

Anisotropic Spatiotemporal Regularization in Compressive Video Recovery by Adaptively Modeling the Residual Errors as Correlated Noise

Nasser Eslahi and Alessandro Foi
Laboratory of Signal Processing
Tampere University of Technology, Finland

Abstract—Many approaches to compressive video recovery proceed iteratively, treating the difference between the previous estimate and the ideal video as residual noise to be filtered. We go beyond the common white-noise modeling by adaptively modeling the residual as stationary spatiotemporally correlated noise. This adaptive noise model is updated at each iteration and is highly anisotropic in space and time; we leverage it with respect to the transform spectra of a motion-compensated video denoiser. Experimental results demonstrate that our proposed adaptive correlated noise model outperforms state-of-the-art methods both quantitatively and qualitatively.

I. INTRODUCTION

High-speed motion capture has many important applications. However, direct high frame-rate capture is severely restricted by hardware constraints, often inherent to the read-out electronics. This notwithstanding, the spatial and temporal regularity of natural scenes makes it possible to recover high frame-rate video indirectly from temporally multiplexed measurements at a low frame-rate [1]–[9]. The forward model of a temporal multiplexing camera maps a high frame-rate video sequence composed of pq frames into q frames, where $p > 1$ is the temporal compression factor. This can be written in matrix notation as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^{n_1 n_2 p \times q}$ is a matrix whose columns are the vectorization of p consecutive $n_1 \times n_2$ spatial frames of the high frame-rate video, $\mathbf{A} \in \mathbb{R}^{n_1 n_2 \times n_1 n_2 p}$ is a coding matrix, $\mathbf{e} \in \mathbb{R}^{n_1 n_2 \times q}$ is a random measurement error, and $\mathbf{y} \in \mathbb{R}^{n_1 n_2 \times q}$ is a matrix whose columns are the vectorized temporally compressed measurements of the columns of \mathbf{x} . Typical temporal multiplexing cameras such as streak cameras [7] and the coded aperture compressive temporal imager (CACTI) [3] leverage translating masks during exposure. Assuming the vectorization of each video chunk is done first in time and then in the direction of mask translation, then the coding matrix \mathbf{A} has a $n_1 n_2 \times n_1 n_2$ block diagonal structure, where the $n_1 n_2$ blocks along the diagonal are $1 \times p$ row vectors, whose vertical concatenation yields a stack of n_2 Toeplitz matrices, each of size $n_1 \times p$.

After capture, the high-speed video can be recovered using a nonlinear sparsity-promoting algorithm, under the assumption that \mathbf{x} is sparse or compressible with respect to a given basis or a redundant dictionary mutually incoherent with \mathbf{A} [10], [11]. This process is termed compressive video recovery (CVR).

CVR is typically iterative, where each iteration includes a filtering step in which the current/previous estimate is treated as a degraded observation of the video to be recovered [1]–[6]. A common feature

of these approaches is that the degradation to be filtered at each iteration is modeled as additive white Gaussian noise (AWGN) which is alleviated by shrinkage or denoising. However, the common assumption of uncorrelated white noise holds only under special conditions that are hardly met in practice, e.g., the coding matrix \mathbf{A} being itself random i.i.d. Gaussian [12].

A more general model, which we advocate in this paper, allows spatial and temporal correlation within such degradations. This correlation can be result of multiple contributors: the structure of \mathbf{A} , the statistics of \mathbf{e} , as well as their interaction with the structure of \mathbf{x} and the effect of denoisers during the previous iterations. Since denoising filters take advantage of spatial and temporal redundancy within the video, the filtering can introduce various forms of correlation between adjacent samples in space and time.

In contrast to AWGN noise, correlated noise can lead to errors that are disproportionate across the data spectrum, to an extent that AWGN denoisers may not effectively discern between the true signal and noise in regularization via shrinkage. Hence, ignoring such correlation in the denoising step can lead to ineffective filtering and also distortion to the underlying video, thus impairing the accurate high-quality recovery of \mathbf{x} .

In this work we leverage the RF3D filter [13], which natively supports different forms of spatiotemporal noise correlation.

Our contributions are summarized as follows:

- We model the noise at each iteration of CVR as *stationary spatiotemporally correlated noise*, and adopt this model within the iterative/shrinkage thresholding (IST) [14] approach (Section II-A).
- We describe the noise correlation through the noise *power spectral density* (PSD), which we estimate at every iteration from the *median absolute deviation* (MAD) [15] of the transform spectra of the *residual error signal* (Section II-C), with respect to the sparsifying transform used internally by the RF3D filter (Section II-B): these PSDs modulate the shrinkage thresholds, i.e. allow to compare the magnitude of each transform coefficient against that of the corrupting noise.
- We develop a method for CVR based on the IST framework via the RF3D filter as the denoiser/regularizer under the modeling assumption of either i.i.d. (referred to as IST-wRF3D) or spatiotemporally correlated noise model (referred to as IST-cRF3D). We compare the state-of-the-art CVR techniques with our developed method, showing a significant advantage of our adaptive correlated noise model in terms of both objective and subjective visual quality (Section III).

The research leading to these results has received funding from the European Union’s H2020 Framework Programme (H2020-MSCA-ITN-2014) under grant agreement no. 642685 MacSeNet, and by the Academy of Finland (project no. 310779).

II. ANISOTROPIC SPATIOTEMPORAL REGULARIZATION IN COMPRESSIVE VIDEO RECOVERY

A. Denoising-based Iterative Shrinkage Thresholding

The IST framework [14] follows the two-step iterative procedure

$$\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k, \quad (2a)$$

$$\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k + \rho_k \mathbf{A}^\dagger \mathbf{r}_k), \quad (2b)$$

where \mathbf{x}_k and \mathbf{r}_k are, respectively, the estimate of the underlying video \mathbf{x} and the residual measurement at iteration $k \geq 0$, $\mathbf{x}_0 = 0$, Φ is a sparsity-promoting filter, \mathbf{A}^\dagger is the pseudo-inverse of \mathbf{A} , and $\rho_k > 0$ is the step size.

The action of Φ on its input can be regarded as a denoiser seeking to recover \mathbf{x} from the noisy observation [12]

$$\mathbf{z}_k = \mathbf{x} + \boldsymbol{\eta}_k = \mathbf{x}_k + \rho_k \mathbf{A}^\dagger \mathbf{r}_k, \quad (3)$$

where \mathbf{z}_k and $\boldsymbol{\eta}_k$ respectively represent the degraded video and the effective noise at each iteration of (2a)-(2b). We dub $\rho_k \mathbf{A}^\dagger \mathbf{r}_k$ the *residual signal*.

B. Stationary Colored Gaussian Noise Denoising

We use the RF3D filter Φ in the denoising step (2b), tacitly assuming that its input \mathbf{z}_k is reshaped so to reconstitute the three dimensions of the video and that its output is vectorized back. RF3D models the noise $\boldsymbol{\eta}_k$ as a combination of two frame-wise noise components: a *random noise* $\boldsymbol{\eta}_{\text{RND}_k}$ that is independently realized at every frame, and a *fixed-pattern noise* (FPN) $\boldsymbol{\eta}_{\text{FPN}_k}$ that is constant in time. Thus, in a set of pq frames, there are pq independent realizations of $\boldsymbol{\eta}_{\text{RND}_k}$ and pq copies of a unique realization of $\boldsymbol{\eta}_{\text{FPN}_k}$. Both $\boldsymbol{\eta}_{\text{RND}_k}$ and $\boldsymbol{\eta}_{\text{FPN}_k}$ are zero-mean processes and each features its own spatial correlation. Overall this corresponds to a spatiotemporally correlated noise whose 3D-FFT PSD is composed of

- $pq - 1$ temporal-AC planes that being characterized exclusively by $\boldsymbol{\eta}_{\text{RND}_k}$ are all equal to each other, and
- a temporal-DC plane that encompasses both $\boldsymbol{\eta}_{\text{RND}_k}$ and pq copies of $\boldsymbol{\eta}_{\text{FPN}_k}$.

Leveraging this noise model into the filter Φ (2b) is equivalent to an anisotropic spatiotemporal regularization in CVR.

RF3D aggregates a multitude of motion-compensated spatiotemporal volumes which are filtered in a transform domain. Specifically, a generic spatiotemporal volume \mathbf{v} extracted from \mathbf{z}_k is formed by concatenating $b \times b$ blocks extracted from $h = h^+ - h^- + 1$ consecutive frames into a $b \times b \times h$ array, i.e. $\mathbf{v} = [\mathbf{z}_k[\mathbf{d}_i, i]]_{i=h^-}^{h^+}$, where \mathbf{d}_i are the spatial coordinates of a block within i -th frame, and h^- and h^+ are the first and last frame indices of the spatiotemporal volume. The spatial coordinates are adaptively chosen so that concatenated blocks follow a motion trajectory in the scene. The transform-domain filtering of \mathbf{v} can be defined as

$$\tilde{\mathbf{v}} = \mathcal{T}_{3\text{D}}^{-1} \left(\Upsilon \left(\mathcal{T}_{3\text{D}}(\mathbf{v}); \tau_k \right) \right), \quad (4)$$

where Υ is a shrinkage operator, e.g., hard thresholding, τ_k is a threshold value depending on the statistics of $\boldsymbol{\eta}_k$ and \mathbf{x} , $\mathcal{T}_{3\text{D}}$ is the separable 3D discrete cosine transform (DCT) operating on 3D spatiotemporal volumes and $\mathcal{T}_{3\text{D}}^{-1}$ is its inverse. The blocks forming $\tilde{\mathbf{v}}$ are estimates of the noise-free blocks $\{\mathbf{x}[\mathbf{d}_i, i]\}_{i=h^-}^{h^+}$ and they are therefore returned and adaptively aggregated to their original positions $\{\mathbf{d}_i, i\}_{i=h^-}^{h^+}$ in the video.

In the simplest case when $\boldsymbol{\eta}_k$ is a zero-mean AWGN with variance $\sigma_{\boldsymbol{\eta}_k}^2$, τ_k could have been fixed as a multiple of standard deviation of the noise, i.e. $\tau_k = \lambda \sigma_{\boldsymbol{\eta}_k}$, $\lambda > 0$.

In the spatially-correlated random+fixed noise model of RF3D, the coefficient-invariant threshold has to be changed to an adaptively varying one:

$$\tau_k[\xi_1, \xi_2, \xi_3] = \lambda \sigma_{\mathcal{T}_{3\text{D}}(\mathbf{v})}[\xi_1, \xi_2, \xi_3], \quad (5)$$

where $\sigma_{\mathcal{T}_{3\text{D}}(\mathbf{v})}^2$ is the $\mathcal{T}_{3\text{D}}$ noise PSD of \mathbf{v} , and ξ_1 , ξ_2 , and ξ_3 are, respectively, the two spatial frequencies and the temporal frequency in the separable $\mathcal{T}_{3\text{D}}$ domain.

The $\mathcal{T}_{3\text{D}}$ noise PSD of \mathbf{v} can vary not only with respect to ξ_1 , ξ_2 , and ξ_3 , but also depending on the spatial alignment of the blocks, i.e. $\{\mathbf{d}_i\}_i$, due to the presence of FPN (see [13, Section IV-D]). In particular, there are two extreme cases: 1) all blocks in \mathbf{v} share the same spatial position, hence the FPN accumulates in the temporal-DC component; 2) all blocks have sufficiently different positions so that the FPN at \mathbf{d}_i is not correlated with that at \mathbf{d}_j if $i \neq j$, hence the FPN behaves like another independent noise at every frame. The mathematical expressions of the $\mathcal{T}_{3\text{D}}$ noise PSD of \mathbf{v} for these two cases are

$$\begin{aligned} 1) \quad \sigma_{\mathcal{T}_{3\text{D}}(\mathbf{v})}^2[\xi_1, \xi_2, \xi_3] &= \begin{cases} \Psi_{\text{RND}_k}[\xi_1, \xi_2] + h \Psi_{\text{FPN}_k}[\xi_1, \xi_2] & \xi_3 = 1, \\ \Psi_{\text{RND}_k}[\xi_1, \xi_2] & \xi_3 \neq 1, \end{cases} \\ 2) \quad \sigma_{\mathcal{T}_{3\text{D}}(\mathbf{v})}^2[\xi_1, \xi_2, \xi_3] &= \Psi_{\text{RND}_k}[\xi_1, \xi_2] + \Psi_{\text{FPN}_k}[\xi_1, \xi_2], \end{aligned}$$

where Ψ_{RND_k} and Ψ_{FPN_k} are $\mathcal{T}_{2\text{D}}$ -PSDs of $\boldsymbol{\eta}_{\text{RND}_k}$ and $\boldsymbol{\eta}_{\text{FPN}_k}$, and $\xi_3 = 1$ corresponds to the temporal DC. RF3D adaptively defines $\sigma_{\mathcal{T}_{3\text{D}}(\mathbf{v})}$ also for all intermediate cases based on Ψ_{RND_k} , Ψ_{FPN_k} , and the specific $\{\mathbf{d}_i\}_i$ [13]. The subindex k in the notation for Ψ_{RND_k} and Ψ_{FPN_k} emphasizes that these PSDs may vary at each iteration of CVR.

C. Adaptive Correlated Noise Estimation

Even though $\boldsymbol{\eta}_k$ does not coincide with the residual signal $\rho_k \mathbf{A}^\dagger \mathbf{r}_k$ (3), we argue that the statistics of Ψ_{RND_k} and Ψ_{FPN_k} for $\boldsymbol{\eta}_k$ can be more conveniently estimated from $\rho_k \mathbf{A}^\dagger \mathbf{r}_k$, since the residual signal has the benefit of being readily computed in IST and, compared to \mathbf{z}_k , it does not feature a dominating \mathbf{x} component.

Thus, since the FPN component $\boldsymbol{\eta}_{\text{FPN}_k}$ is temporally invariant, we estimate it as the temporal average of $\rho_k \mathbf{A}^\dagger \mathbf{r}_k$ over all the frames in the video. The random noise component $\boldsymbol{\eta}_{\text{RND}_k}$ can be then captured by subtracting the estimated $\boldsymbol{\eta}_{\text{FPN}_k}$ from $\rho_k \mathbf{A}^\dagger \mathbf{r}_k$. Finally, we can compute the 2D root-PSDs $\Psi_{\text{RND}_k}^{1/2}$ and $\Psi_{\text{FPN}_k}^{1/2}$ through the MAD over all blocks extracted from estimated $\boldsymbol{\eta}_{\text{RND}_k}$ and $\boldsymbol{\eta}_{\text{FPN}_k}$ at each iteration.

III. EXPERIMENTS AND DISCUSSION

We compare the CVR results by our proposed IST-wRF3D/-cRF3D with those of two state-of-the-art methods:

GMM-TP [8]: A Gaussian mixture model (GMM) approach where the priors of the GMM are learned by the patches extracted from the training datasets.

MMLE-GMM [9]: A maximum marginal likelihood estimator (MMLE) that maximizes the marginal likelihood of the GMM of \mathbf{x} given only the measurement \mathbf{y} when GMM-TP with full overlapping patches is used to initialize it.

We measure the objective quality of the recovery by the peak signal-to-noise ratio (PSNR) of the estimate $\tilde{\mathbf{x}}$, i.e. $\text{PSNR}(\mathbf{x}, \tilde{\mathbf{x}}) = 10 \log_{10} (n I_{\text{max}}^2 / \|\mathbf{x} - \tilde{\mathbf{x}}\|^2)$ where I_{max} is the peak of the noise-free video \mathbf{x} . We consider the $256 \times 256 \times 32$ *NBA* video sequence (both ground-truth and compressive measurements) used in [9]. The compressive measurements are acquired by the CACTI system: each video frame is masked with a shifted version of single random binary mask drawn from the Bernoulli distribution with probability

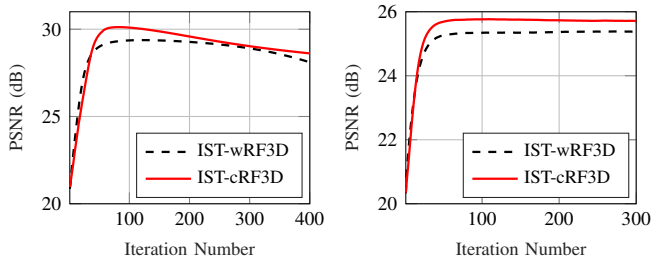


Figure 1. Evolution of PSNR versus iteration number for the case of noise-free (left plot) and noisy measurements (right plot).

0.5, and the temporally-compressed $256 \times 256 \times 4$ measurements are constituted by summing such masked frames in consecutive groups of $p=8$. The aggressive compressive acquisition of a dynamic scene with complex motions of non-rigid bodies makes this a challenging and representative benchmark for testing CVR. We consider CVR reconstruction from both noise-free and noisy measurements (AWGN, SNR = 15 dB). Since the available codes for GMM-TP [16] and MMLE-GMM [17] methods deal with CVR from noise-free measurements while producing poor results in the case of noisy measurements, for the sake of fair comparison we consider these methods only with noise-free measurements.

For our IST-based RF3D CVR method, we fix $\rho_k = 2$ and adopt spatiotemporal volumes composed of 8×8 blocks and spanning a temporal extent of $h = 9$ frames. In GMM-TP and MMLE-GMM, we follow the settings used in [9] by the same authors to train the underlying GMM parameters and use them to initialize the MMLE.

Fig. 1 draws the evolution of PSNR versus iteration number under noise-free and noisy measurements. At least in terms of PSNR, most of recovery happens during the first 80 iterations, consistently for the four different cases. We thus compare individual reconstructed

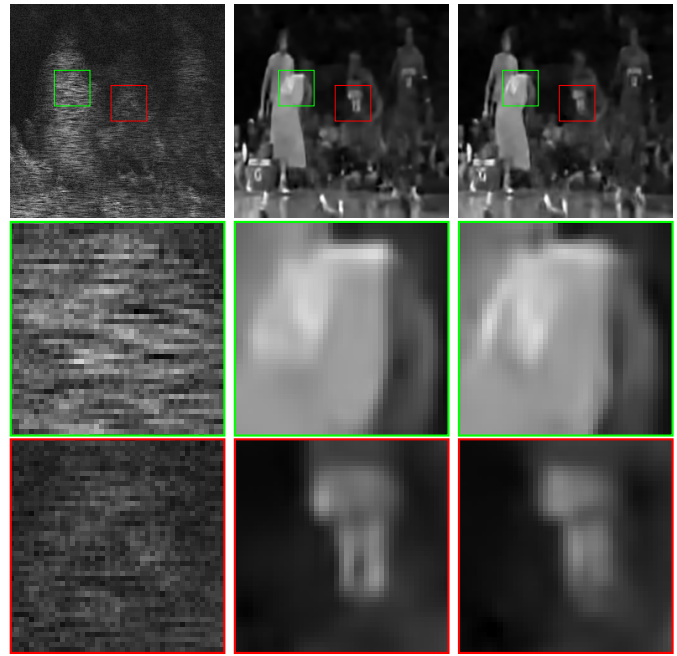


Figure 3. From left to right, 1st row: temporally encoded frame #2 obtained from combining masked original frames #9, ..., #16 which is then corrupted by AWGN (SNR = 15 dB); the recovered frames via IST-wRF3D (25.20 dB) and IST-cRF3D (25.76 dB), respectively. 2nd and 3rd rows: cropped+zoomed portions of 1st row.

frames at $k = 80$ in Figs. 2 and 3, leaving the study of adaptive stopping criteria to future work. It can be seen that the proposed IST-cRF3D gives the best performance with a better detection of fast and complicated motions with less artifacts around moving objects in the video and without excessive loss of details.

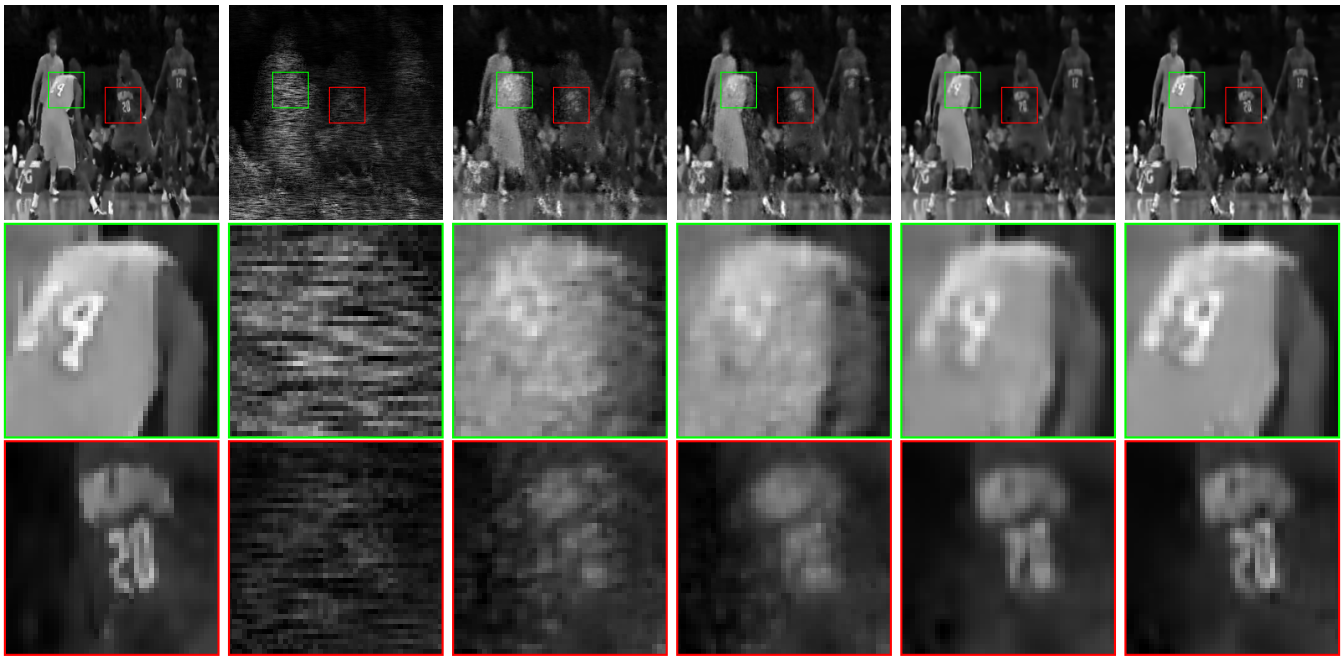


Figure 2. From left to right, 1st row: original frame #15; temporally encoded frame #2 obtained from combining masked original frames #9, ..., #16; the recovered frames from noiseless measurements via GMM-TP [8] (PSNR: 24.13 dB), MMLE-GMM [9] (26.96 dB), IST-wRF3D (29.14 dB) and IST-cRF3D (30.12 dB), respectively. 2nd and 3rd rows are respectively cropped+zoomed portions of the 1st row.

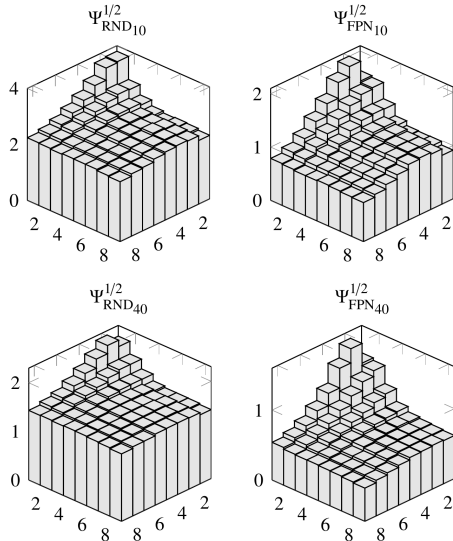


Figure 4. 1st (resp. 2nd) row: the 2D root-PSDs of the modeled η_{RND_k} and η_{FPN_k} captured in the residual errors of the reconstructed *NBA* video via IST-cRF3D at iteration $k = 10$ (resp. $k = 40$). These root-PSDs are defined with respect to the 8×8 2D DCT. The DC coefficient and the highest frequency coefficient are diametrically opposed, with the former located in the top corner (1, 1) and latter in the bottom corner (8, 8).

Figs. 4 and 5 show, respectively, the 2D root-PSDs and 3D PSDs of the modeled noise captured in the residual signals at two different iterations of IST-cRF3D. These 2D root-PSDs (resp. 3D PSDs) are defined with respect to the 8×8 2D block DCT (resp. $8 \times 8 \times 9$ 3D DCT) applied on the non-overlapping 2D blocks (resp. 3D volumes) extracted from the residual signal. As can be seen from Figs. 4 and 5, the spectrum of each residual signal of IST-cRF3D has a noticeable anisotropic behavior implying the correlation in the modeled noise.

In the shrinkage step, λ scales the noise root-PSD to modulate the filtering strength (5), and choosing a large (small) λ causes oversmoothing (undersmoothing). While for simplicity we do not vary λ as the recovery progresses, we did experiments for several different values of λ in IST-cRF3D (resp. IST-wRF3D) and found out that the highest PSNR at iteration 80 and further iterations is obtained for λ being 11 and 3.56 (resp. 70 and 12), respectively, in the case of noise-free and noisy measurements. The choice of $\lambda = 3.56$ in IST-cRF3D in the case noisy measurements appears reasonable and roughly matches the universal threshold factor $\sqrt{2 \log(b^2 h)}$ for an array of size $b \times b \times h$ [18]. The best λ values for the other three cases are however larger. In IST-wRF3D, we compute the residual noise variance from the highest-frequency portion of the anisotropic spectrum of the residual signal, which as shown in Fig. 5 includes the weakest components in the spectrum. For an AWGN model, the noise variance coincides with the level of the flat PSD, which means that in practice the algorithm operates with a significant underestimate of the noise. Therefore, an increase of λ in the case of AWGN yields a higher filtering strength to partly compensate the discrepancy between noise models. We observed that in the noise-free case, both IST-cRF3D and IST-wRF3D benefit from a comparably large λ in order to recover the signal during the first few iterations ($k < 20$). This seems consistent with parameter choices in other iterative recovery algorithms, such as [19], [20]. We speculate that the PSNR decay that takes place towards later stages of CVR ($k > 100$, shown in Fig. 1, left) is due to oversmoothing of the data caused by a fixed and large

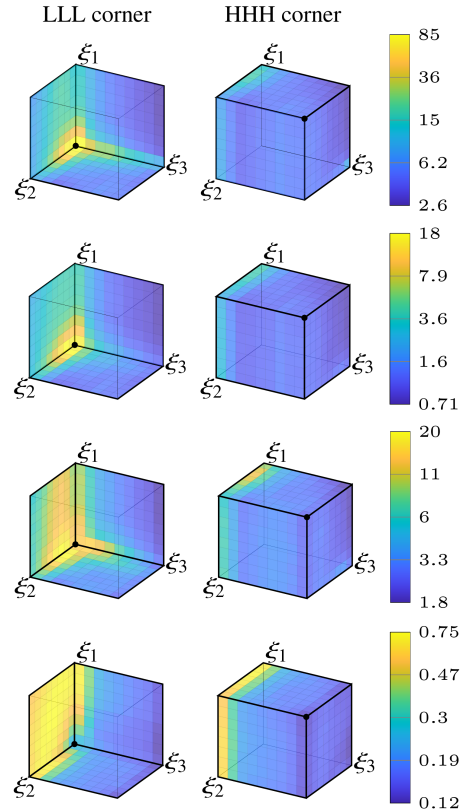


Figure 5. 1st and 2nd (resp. 3rd and 4th) rows: spatiotemporal correlation in the residual errors of the reconstructed *NBA* video via IST-cRF3D (resp. IST-wRF3D) when the corresponding temporally-compressed measurements are noise-free, visualized as noise PSD with respect to the $8 \times 8 \times 9$ 3D DCT applied on the 3D volumes, respectively, extracted from the residual signals at iteration $k = 10$ and $k = 40$. Only the lowest- and highest-frequency faces of each 3D PSD cube are shown.

λ . Hence, for future implementations of IST-cRF3D, we may consider a λ that progressively decreases to $\sqrt{2 \log(b^2 h)}$ as k grows. For a MATLAB/C implementation of the above experiments on a computer equipped with Intel Core i7 2.8-GHz CPU, IST-wRF3D and IST-cRF3D respectively take 10.9 and 13.2 seconds per iteration. Most of the extra time in IST-cRF3D is dedicated to estimating the root-PSDs (2.1 seconds).

IV. CONCLUSIONS

We modeled the noise at each iteration of CVR as stationary spatiotemporally correlated noise, comprised of two spatially correlated random and fixed-pattern noise components that are adaptively estimated from the residual signal. It is worth noting that this is not a perfect model of the actual underlying noise, however it is much more flexible than the common i.i.d. model. We embed the proposed model within the RF3D denoising algorithm, which we adopted as the sparsity-promoting filter in IST. Experimental analysis demonstrates the superior subjective and objective performance of the proposed IST-based CVR method employing RF3D as the denoiser via adaptive correlated noise model versus an i.i.d. model and state-of-the-art techniques.

REFERENCES

- [1] D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *Proc. IEEE CVPR*, 2011, pp. 329–336.
- [2] A. Veeraraghavan, D. Reddy, and R. Raskar, "Coded strobing photography: Compressive sensing of high speed periodic videos," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 4, pp. 671–686, 2011.
- [3] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," *Opt. Express*, vol. 21, no. 9, pp. 10526–10545, 2013.
- [4] X. Yuan, J. Yang, P. Llull, X. Liao, G. Sapiro, D. J. Brady, and L. Carin, "Adaptive temporal compressive sensing for video," in *Proc. IEEE ICIP*, 2013, pp. 14–18.
- [5] X. Liao, H. Li, and L. Carin, "Generalized alternating projection for weighted- $\ell_{2,1}$ minimization with applications to model-based compressive sensing," *SIAM J. Imaging Sci.*, vol. 7, no. 2, pp. 797–823, 2014.
- [6] R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, and A. K. Katsaggelos, "High spatio-temporal resolution video with compressed sensing," *Opt. Express*, vol. 23, no. 12, pp. 15992–16007, 2015.
- [7] L. Gao, J. Liang, C. Li, and L. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," *Nature*, vol. 516, no. 7529, p. 74, 2014.
- [8] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using Gaussian mixture models," *IEEE Trans. Image Process.*, vol. 23, no. 11, pp. 4863–4878, 2014.
- [9] J. Yang, X. Liao, X. Yuan, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Compressive sensing by learning a Gaussian mixture model from measurements," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 106–119, 2015.
- [10] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [11] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [12] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5117–5144, 2016.
- [13] M. Maggioni, E. Sánchez-Monge, and A. Foi, "Joint removal of random and fixed-pattern noise through spatiotemporal video filtering," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4282–4296, 2014.
- [14] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [15] F. R. Hampel, "The influence curve and its role in robust estimation," *J. Am. Stat. Assoc.*, vol. 69, no. 346, pp. 383–393, 1974.
- [16] J. Yang, "GMM-TP (version 1.1)," <https://github.com/jianboyang/GMM-TP>, 2014.
- [17] —, "MMLE-GMM (version 1.1)," <https://github.com/jianboyang/MMLE-GMM>, 2015.
- [18] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [19] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Proc. IEEE ICIP*, 2007, pp. 549–552.
- [20] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, 2013.