative solutions in the least-squares sense for the updated scaling factors using their definition (2) and (3). Note that the updated $\hat{\varsigma}_{\text{FPNnew}}(t)$ can be simply obtained from (21) as

$$\hat{\varsigma}_{\text{FPNnew}}^2(t) = \hat{\varsigma}_{\text{FPNmix}}^2(t) + \hat{\varsigma}_{\text{RNDmix}}^2(t).$$

Lastly, we compute the updated PSD (20) using a parameter γ defined as

$$\gamma(t) = \frac{\hat{\varsigma}_{\text{RNDmix}}^2(t)}{\hat{\varsigma}_{\text{FPNmix}}^2(t) + \hat{\varsigma}_{\text{RNDmix}}^2(t)}.$$

Note also that the updated Ψ_{FPNnew} and $\hat{\varsigma}_{\text{FPNnew}}$ are used for computing the adaptive threshold array (15)–(16) in place of Ψ_{FPN} and $\hat{\varsigma}_{\text{FPN}}$, respectively.

Fig. 7. Frames from the noise-free sequences *Foreman* (left) and *Miss America* (right).

VI. EXPERIMENTS

We compare the filtering results of RF3D and E-RF3D against those obtained using the same spatiotemporal filter but with different *a-priori* assumptions on the observation model:

- *WR*: data corrupted by one additive white random noise component and no FPN component. In this case, $\hat{\varsigma}_{\text{RND}}$ reduces to a weighted average of (9) and (10) over Ξ, because in the non-negative least-squares minimization (11) we assume $\varsigma_{\text{FPN}} = 0$ and $\Psi_{\text{RND}}(\boldsymbol{\xi}) = 1$ for all $\boldsymbol{\xi} \in \Xi$.
- *CR*: data corrupted by one additive colored random noise component and no FPN component. The PSD of such noise is assumed equal to

$$\frac{\varsigma_{\text{RND}}^2 \Psi_{\text{RND}} + \varsigma_{\text{FPN}}^2 \Psi_{\text{FPN}}}{\varsigma_{\text{RND}}^2 + \varsigma_{\text{FPN}}^2},$$

thus treating the FPN as another random component. Again, $\hat{\varsigma}_{\text{RND}}$ reduces to a weighted average of (9) and (10) over Ξ, because we assume $\varsigma_{\text{FPN}} = 0$.
- *WRWF*: data corrupted by two additive white noise components, namely random and fixed-pattern noise, with PSDs assumed as $\Psi_{\text{RND}}(\boldsymbol{\xi}) = \Psi_{\text{FPN}}(\boldsymbol{\xi}) = 1$ for all $\boldsymbol{\xi} \in \Xi$. Under this assumption, $\hat{\varsigma}_{\text{RND}}$ and $\hat{\varsigma}_{\text{FPN}}$ are given by (11).

Each of these simplified –and rough– assumptions reduce RF3D to an elementary algorithm that is unable to deal with the specific features of the actual noise model at hand. In particular, under the *WR* and *CR* assumptions, the FPN component is ignored and thus the filter is not able to account for the possible accumulation of FPN in the DC plane of the 3-D spectrum, which may hence remain unfiltered. Conversely, *WRWF* does model both the RND and FPN but ignores the spatial correlations that exist in the two noise components; thus filtering faces a particularly serious compromise between preservation of details and attenuation of noise. Additionally, we test the denoising performances of the state of the art in video and volumetric data denoising, namely V-BM3D [13] and BM4D [15], which are however designed for AWGN or, equivalently, for the WR assumption with i.i.d. Gaussian noise having standard deviation σ_{AWGN}.

In our experiments both synthetically corrupted sequences and real LWIR thermography data are considered. The objective denoising quality is measured by the peak signal-to-noise ratio (PSNR) of the estimate $\hat{y}$

$$10 \log_{10} \left(\frac{I_{\max}^2 |X||T|}{\sum_{\mathbf{x} \in X, t \in T} \left(\hat{y}(\mathbf{x}, t) - y(\mathbf{x}, t) \right)^2} \right),$$

where $I_{\max}$ is the maximum intensity value (peak) of the signal, y is the noise-free data, and $|X|$, $|T|$ are the cardinality of X and T, respectively. The data is hereafter considered to be in the range $[0, 255]$, i.e. $I_{\max} = 255$. We consider the standard sequences *Foreman*, *Coastguard*, *Miss America*, and *Flower Garden* corrupted as in (1) with different combinations of ς_{RND} and ς_{FPN}. In Fig. 7, we show two noise-free frames of *Foreman* and *Miss America*.

The remainder of this section is organized as follows. In Section VI-A we discuss the implementation details, parameter settings, and computational complexity of the proposed denoiser; in Section VI-B we present the denoising results for synthetic data; then, in Section VI-C, we show the denoising results of real thermography sequences to demonstrate that the proposed model (1) can appropriately describe the output of LWIR imagers.

A. Implementation Details

1) Motion Estimation: The proposed framework is relatively independent from the particular strategies used for the motion estimation. In our implementation, we use a coarse-to-fine two-scale motion estimator. First the sequence is downsampled by a factor of two; then the motion trajectories are computed using a fast diamond search [38] where the distance function is defined as the ℓ_2-norm difference of blocks of size $N \times N$, which thus cover an image area two times larger than that at the original resolution. Note that the downsampling increases the signal-to-noise ratio, and thus makes the motion estimation less impaired by noise. Finally, the found motion trajectories are refined on the full-resolution video using the same search process. For the refinement we employ a penalization term in the distance functional [14] to promote the matching of the blocks at the position predicted within the coarser scale.

2) Two-Stage Filtering: Similar to other algorithms [12], [13], [14], [15], we employ two cascading stages which differ for the particular shrinkage operator Υ (12): specifically we use a hard-thresholding operator in the first stage and an empirical Wiener filter in the second. The hard-thresholding stage is intended to provide a basic estimate which will serve as a pilot for the Wiener-filtering stage and uses an adaptive threshold array equal to the square root of the 3-D variances $s_{\mathbf{x}_R, t_R}^2$ scaled by a constant factor λ_{3D} [22], [12]. In both stages, the estimates of volumes are obtained after applying the inverse 3-D transform on their thresholded spectra, and then are returned in their original location. Overlapping estimates are finally aggregated through an adaptive convex combination using (15)–(16) as in [12]. This implementation can be interpreted either as the V-BM3D algorithm [13] with the block matching performed only along the temporal dimension, or as the V-BM4D algorithm [14] without the 4-D nonlocal grouping.

3) PSD Normalization: Without loss of generality, both PSDs Ψ_{RND} and Ψ_{FPN} are normalized with respect to their highest frequency coefficient. In Fig. 2, the highest frequency coefficients are located at the bottom corner, diametrically opposite to the DC coefficients. Observe in the figure that the magnitude of the highest-frequency coefficients is among the
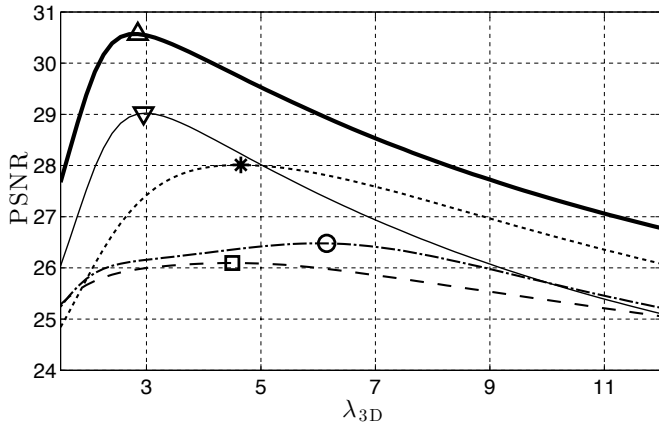
Fig. 8. Average PSNR (dB) obtained by WR (dashed line), CR (dot-dashed line), WRWF (dotted line), RF3D (thin solid line), and E-RF3D (thick solid line) as a function of the threshold factor λ_{3D}. The markers denote the global maxima.
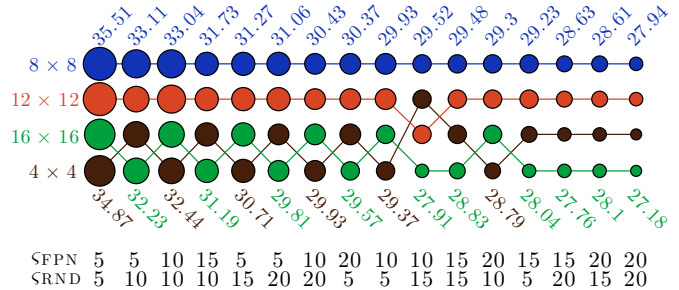

Fig. 9. Effect of different block sizes on the PSNR performance for the proposed method under different noise conditions. The area of the disks represents the average PSNR over different test videos for a specific noise level and block size, and each color represents a particular block size. The disks in each column are ordered in decreasing PSNR value from top to bottom; for each noise level we also report the best and worst PSNR value.

smallest of their respective PSDs Ψ_{RND} and Ψ_{FPN}; thus, the values of ς_{RND} and ς_{FPN} constitute only a rough quantitative indication of the actual strength of the two noise components, whose average standard-deviation can in fact be much larger than ς_{RND} and ς_{FPN}.

4) Parameter Settings: We set the maximum temporal extent of the spatiotemporal volumes to $H = t_{h^+} - t_{h^-} = 9$ with the reference block located in the middle, the size of the the 2-D blocks to $N \times N = 8 \times 8$, and the threshold factor to $\lambda_{3D} = 2.7$. As transform $\mathcal{T}_{3D}$ we utilize a separable 3-D DCT of size $N \times N \times H$.

The factor λ_{3D} is crucial: a too small or too large value may cause undersmoothing or oversmoothing of the data. In Fig. 8 we show the average PSNR obtained by the different methods for the denoising of the considered test videos and noise levels as λ_{3D} varies. We exclude *Miss America* from such average-value analyses because most of the sequence consists of a large smooth stationary background and thus its PSNR remains high even when a large λ_{3D} causes oversmoothing. The chosen $\lambda_{3D} = 2.7$ approximately yields the PSNR peak for both RF3D and E-RF3D; conversely, for WR, CR, and WRWF the best λ_{3D} needs to be larger (4.5, 6.15, and 4.65, respectively) to compensate the deficiencies of their assumed observation models. Note that $\lambda_{3D} = 2.7$ is equal to that used in [12] and is also not far from the universal threshold $\sqrt{2 \log(NNH)}$ [22].

Both BM4D [15] and V-BM3D [13] modulate their filtering strength with the standard deviation σ_{AWGN} of the i.i.d. Gaussian noise assumed to corrupt the data; however, because of the mismatch between the AWGN model and the actual observations (1), there is no ideal value of σ_{AWGN}. We aim to compare the proposed algorithm against the best possible BM4D and V-BM3D results; thus, we use "oracle" σ_{AWGN}^* values that maximize the output PSNR individually in each experiment. Details are given in the Appendix.

The block size 8×8 is widely used in many image-processing applications because it enables the use of fast transform implementation (e.g., DCT or FFT) and also allows

for a good data sparsification (e.g., BM3D denoising [12] or JPEG/MPEG compression). The denoising performances of the proposed E-RF3D using different block sizes are evaluated in Fig. 9; the performance is measured as average PSNR over the test sequences considered in our experiments, again minus Miss America, using $\lambda_{3D} = 2.7$. The area of each disk is proportional to the average PSNR (bigger disks indicate higher PSNR), and each color represents a particular block size. The disks in each column are ordered in descending PSNR value, and as one can clearly see, the best performance is always attained by 8×8 blocks (blue disks) with PSNR improvements ranging between 0.5dB and 1.5dB with respect to the worst case. Note that also V-BM3D as well as all others considered algorithms employ 8×8 blocks as basic data structures, whereas BM4D uses cubes of size $4 \times 4 \times 4$.

Our single-threaded MATLAB implementation[1] of the proposed algorithm used for the reported experiments processes a CIF-resolution sequence (i.e. 352×288) at approximately 1 frame per second on an Intel® i7-2640M CPU at 2.80-GHz.

B. Synthetic Data

The synthetic noisy sequences are generated according to the observation model (1) with the PSDs defined in (2) and (3) and shown in Fig. 2; ς_{RND} and ς_{FPN} are both simulated to remain constant in time. In order to present the best possible performances, every compared method use the optimized value of λ_{3D} discussed in Section VI-A4.

1) Joint Random and Fixed-Pattern Noise Removal: The PSNR denoising results under static and drifting FP are reported in Table I and Table II, respectively. Table II only includes E-RF3D because the other methods only exploit the PSD of the FPN, and not the actual realization FP, and thus are unaffected by the drift. In fact, the PSNR of such methods under static or drifting FP only differ by ± 0.1dB. Observe that a drift in the FP complicates the estimation (18), and thus the results of E-RF3D reported in Table II are not always as good as those obtained in case of static FP.

Referring to the PSNR results in Table I, RF3D and E-RF3D consistently outperform the results obtained under the less

---

[1]MATLAB code downloadable at http://www.cs.tut.fi/~foi/GCF-BM3D/.

TABLE I
PSNR (DB) DENOISING PERFORMANCE OF V-BM3D [13], BM4D [15], AND THE PROPOSED RF3D AND E-RF3D APPLIED TO DATA CORRUPTED BY SYNTHETIC NOISE AS IN (1) HAVING DIFFERENT COMBINATIONS OF ς_{FPN} AND ς_{RND}. THE SAME DATA IS ALSO FILTERED ASSUMING WHITE RANDOM NOISE (WR), COLORED RANDOM NOISE (CR), OR WHITE RANDOM AND WHITE FIXED-PATTERN NOISE (WRWF). THE FP IS STATIC IN TIME.

Video		Foreman				Coastguard				Miss America				Flower Garden			
Resolution		352 × 288				176 × 144				360 × 288				352 × 240			
Frames		300				300				150				150			
ς_{FPN}	Filter	ς_{RND}															
		5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
5	V-BM3D	33.89	33.20	32.11	30.88	32.11	31.47	30.59	29.58	37.22	36.99	36.58	35.75	32.25	30.09	28.25	26.73
	BM4D	33.18	32.72	31.84	30.83	32.27	31.66	30.77	29.86	35.64	36.11	36.10	35.45	31.37	29.18	27.29	25.75
	WR	34.41	33.26	31.94	30.80	32.27	31.26	30.18	29.16	35.85	37.30	37.04	36.20	27.02	25.83	24.65	23.58
	CR	34.42	32.73	31.22	30.03	32.03	30.90	29.77	28.75	37.91	37.65	36.92	36.08	26.55	25.27	24.09	23.04
	WRWF	35.32	33.71	32.32	31.15	33.45	31.96	30.79	29.76	37.68	37.32	36.98	36.10	31.36	29.19	27.35	25.87
	RF3D	36.14	34.52	33.16	32.00	34.02	32.75	31.65	30.68	38.14	37.39	37.10	36.48	32.23	30.04	28.26	26.87
	E-RF3D	**38.52**	**35.44**	**33.53**	**32.15**	**35.74**	**33.83**	**32.22**	**31.03**	**38.80**	**38.17**	**37.38**	**36.54**	**33.12**	**30.42**	**28.44**	**26.92**
10	V-BM3D	29.87	29.77	29.67	29.50	28.35	28.27	28.14	27.96	34.83	34.75	34.62	34.46	28.04	27.41	26.61	25.73
	BM4D	29.12	29.12	29.11	29.02	27.70	27.68	27.67	27.57	34.36	34.31	34.12	33.80	26.52	25.99	25.29	24.49
	WR	29.40	29.84	30.03	29.89	28.09	28.25	28.21	27.99	29.03	30.55	32.25	33.97	25.46	24.78	23.95	23.12
	CR	30.72	30.58	30.00	29.30	28.77	28.68	28.32	27.83	32.01	33.92	34.81	34.90	25.20	24.39	23.47	22.61
	WRWF	31.92	31.33	30.68	30.09	29.95	29.26	28.68	28.20	34.82	34.55	34.35	34.30	27.94	27.02	25.96	24.95
	RF3D	33.01	32.34	31.55	30.81	30.55	30.04	29.47	28.97	36.30	35.79	35.28	34.76	28.64	27.80	26.82	25.89
	E-RF3D	**37.10**	**34.78**	**33.12**	**31.82**	**33.01**	**32.17**	**31.22**	**30.34**	**36.74**	**36.45**	**35.87**	**35.61**	**30.76**	**29.27**	**27.81**	**26.43**
15	V-BM3D	27.83	27.81	27.77	27.72	26.23	26.20	26.16	26.10	32.94	32.91	32.84	32.75	25.01	24.79	24.51	24.13
	BM4D	26.54	26.55	26.59	26.67	25.18	25.19	25.25	25.26	32.52	32.47	32.37	32.20	23.19	23.05	22.81	22.51
	WR	25.62	26.18	26.79	27.31	24.62	25.02	25.43	25.73	25.07	25.97	27.18	28.51	23.75	23.41	22.92	22.38
	CR	27.46	27.96	28.15	28.02	25.80	26.20	26.41	26.40	27.77	29.39	31.08	32.15	23.75	23.28	22.66	22.00
	WRWF	29.68	29.34	28.98	28.64	27.77	27.36	26.97	26.64	32.44	32.28	32.16	32.06	25.40	24.95	24.36	23.73
	RF3D	31.06	30.64	30.13	29.59	28.58	28.27	27.89	27.51	34.32	34.02	33.71	33.35	26.05	25.69	25.20	24.65
	E-RF3D	**35.24**	**33.93**	**32.56**	**31.42**	**30.84**	**30.70**	**30.19**	**29.44**	**34.84**	**34.61**	**34.24**	**33.91**	**29.07**	**27.99**	**26.99**	**25.99**
20	V-BM3D	26.50	26.50	26.48	26.46	24.83	24.83	24.80	24.78	31.46	31.45	31.40	31.33	22.72	22.63	22.50	22.32
	BM4D	26.24	26.23	26.18	26.12	23.34	23.35	23.40	23.42	30.93	30.91	30.84	30.73	20.76	20.72	20.64	20.53
	WR	22.73	23.16	23.81	24.47	21.96	22.31	22.79	23.27	22.37	22.92	23.75	24.75	22.11	21.95	21.71	21.39
	CR	24.81	25.40	26.02	26.37	23.34	23.82	24.36	24.74	24.91	26.00	27.49	28.91	22.42	22.14	21.74	21.27
	WRWF	28.15	27.87	27.59	27.35	26.19	25.93	25.65	25.41	30.55	30.43	30.35	30.29	23.53	23.26	22.90	22.49
	RF3D	29.78	29.45	29.02	28.60	27.23	27.02	26.73	26.44	**32.74**	**32.52**	**32.31**	**32.07**	24.20	24.01	23.74	23.42
	E-RF3D	**33.77**	**32.98**	**31.93**	**30.93**	**29.98**	**30.03**	**29.40**	**28.86**	**32.74**	**32.52**	**32.31**	**32.07**	**27.60**	**26.94**	**26.17**	**25.43**

TABLE II
PSNR (DB) DENOISING PERFORMANCE OF E-RF3D APPLIED TO DATA CORRUPTED BY SYNTHETIC NOISE AS IN (1) HAVING DIFFERENT COMBINATIONS OF ς_{FPN} AND ς_{RND}. THE FP PRESENTS A DRIFT IN TIME. IN THIS CONDITION V-BM3D, BM4D WR, CR, WRWF, AND RF3D OBTAIN RESULTS COMPARABLE (± 0.1dB) TO THE ONES REPORTED IN TABLE I, AND THUS ARE NOT SHOWN.

Video		Foreman				Coastguard				Miss America				Flower Garden			
Resolution		352 × 288				176 × 144				360 × 288				352 × 240			
Frames		300				300				150				150			
ς_{FPN}	Filter	ς_{RND}															
		5	10	15	20	5	10	15	20	5	10	15	20	5	10	15	20
5	E-RF3D	37.87	35.10	33.32	32.00	35.30	33.43	31.95	30.80	38.02	37.89	37.28	36.52	32.88	30.30	28.35	26.90
10	E-RF3D	35.61	34.07	32.61	31.43	31.97	31.40	30.53	29.69	36.37	36.11	35.77	35.34	30.46	28.95	27.58	26.32
15	E-RF3D	33.28	32.76	31.75	30.76	29.70	29.67	29.18	28.59	34.36	34.29	34.20	33.79	28.57	27.58	26.64	25.71
20	E-RF3D	31.31	31.26	30.74	30.04	28.29	28.33	28.10	27.65	32.73	32.55	32.33	32.09	26.87	26.47	25.75	24.97

accurate WR, CR, and WRWF assumptions with a substantial PSNR improvement in almost every experiment. Similarly, the state-of-the-art V-BM3D and BM4D filters (which we remark are designed for AWGN) are outperformed by the RF3D and E-RF3D methods. This demonstrates the importance of correctly modeling and appropriately filtering the two different components of the noise. It is interesting to notice that whenever ς_{FPN} is large enough (≥ 10), the PSNR of WR and CR increase as ς_{RND} increases. This apparent counterintuitive be-

havior is explained by the fact that neither WR nor CR model the FPN component, which may accumulate in the temporal DC plane of the 3-D volume spectrum. Such accumulation is particularly significant when motion is absent, as shown by (13), and corresponds to DC-plane coefficients having much larger noise variance than the rest of the spectrum. WR and CR make no distinction between DC-plane coefficients and AC coefficients, thus an increase of the RND noise component results in a higher filtering strength, which partly compensates
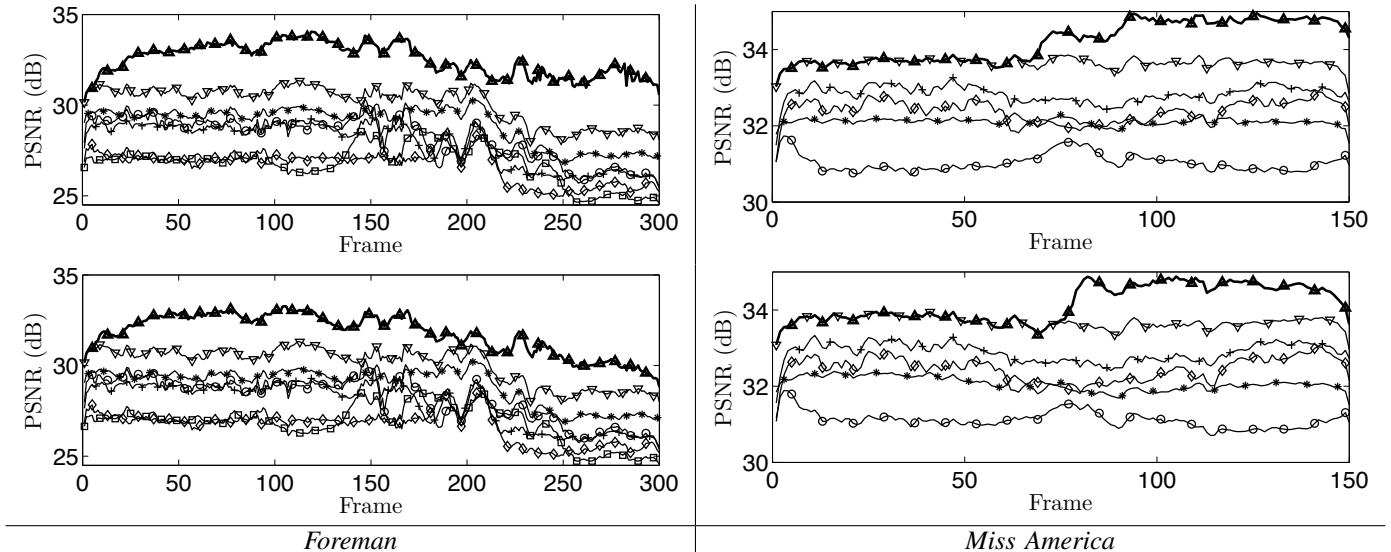
Fig. 10. Frame-by-Frame PSNR (dB) output of the videos *Foreman* and *Miss America* corrupted by synthetic noise having $\varsigma_{RND} = \varsigma_{FPN} = 15$ with either static FP (top row) or drifting FP (bottom row). We show the results of V-BM3D (+), BM4D ($\diamond$), WR ($\square$), CR ($\circ$), WRWF ($\ast$), RF3D ($\triangledown$), and E-RF3D ($\triangle$).

their model deficiency. The sequence *Flower Garden* is an exception: being a fast moving scene there is no accumulation of FPN and thus the PSNR naturally decreases with the increase of ς_{RND}. An additional remark about Table I regards the results of RF3D and E-RF3D for *Miss America* under high levels of ς_{FPN}: since the sequence presents little motion, E-RF3D is challenged to get a reliable estimate of the FP under strong FPN, and thus it is not able to provide the same performance gain as that of the other cases. As a matter of fact RF3D and E-RF3D provide the same PSNR results at $\varsigma_{FPN} = 20$.

Fig. 10 shows the frame-by-frame PSNR of *Foreman* and *Miss America* corrupted by random and fixed-pattern noise having $\varsigma_{RND} = \varsigma_{FPN} = 15$. *Miss America* and the first half of *Foreman* have low motion activity, whereas the second half of *Foreman* exhibits a high motion activity because of a fast transition in the scene. In good accord with the numerical results of Table I and Table II, E-RF3D ($\triangle$) always outperform the results obtained under WR ($\square$), CR ($\circ$), and WRWF ($\ast$) assumptions, as well as those of V-BM3D (+) and BM4D ($\diamond$). RF3D ($\triangledown$) is in few cases marginally inferior to V-BM3D (+). The advantage of the enhanced fixed-pattern suppression is clearly visible in all experiments, with the immediate and substantial PSNR improvement after the first estimate of the FP is subtracted (around the 10th frame in *Foreman* and between the 50th and the 75th frame in *Miss America*).

In Fig. 11, we show a denoised frame from *Foreman* and *Miss America* corrupted by synthetic noise having $\varsigma_{RND} = \varsigma_{FPN} = 15$, as well as the FP estimate obtained by E-RF3D. The noise-free data is shown in Fig. 7. Under the WR and CR assumptions the filter is unable to properly remove the FPN component, whose residual artifacts can be easily spotted within the denoised frames. In the WRWF results, we notice a good suppression of the random noise, but the structures of the FPN are still clearly visible. Conversely, RF3D and E-

RF3D generate more visually pleasant images, as the artifacts of the FPN are dramatically reduced and many high-frequency features, such as the hair and facial features of *Foreman* or the wrinkles in the clothes of *Miss America*, are nicely preserved. The results obtained by the V-BM3D and BM4D algorithms are separately presented in Fig. 12: as one can clearly see, the visual quality is significantly inferior those of RF3D and E-RF3D because of the remaining artifacts due to the FPN and the excessive loss of details.

2) Separate Random and Fixed-Pattern Noise Removal: The proposed filter is designed to jointly remove the random and fixed-pattern noise components, but for this set of experiments we modify it such that the two noise components are suppressed one at a time in two cascading passes. In other words the modified filter is applied twice on the observed data, first suppressing the random noise and then the FPN, or viceversa. From Fig. 13 it can be seen that whenever the FPN is suppressed before the random noise, the visual quality of the denoised videos is comparable or even slightly better to that obtained by the joint denoising strategy (at the obvious expense of a doubled computational load). The improvement is due to the assumption of zero random noise made in the first pass: if $\varsigma_{RND} = 0$ the number M of frames required for the FP estimation is small and thus the FP estimate can be obtained faster. Conversely, the reversed schema, implemented by suppressing the FPN after the random noise, is not as effective. In fact, as can be seen from the cheek of *Foreman* in Fig. 13, the corresponding denoising results exhibit significant FP artifacts.

3) Additive White Gaussian Noise Removal: In the final set of experiments using synthetic noise, we evaluate the proposed method against sequences corrupted solely by i.i.d. additive (white) Gaussian random noise with standard deviation σ_{AWGN}, which is assumed to be known. The proposed RF3D operates according to the WR assumption with $\hat{\varsigma}_{RND} = \sigma_{AWGN}$. The

Fig. 11. From top to bottom: denoising results of WR, CR, WRWF, RF3D, E-RF3D, and the FP estimate obtained from E-RF3D for *Foreman* and *Miss America* corrupted by synthetic noise having $\varsigma_{RND} = \varsigma_{FPN} = 15$.

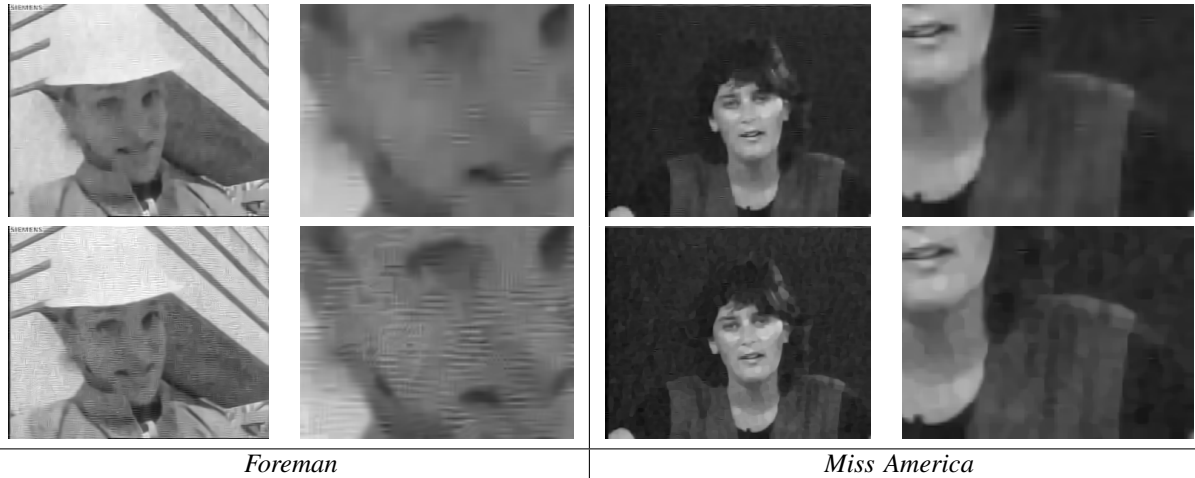*Foreman*   *Miss America*

Fig. 12. From top to bottom: denoising results of V-BM3D (*Foreman* 27.77 dB, *Miss America* 32.84 dB) and BM4D (*Foreman* 26.59 dB, *Miss America* 32.37 dB). The synthetic correlated noise is characterized by $\varsigma_{RND} = \varsigma_{FPN} = 15$.

Foreman   *Miss America*

Fig. 13. Denoising results for *Foreman* and *Miss America* corrupted by synthetic correlated noise having $\varsigma_{RND} = \varsigma_{FPN} = 15$ using E-RF3D to separately remove the two noise components. Top: first suppression of random noise and then FPN (*Foreman* 31.23dB, *Miss America* 33.37dB); bottom first suppression of the FPN and then random noise (*Foreman* 32.02dB, *Miss America* 34.29dB). For comparison, as can be seen in Table I, E-RF3D with joint-noise suppression provides 32.56dB for *Foreman* and 34.24dB for *Miss America*.

TABLE III
PSNR (dB) DENOISING PERFORMANCE OF V-BM3D, BM4D, AND RF3D
FOR DATA CORRUPTED BY I.I.D. GAUSSIAN NOISE WITH STANDARD
DEVIATION σ_{AWGN}.

σ_{AWGN}	Video Res. Frames	Foreman 352 × 288 300	Coastg. 176 × 144 300	Miss Am. 360 × 288 150	Fl. Gard. 352 × 240 150
5	V-BM3D	39.84	38.33	41.50	36.53
	BM4D	39.77	38.87	**42.02**	36.09
	RF3D	**40.27**	**39.43**	41.98	**36.58**
10	V-BM3D	36.55	34.82	39.64	**32.15**
	BM4D	36.38	35.31	**40.28**	31.39
	RF3D	**36.88**	**35.77**	40.19	32.06
20	V-BM3D	33.40	31.76	37.95	**28.30**
	BM4D	33.27	32.13	38.33	27.27
	RF3D	**33.72**	**32.36**	**38.40**	28.00
40	V-BM3D	29.99	28.28	35.46	**24.34**
	BM4D	30.39	29.08	36.03	23.40
	RF3D	**30.61**	**29.09**	**36.23**	24.21

rationale of these experiments is to compare RF3D against V-BM3D and BM4D on data where the latter two methods operate in ideal conditions; the results for different values of σ_{AWGN} are reported in Table III. From the table we can notice that the best-performing method is not the same for all experiments: while RF3D yields the best results in most of the cases, it also sometimes falls behind. The gap between the highest and lower PSNR values is at most 1.1dB, and typically much smaller; overall, these three methods perform comparably. Thus, the significant advantage (often several dB) of RF3D and especially E-RF3D in the case of correlated and fixed-pattern noise reported in Table I is a result of a correct modeling of the observed data, and not of an intrinsically more powerful algorithm.

C. LWIR Thermography Data

In this section we demonstrate the appropriateness of the proposed method through the denoising of two real LWIR thermography sequences acquired using a FLIR Tau 320 camera: the first sequence, *Matteo*, is characterized by high
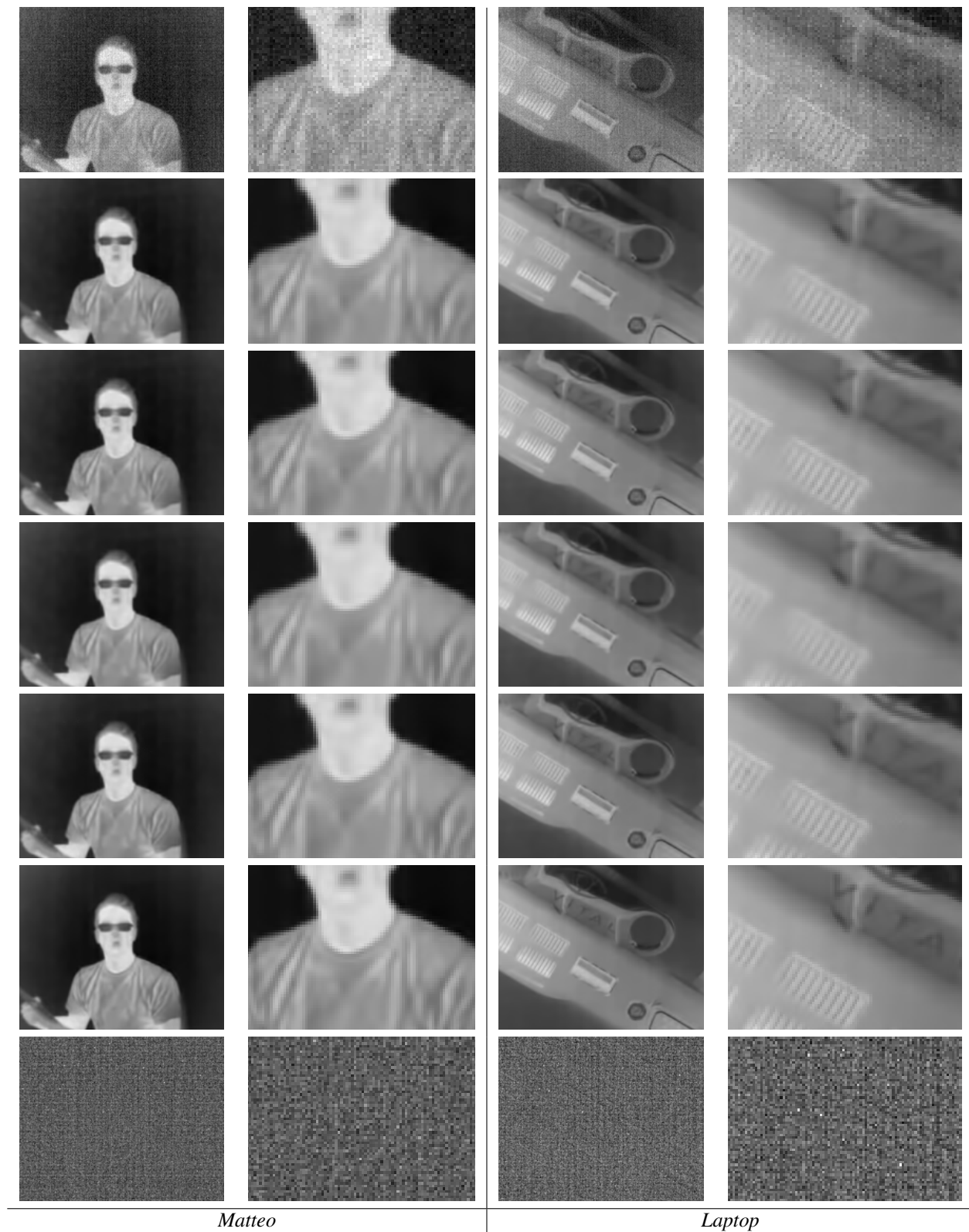
Fig. 14. From top to bottom: denoising results of WR, CR, WRWF, RF3D, E-RF3D, and the FP estimate obtained from E-RF3D for LWIR thermography sequences *Matteo* and *Laptop* acquired by a FLIR Tau 320 camera.

Matteo   *Laptop*
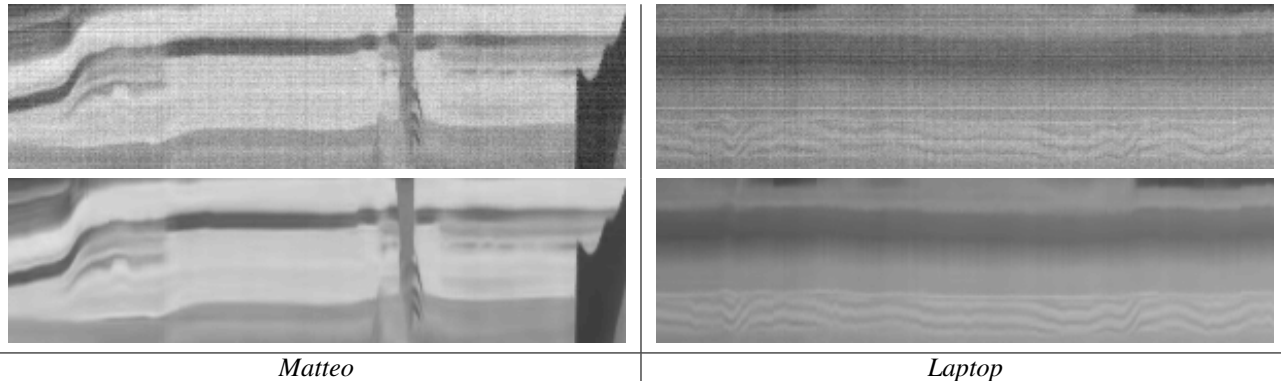
Fig. 15. Temporal cross-section of the noisy (top row) and E-RF3D denoised (bottom row) *Matteo* and *Laptop* sequences acquired by a FLIR Tau 320 camera. Both sequences consists of 300 frames. The artifacts of the FPN and the random noise are evident from the roughly constant streaks in time (horizontal direction) and space (vertical direction), respectively.

motion activity, whereas the second, *Laptop*, contains a more static scene[2]. The noise in the acquired data is characterized by $\varsigma_{RND} \approx 2.3$ and $\varsigma_{FPN} \approx 1.5$ over a $[6010, 6100]$ range, which corresponds to $\varsigma_{RND} \approx 6.5$ and $\varsigma_{FPN} \approx 4.3$ for a $[0, 255]$ range.

Objective assessments cannot be made because the ground-truth is not available, however, referring to Fig. 14, we can observe that under the WR, CR and WRWF assumptions the filter is not able to remove the noise, and that the best visual quality is obtained by the proposed RF3D and its enhancement E-RF3D. In particular, E-RF3D provides the best FPN suppression, which is evident from smooth areas such as the background of *Matteo*, and the best detail preservation, as can be seen from the folds in the tee-shirt of *Matteo* or the grid and letters in *Laptop*.

In the last row of Fig. 14 we show the FP estimate obtained from E-RF3D. As can be noticed, in the case of the static sequence *Laptop* part of the signal leaks into the residuals and is accumulated into the FP estimate. This is explained by the difficulty of unambiguously distinguishing the static information of the signal from the pattern of the FPN without the aid of motion (as described in Section V-A). In such cases the estimate of the FP (18) is likely to be less accurate, and thus isolating the noise component may be challenging. However, in spite of this mild leakage, the quality of the E-RF3D estimate is clearly superior to that of the compared methods (including RF3D), with better preservation of details and suppression of noise. In Fig. 15, we illustrate the effects of the random and fixed-pattern noise from the temporal cross-section of *Matteo* and *Laptop* (i.e. the horizontal dimension represents time, and the vertical dimension represents a particular cross-section of each frame). The effects of the noise structure of the FPN and RND can be respectively noticed from the horizontal and vertical streaks in the noisy data, whereas in the denoised counterparts these artifacts are effectively removed while preserving the fine (temporal) details, such as the three "claws" in the second half of *Matteo* and the "waves" in *Laptop*.

[2]This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. This includes the raw and filtered LWIR sequences of *Matteo* and *Laptop* as uncompressed AVI format movie clips. The material as GZIP Tar Archive file is 143 MB in size.

TABLE IV
MINIMUM (LEFT VALUE IN EACH CELL) AND MAXIMUM (RIGHT VALUE IN EACH CELL) VALUES OF THE ORACLE σ^*_{AWGN} PARAMETERS OF BM4D AND V-BM3D FOR EACH COMBINATION OF NOISE SCALING FACTORS ς_{FPN} AND ς_{RND}.

ς_{FPN}	Filter	ς_{RND}							
		5		10		15		20	
		min	max	min	max	min	max	min	max
5	V-BM3D	10	26	14	27	19	30	25	38
	BM4D	10	28	14	29	19	33	25	41
10	V-BM3D	16	57	19	60	23	60	27	59
	BM4D	17	189	19	160	23	152	27	150
15	V-BM3D	24	93	26	92	30	91	32	90
	BM4D	25	274	27	272	30	268	32	264
20	V-BM3D	34	124	35	123	39	120	40	120
	BM4D	35	385	37	380	39	371	40	352

VII. CONCLUSION

The contribution of this work is twofold. First, we developed an observation model for data corrupted by a combination of two spatially correlated components, i.e. random and fixed-pattern noise, each having its own non-flat PSD. This observation model can characterize several imaging sensors, and is particularly successful in describing the output of LWIR imagers. Second, we embed such observation model within a filtering framework based on 3-D spatiotemporal volumes built by stacking a sequence of blocks along the motion trajectories of the video. The volumes are then sparsified by a decorrelating 3-D transform, and then filtered in 3-D transform domain through a shrinkage operator based on both the PSDs of the noise components and on the relative spatial position of the blocks in the volume. Extensive experimental analysis demonstrates the subjective and objective (PSNR) effectiveness of the proposed framework for the denoising of synthetically corrupted videos, as well as the high visual quality achieved by the filtering of real LWIR thermography sequences. We further showed the capabilities of online FP estimation and subtraction to improve the denoising results.

APPENDIX

The denoising results of V-BM3D and BM4D in Table I are obtained with a default implementation of those algorithms

[13], [15] and an "oracle" value σ^*_{AWGN} of the assumed noise standard deviation. In particular, for each video and for each separate combination of ς_{RND} and ς_{FPN} under either static or drifting FPN, we have optimized σ^*_{AWGN} such that it yields the maximum PSNR value in each individual experiment. Due to length limitation and for the sake of illustration simplicity, in Table IV we report only the minimum and maximum of such optimum σ^*_{AWGN} values for all combination of noise scaling factors. As can be clearly seen, the difference between the maximum and minimum values notably increases with ς_{FPN}, thus indicating the impossibility of compensating the mismatch in the observation model by a simple tuning of the filter's parameters. Also, note how the maximum values tend to be very large in order to compensate the accumulated FPN in the volume spectra as quantified in (13).

REFERENCES

[1] A. F. Milton, F. R. Barone, and M. R. Kruer, "Influence of nonuniformity on infrared focal plane array performance," *Optical Engineering*, vol. 24, no. 5, pp. 245 855–245 855, Aug. 1985.

[2] M. T. Eismann and C. Schwartz, "Focal plane array nonlinearity and nonuniformity impacts to target detection with thermal infrared imaging spectrometers," in *Proceedings of the SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing*, vol. 3063, Jun. 1997, pp. 164–173.

[3] A. El Gamal, B. A. Fowler, H. Min, and X. Liu, "Modeling and estimation of FPN components in CMOS image sensors," in *Proceedings of the SPIE Solid State Sensor Arrays: Development and Applications*, vol. 3301, 1998, pp. 168–177.

[4] M. J. Schulz and L. V. Caldwell, "Nonuniformity correction and correctability of infrared focal plane arrays," in *Proceedings of the SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing*, vol. 2470, May 1995, pp. 200–211.

[5] A. Kumar, S. Sarkar, and R. P. Agarwal, "A novel algorithm and hardware implementation for correcting sensor non-uniformities in infrared focal plane array based staring system," *Infrared Physics & Technology*, vol. 50, no. 1, pp. 9–13, Mar. 2007.

[6] P. M. Narendra, "Reference-free nonuniformity compensation for IR imaging arrays," in *Proceedings of the SPIE Smart Sensors*, vol. 252, Jan. 1980, pp. 10–17.

[7] D. A. Scribner, K. A. Sarkady, J. T. Caulfield, M. R. Kruer, G. Katz, C. J. Gridley, and C. Herman, "Nonuniformity correction for staring IR focal plane arrays using scene-based techniques," in *Proceedings of the SPIE Infrared Detectors and Focal Plane Arrays*, vol. 1308, Apr. 1990, pp. 224–233.

[8] J. Harris and C. Yu-Ming, "Nonuniformity correction of infrared image sequences using the constant-statistics constraint," *IEEE Transactions on Image Processing*, vol. 8, no. 8, pp. 1148–1151, Aug 1999.

[9] S. N. Torres, J. E. Pezoa, and M. M. Hayat, "Scene-based nonuniformity correction for focal plane arrays by the method of the inverse covariance form," *Applied Optics*, vol. 42, no. 29, pp. 5872–5881, Oct. 2003.

[10] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[11] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[13] K. Dabov, A. Foi, and K. Egiazarian, "Video denoising by sparse 3D transform-domain collaborative filtering," in *Proceedings of the European Signal Processing Conference*, Sep. 2007. [Online]. Matlab code available: http://www.cs.tut.fi/~foi/GCF-BM3D/

[14] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 3952–3966, Sep. 2012.

[15] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 119–133, Jan. 2013. [Online]. Matlab code available: http://www.cs.tut.fi/~foi/GCF-BM3D/

[16] E. P. Simoncelli and B. Olshausen, "Natural image statistics and neural representation," *Annual Review of Neuroscience*, vol. 24, pp. 1193–1216, May 2001.

[17] J. S. De Bonet, "Noise reduction through detection of signal redundancy," Rethinking Artificial Intelligence, MIT AI Lab, Tech. Rep., 1997.

[18] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[19] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, "From local kernel to nonlocal multiple-model image denoising," *International Journal of Computer Vision*, vol. 86, pp. 1–32, Jan. 2010.

[20] H. Ji, S. Huang, Z. Shen, and Y. Xu, "Robust video restoration by joint sparse and low rank matrix approximation," *SIAM Journal on Imaging Sciences*, vol. 4, no. 4, pp. 1122–1142, Nov. 2011.

[21] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 106–128, Jan. 2013.

[22] D. Donoho, I. Johnstone, and I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1993.

[23] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[24] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for MRI denoising based on sparseness and self-similarity," *Medical Image Analysis*, vol. 16, no. 1, pp. 18–27, 2012.

[25] L. Jovanov, A. Pizurica, S. Schulte, P. Schelkens, A. Munteanu, E. Kerre, and W. Philips, "Combined wavelet-domain and motion-compensated video denoising based on video codec motion estimation methods," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 417–421, Mar. 2009.

[26] Z. Wang and Q. Li, "Statistics of natural image sequences: temporal motion smoothness by local phase correlations," in *Proceedings of the SPIE Human Vision and Electronic Imaging*, vol. 7240, Jan. 2009, pp. 1–12.

[27] R. Kleihorst, R. Lagendijk, and J. Biemond, "Noise reduction of image sequences using motion compensation and signal decomposition," *IEEE Transactions on Image Processing*, vol. 4, no. 3, pp. 274–284, 1995.

[28] A. Buades, B. Coll, and J. M. Morel, "Denoising image sequences does not require motion estimation," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, Sep. 2005, pp. 70–74.

[29] J. Boulanger, C. Kervrann, and P. Bouthemy, "Space-time adaptation for patch-based image sequence restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1096–1102, Jun. 2007.

[30] M. Protter and M. Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp. 27–35, Jan. 2009.

[31] E. Dubois and S. Sabri, "Noise reduction in image sequences using motion-compensated temporal filtering," *IEEE Transactions on Communications*, vol. 32, no. 7, pp. 826–831, Jul. 1984.

[32] C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Proceedings of the European conference on Computer vision*, 2010, pp. 706–719.

[33] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869–889, Aug. 1988.

[34] J. Brailean, R. Kleihorst, S. Efstratiadis, A. Katsaggelos, and R. Lagendijk, "Noise reduction filters for dynamic image sequences: A review," *Proceedings of the IEEE*, vol. 83, no. 9, pp. 1272–1292, Sep. 1995.

[35] A. Foi and M. Maggioni, "Methods and systems for suppressing noise in images," Patent Application US 13/943,035, Filed Jul. 16, 2013.

[36] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, Jun. 1974.

[37] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[38] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 287–290, Feb. 2000.

Matteo Maggioni received the B.Sc. and M.Sc. degree in Computer Engineering from Politecnico di Milano, Italy, in 2007 and 2010, respectively. Currently, he is pursuing a Ph.D. degree with the Department of Signal Processing, Tampere University of Technology, Finland. His research interests are focused on nonlocal adaptive transform-domain signal-restoration techniques.

Enrique Sánchez-Monge graduated in 2009 as a Telecommunications Engineer from Universidad Pública de Navarra, Spain, and received the M.Sc. in Signal Processing and Positioning and Navigation from Tampere University of Technology, Tampere, Finland in 2012. He is with Noiseless Imaging Ltd, Finland, as Senior Image Processing Engineer.
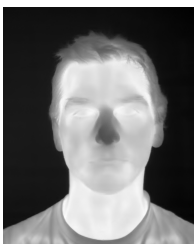
Alessandro Foi received the M.Sc. degree in Mathematics from the Università degli Studi di Milano, Italy, in 2001, the Ph.D. degree in Mathematics from the Politecnico di Milano in 2005, and the D.Sc.Tech. degree in Signal Processing from Tampere University of Technology, Finland, in 2007. He is currently an Academy Research Fellow with the Academy of Finland, at the Department of Signal Processing, Tampere University of Technology, where he is also Adjunct Professor.

His research interests include mathematical and statistical methods for signal processing, functional and harmonic analysis, and computational modeling of the human visual system. His recent work focuses on spatially adaptive (anisotropic, nonlocal) algorithms for the restoration and enhancement of digital images, on noise modeling for imaging devices, and on the optimal design of statistical transformations for the stabilization, normalization, and analysis of random data.

He is a Senior Member of the IEEE and an Associate Editor for the IEEE Transactions on Image Processing.