Fully Decomposable Compressive Sampling With Joint Optimization for Multidimensional Sparse Representation

Wenrui Dai, Member, IEEE, Yong Li, Student Member, IEEE, Junni Zou, Member, IEEE, Hongkai Xiong, Senior Member, IEEE, and Yuan F. Zheng, Fellow, IEEE

Abstract—Conventional compressive sampling methods cannot efficiently exploit structured sparsity for sampling multidimensional signals like video sequences. In this paper, we propose a fully decomposable compressive sampling model that adopts the Kronecker product framework to exploit the structured sparsity spanning multidimensional signals. It enables efficient sampling in a progressive fashion by retaining the block-diagonal feature of Kronecker products. A synthetic sensing matrix is developed for joint optimization over sampling signals with multiple dimensions. Instead of adjusting global Gram matrix, separable minimization of mutual coherence in multiple dimensions is jointly formulated for a stable recovery with enhanced convergence rate. Sampling rate allocation is considered to improve recovery performance based on the decomposable compressive sampling. The proposed model is employed in video acquisition for temporal sparsity along motion trajectory. Experiment results show that the proposed model can improve the recovery performance with a reduced number of necessary samples in comparison to the state-of-the-art methods.

Index Terms—Sparse representation, matrix optimization, compressive video sampling, video acquisition.

I. INTRODUCTION

MULTIMEDIA signal acquisition usually requires a large number of sensors to sample the raw data. Shannon’s celebrated theorem demonstrates that traditional sampling requires a minimum rate twice the highest frequency of the signal, namely the Nyquist rate, to achieve sufficient sampling without loss of any information. Therefore, traditional data acquisition systems commonly employ a sample-then-compress framework to acquire signals. However, most signals of interest have concise (sparse) linear representations with respect to some basis, e.g., Fourier or wavelet basis, which means that the transformed (compressed) coefficients equal or close to zero could be discarded. Recognizing the deficiency of directly sampling raw data, compressive sampling (CS) [1], [2]) has been developed to make sub-Nyquist sampling by recording fewer projections into an incoherent set of measurement vectors.

The compression-into-sampling system can generate irregular and inadequate measurements conditioned on the actual sparsity of signals, so that the necessary sampling rate can be substantially reduced. Since exact reconstruction via nonlinear recovery regularized by $l_0$ quasi-norm is an NP-hard problem, the $l_1$ minimization [3] was commonly adopted as a tractable alternative for sparse recovery. Its recovery performance is demonstrated to be guaranteed under the constraints of the null space property (NSP) [4]) and the restricted isometry property (RIP) [5]) for measurement vectors. However, conventional CS [6], [7] relies on a vectorized representation for the whole signal of interest, which would obscure the intrinsic structure of multidimensional signals. As a result, these methods are restricted in applications for multidimensional signals, e.g., hyperspectral imaging [8]–[10], magnetic resonance imaging (MRI [11]), and image/video acquisition [12].

Recently, compressive sampling has been inspired to investigate high-speed spatial and temporal imaging resolution with a reversed-complexity manner, which would favor the ubiquitous and interactive multimedia access for increasing mobile communication. To facilitate a distributed CS protocol, the sparsifying basis and synthetic sensing matrix for sampling are optimized for compact representation of varying sparsities. Traced back to [13], compressive imaging was first employed into video representation and coding based on a single-pixel camera model, where 3-D wavelet transform was adopted to improve the frame-by-frame reconstruction with its 2-D counterpart. In [14], an iterative multiscale framework was developed to exploit motion estimation and compensation for enhanced sparsity at finer scales in the 3-D wavelet domain. Cossalter et al. [15] combined compressive sensing and the information obtained by tracking the motion objects. To improve sampling efficiency, each frame
was further segmented into non-overlapping blocks which were estimated with a linear combination of blocks in previous frames [16], [17]. For distributed compressive video sensing, an efficient modified gradient projection algorithm was developed for sparse reconstruction at the decoder [18]. To further reduce random measurements, Ma et al. [19] adopted a pseudorandom downsampling of the 2-D Fourier transform for fast online encoding of each frame and the approximate message passing approach for offline decoding with rapid convergence.

To improve the recovery performance, motion estimation and compensation were introduced into the block-based CS framework. In essence, a video sequence was decomposed into reference frames and the other CS frames reconstructed based on adaptive bases extracted from adjacent reconstructed reference frames. Liu et al. [20] developed an adaptive block-based CS framework for video, which classified blocks with their interframe correlations and texture region. Mum et al. [21] improved the quality of reconstructed frames by combining the motion-compensated prediction from the reference frame and the recovered residual with the smooth projected Landweber reconstruction. Later, this method was extended to multi-hypothesis prediction based on multiple reference frames [22]. In [23], Karhunen-Loève transform (KLT) bases were adopted to exploit long-term inter-frame correlations, which adaptively represented sparsities for each block based on reference frames. These methods achieved a potential gain over conventional video acquisition methods. However, they cannot provide theoretical tractability and performance guarantees, as explicit utilization of motion information hampers the derivation of an adaptive sensing matrix to ensure strict sparsity property for the spatio-temporal correlations.

Kronecker product framework was first introduced for multidimensional signals in Kronecker compressive sensing (KCS [24]), where spatio-temporal sparsity was jointly modeled with the Kronecker product of a 2-D spatial sparsifying basis and 1-D temporal sparsifying basis. KCS achieved the stable recovery for compressive video sampling with a block-diagonal sensing matrix synthesized by sensing matrix and full-rank identity matrix for spatial and temporal dimension. Thus, the recovery performance is degraded in KCS due to its redundant sampling in temporal dimension. To improve the recovery efficiency, multi-way compressed sensing (MWCS [25]) developed a two-step process to utilize the Kronecker sensing structure, including fitting a low-rank model in compressed domain and per-mode decompression. However, MWCS is an NP-hard problem, as it heavily relies on tensor rank estimation. Recently, generalized tensor compressive sensing (GTCS [26]) was developed to employ the Kronecker structure for simultaneous acquisition and representation from all tensor modes, which considerably reduces the computational complexity for recovery by skipping the vectorization of video frames and tensor rank estimation. However, its recovery performance is degraded over the fixed basis (e.g. 3D DCT basis) due to the error accumulation with tensor mode continuing.

Furthermore, sensing matrix optimization is not considered in [24]–[26], which would substantially enhance the CS reconstruction with derived basis. However, existing optimization methods [27]–[30] cannot perform efficiently for synthetic sensing matrix generated by Kronecker products. Tremendous computation burden is led by iteratively optimizing the sensing matrix to force the Gram matrix to identity matrix. Considering the separability of Kronecker matrix, it would be desirable to design an efficient algorithm based on separable metrics like mutual coherence.

This paper proposes a fully decomposable compressive sampling model to exploit the structured sparsity in multidimensional signals. The contribution of this paper is two-fold. Firstly, the proposed model adopts a rank-deficient matrix to improve compressive sampling in multiple dimensions. Integrating with the Kronecker product framework, it can achieve a higher compression ratio with a guarantee of recovery performance. Secondly, the proposed model optimizes the synthetic sensing matrix by jointly minimizing the mutual coherence of the projection matrix. Decomposable optimization is allowed for sampling signals with multiple dimensions using sampling rate-distortion optimization. Given sparsifying basis, the optimized sensing matrix can achieve both reliable performance and improved convergence rate in reconstruction.

To validate the efficacy of the proposed model, we employed it into video acquisition to evaluate its recovery performance under various sampling rates. It improves the compression ratio with a guaranteed recovery performance by performing compressive sampling in both spatial and temporal dimensions. The proposed model is shown to outperform conventional CS schemes and latest Kronecker product based methods in terms of PSNR and visual quality. We also evaluated the sensing matrix optimization and sampling rate allocation scheme, which can effectively improve the recovery performance and computational complexity based on given number of measurements.

The reminder of this paper is organized as follows. In Section II, the relevant preliminaries and analysis on Kronecker compressive sensing (KCS) are introduced. Section III proposes the fully decomposable compressive sampling model, including its formulation, upper bound of mutual coherence, stable recovery and advantages to KCS. Sensing matrix optimization for fully decomposable CS is developed in Section IV with a well-designed algorithm for sampling rate allocation. In Section V, fully decomposable CS is employed into video acquisition with joint optimization for spatio-temporal sparsity. Section VI provides extensive experiment results to evaluate the proposed video acquisition schemes. Finally, we conclude this paper in Section VII.

II. PRELIMINARIES AND MOTIVATION

In the rest of this paper, we reserve normal symbols to scalar variables and boldface symbols to vector variables for clarity. Table I specifies the frequently-used variables in this paper.

A. Compressive Sampling

A signal $x \in \mathbb{R}^N$ is said to be $K$-sparse over the basis $\Psi$, if $x$ can be decomposed by $\Psi \theta$ with $\|\theta\|_0 = K \ll N$. According to CS theory, the $K$-sparse signal $x$ can be exactly reconstructed from $M = O(K \log(N/K))$ linear projections onto the
sensing matrix $\Phi$. Here, the nonadaptive sensing matrix $\Phi \in \mathbb{R}^{M \times N}$ is designed to satisfy the incoherent condition with fixed sparsifying basis $\Psi$ [31]. It has been shown that the independent and identically distributed (i.i.d.) Gaussian or Bernoulli random matrices can facilitate a simple construction of nonadaptive sensing matrices [32]. The signal $x$ is recovered by finding the sparsest vector $\hat{\theta}$ consistent with the measurement vector $y = \Phi x = \Phi \Psi \theta$, which can be solved by the $\ell_0$-minimization problem [33], [34] or the computationally tractable $\ell_1$ optimization problem.

$$\hat{\theta} = \arg \min_{|\theta|_1} \| y - \Phi \Psi \theta \|_2 \leq \sigma. \quad (1)$$

Commonly, (1) can be solved through Basis Pursuit [35]–[38] at the computational complexity of $O(N^3)$.

Compressive sampling relies on the fundamental principle of incoherent sampling, which makes the sensing matrix $\Phi$ and the sparsifying basis $\Psi$ to be as incoherent as possible. This principle can be guaranteed by the restricted isometry property (RIP [5], [39]) under constraints on $\Phi$ and $\Psi$. However, it requires a combinatorial computational complexity to verify a general matrix $D = \Phi \Psi$. Thus, mutual coherence is introduced for more concrete recovery based on $\Phi$ and $\Psi$ [40].

**Definition 1 (Mutual Coherence):** The mutual coherence of the orthonormal bases $\Phi \in \mathbb{R}^{N \times N}$ and $\Psi \in \mathbb{R}^{N \times N}$ is the maximum absolute value for the inner product between elements of the two bases.

$$\mu(\Phi, \Psi) = \max_{1 \leq i,j \leq N} \left| \langle \phi_i, \psi_j \rangle \right|, \quad (2)$$

Under lower coherence between $\Phi$ and $\Psi$, the signal $x$ can be recovered by a smaller set of CS samples with less mutual information.

**B. Compressive Sampling for Multi-Dimensional Signals**

Compressive sampling for multi-dimensional signals involves complex models for high-dimensional sparsity. Conventional (global) CS methods reshape the multi-dimensional signal into a single 1-D vector without considering the space $X$ spans.

Thus, $X$ is sampled by directly multiplying the 1-D vector with a global dense matrix $\Phi$.

For multi-dimensional signals, it is critical to find appropriate bases to jointly model the high-dimensional sparsity. One promising way is to combine the bases for sparsity in various dimensions with Kronecker product. Here, we define the Kronecker product $(A \otimes B)_{P \times QS}$ of two matrices $A$ and $B$ with sizes $P \times Q$ and $R \times S$.

$$A \otimes B = \begin{bmatrix} A(1,1)B & A(1,2)B & \cdots & A(1,Q)B \\ A(2,1)B & A(2,2)B & \cdots & A(2,Q)B \\ \vdots & \vdots & \ddots & \vdots \\ A(P,1)B & A(P,2)B & \cdots & A(P,Q)B \end{bmatrix}, \quad (3)$$

where $A(p,q)$ is the element at $p$-th row and $q$-th column in matrix $A$. In the Kronecker product framework, a multi-dimensional sparsifying basis $\Psi$ is constructed under the assumption that multi-dimensional signal $X$ is sparse or compressible conditioned on basis $\Psi_k$ along its $k$-th dimension. Thus, $\Psi$ can be synthesized with the Kronecker product of $\Psi_k$, $1 \leq k \leq E$ for all the $E$ dimensions of $X$.

$$\Psi = \Psi_1 \otimes \Psi_2 \otimes \cdots \otimes \Psi_E. \quad (4)$$

Thus, the $i$-th column $\psi^i$ of $\Psi$ can be represented by

$$\psi^i = \psi_1^i \otimes \psi_2^i \otimes \cdots \otimes \psi_E^i, \quad (5)$$

where $\psi_k^i$ is the corresponding column in sparsifying basis $\Psi_k$ for the $k$-th dimension. It is demonstrated in [41] and [42] that hyperbolic basis is optimal for (4) among bases derived from its universality along different dimensions. For example, a 2-D wavelet transform with basis $\Psi_s$ and a 1-D wavelet transform with basis $\Psi_t$ can sparsely represent an image and a smooth or piecewise smooth sequence of pixels, respectively. The Kronecker product of $\Psi_s$ and $\Psi_t$ would construct a hyperbolic wavelet basis with different degrees of smoothness for each dimension [43]. On the contrary, global CS would require multiplexing sensors to operate simultaneously along all dimensions, which are prohibitive and impractical for video acquisition due to the physical complexity and latency.

Without loss of generality, we consider the signal $X$ of $E$ dimensions. The mutual coherence of the sensing matrix $\hat{\Phi} = \Phi_1 \otimes \cdots \otimes \Phi_E$ and sparsifying basis $\hat{\Psi} = \Psi_1 \otimes \cdots \otimes \Psi_E$ can be formulated as

$$\mu(\hat{\Phi}, \hat{\Psi}) = \prod_{k=1}^{E} \mu(\Phi_k, \Psi_k) \leq \min_{1 \leq k \leq E} \mu(\Phi_k, \Psi_k). \quad (6)$$

Equation(6) implies that KCS requires less number of samples than performing CS independently on each dimension (aka, independent CS) to recover the signals of interest, as $\mu(\hat{\Phi}, \hat{\Psi})$ will not exceed $\mu(\Phi_k, \Psi_k)$, $1 \leq k \leq E$.

To summarize, global CS directly performs compressed sampling and global reconstruction for the entire signals. However, it is computationally intensive in practice due to its exponentially increasing data volume and complexity. Independent CS employs individual compressed sampling and individual reconstruction for each dimension. For multi-dimensional signal $X \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_E}$, independent CS performs sampling and
recovery for \( i = 1, \ldots, N_k \) along the \( k \)-th dimension in an isolated manner.

\[
\text{vec}(Y_i) = \Phi_k \text{vec}(X_i) = \Phi_k \Psi_i \theta_i,
\]

where \( X_i \in \mathbb{R}^{N_1 \times N_{i-1} \times N_{i+1} \times \cdots \times N_E} \) and \( \Phi_k \in \mathbb{R}^{M_k \times (N_1 \times N_{i-1} \times N_{i+1} \times \cdots \times N_E)} \). For example, independent CS separately handles each frame of video sequences with \( \Psi_i \) and \( \Phi_i \). Thus, it degrades the sampling and recovery performance by neglecting the temporal sparsity. KCS provides a tractable alternative to balance the complexity and performance for sampling and recovery. It employs the Kronecker product framework to generate sensing matrix for joint recovery.

\[
Y = (\Phi_1 \otimes \cdots \otimes \Phi_E) X = (\Phi_1 \otimes \cdots \otimes \Phi_E) (\Psi_1 \otimes \cdots \otimes \Psi_E) \theta.
\]

For compact representation, \( \Phi_k, k = 2, \ldots, E \) are commonly set to identity matrices \( \mathbf{I} \) in KCS. Thus, sampling is not performed along these dimensions. However, KCS cannot sufficiently capture the temporal sparsity in video sequences, as the identity sensing matrix is adopted for temporal sampling, as shown in (21). This fact means that the measurements \( Y \) individually sampled from frames \( F_1, \ldots, F_T \) are redundant due to the temporal correlations in video sequences. The spatial measurements for adjacent frames could be further clustered with temporal sparsity to reduce the necessary samples for recovery.

III. FULLY DECOMPOSABLE COMPRESSIVE SAMPLING (FDCS)

Recall the synthetic sensing matrix \( \Phi = \Phi_\ast \otimes \mathbf{I} \) in KCS, where \( \mathbf{I} \) and \( \Phi_\ast \) can be regarded as factorized sensing matrices for temporal and spatial dimension, respectively. The identity matrix \( \mathbf{I} \) offers complete temporal components, while \( \Phi_\ast \) determines the overall sampling rate by compressing the spatial measurements. In this section, we propose a general CS framework for multidimensional signals to perform decomposable sampling in each of its subspaces.

A. Formulation

Consider a multidimensional signal \( X \) represented by a union of \( E \) subspaces \( S_1, S_2, \ldots, S_E \), where \( S_i \) is spanned by the sparsifying basis \( \Psi_i \). It is obvious that KCS cannot sufficiently exploit the sparsity in the multidimensional signals with \( E \geq 2 \), as its measurements sampled with the full-rank identity matrix are redundant in the temporal dimension. To develop an efficient sampling model, we redefine the overall sensing matrix \( \Phi \) based on the rank-deficient sensing matrices \( \Phi_i \) for \( S_i \). The measurements \( Y \) are obtained by sampling \( X \) based on a Kronecker product framework.

\[
\text{vec}(Y) = (\Phi_1 \otimes \cdots \otimes \Phi_E) \cdot \text{vec}(X),
\]

where \( \Phi_i \) is the \( M_i \times N \) sensing matrix for \( S_i \) and \( M = \prod_{i=1}^{E} M_i \) for all the \( E \) dimensions, respectively. According to [44], (7) can be rewritten in matrix product from.

\[
Y = \bigotimes_{i=1}^{E} \Phi_i \Psi_i \theta_i,
\]

where \( \Psi_i \) is the vectorization operator for the \( i \)-th subspace \( S_i \) that \( X \) lives in. Consequently, we define the fully decomposable compressive sampling for multidimensional signal \( X \).

Definition 2 (Fully Decomposable Compressive Sampling (FDCS)): A multidimensional signal \( X \) is called to be fully decomposed over a union of \( E \) subspaces \( S_i \), when there exists sensing matrices \( \Phi_i \) and sparsifying bases \( \Psi_i, i = 1, \ldots, E \) satisfying that

\[
Y = \bigotimes_{i=1}^{E} \Phi_i \Psi_i \theta_i,
\]

where \( \theta_i \) is the decomposable sparse representation for the projection of \( X \) onto subspace \( S_i \).

Here, the transform \( \Phi_i \Psi_i \theta_i \) refers to the measurements corresponding to the subspace \( S_i \) spanned by the sparsifying basis \( \Psi_i \). To be concrete, \( \Phi_i \in \mathbb{R}^{M_i \times N} \) in (8) compresses projection in subspace \( S_i \) into a vector of \( M_i \) measurements, as it is always underdetermined for \( M_i < N \). When \( X \) is decomposable over the \( E \) subspaces, operation \( \Phi X \) obtains CS measurements of \( X \) in a progressive fashion based on \( \Phi_i \) and \( \Psi_i \) for \( S_i \). When progressively performing compressive sampling from \( i = 1 \) to \( E \), \( \Phi \in \mathbb{R}^{M_\ast \times N} \) clusters projection of signals at corresponding position in a partial union \( S_1 \otimes \cdots \otimes S_{i-1} \) of subspaces into an \( M_\ast \)-tuple vector.

In comparison to KCS, FDCS adopts an underdetermined matrix \( \Phi_i \) instead of the identity matrix \( \mathbf{I} \) for each subspace, so that it can simultaneously exploit the structured sparsity of these subspaces to obtain the compressed measurements \( Y \).
into a matrix $Y$ with 8 measurements, while KCS requires 
12 measurements to recover the cube. FDCS simultaneously 
exploits spatial and temporal correlations to generate fewer 
measurements that are randomly distributed.

B. Upper Bound for Mutual Coherence

In this section, we estimate the redundancy of $X$ in the 
sense of mutual coherence derived by the FDCS. Given 
sensing matrix $\Phi$ and sparsifying basis $\Psi$, the upper bound 
of mutual coherence is developed to guarantee the recovery performance 
with Kronecker product framework. Given sensing 
matrix $\Phi \in \mathbb{R}^{M \times N}$, the sampling rate is defined as $r = M/N$.

In the Kronecker product framework, the sparsifying basis for 
$X$ is $\Psi = \Psi_1 \otimes \cdots \otimes \Psi_E$. The corresponding sensing matrices 
for fully decomposable CS is $\Phi^{(F)} = \Phi_1 \otimes \cdots \otimes \Phi_E$, while $\Phi_i$, 
$i = 2, \ldots, E$ are set to identity matrices $I$ for KCS. In Theorem 
1, the mutual coherence for the fully decomposable CS model 
is demonstrated to be upper-bounded by the one derived from 
KCS.

Theorem 1: Given corresponding sampling rates $r_i \in (0, 1)$ 
for $\Phi_i$, $i = 1, \ldots, E$, in the fully decomposable CS model 
and the overall sampling rate $r = \prod_{i=1}^E r_i$, under Gaussian random 
sampling, with a high probability asymptotically approaching 1, 
the mutual coherence $\mu(\hat{\Phi}^{(F)}, \hat{\Psi})$ for FDCS is upper-bounded 
by the mutual coherence $\mu(\hat{\Phi}^{(K)}, \hat{\Psi})$ for KCS.

$$\mu(\hat{\Phi}^{(F)}, \hat{\Psi}) \leq \mu(\hat{\Phi}^{(K)}, \hat{\Psi}) \quad (10)$$

Proof: Please refer to Appendix A.

C. Stable Recovery

In this section, we demonstrate that stable recovery can be 
achieved for fully decomposable CS. Given the multidimensional 
$X$, denote $D = \Phi \Psi$ and $D_i = \Phi_i \Psi_i$ the projection 
dictionary for the multidimensional signals $X$ and its $i$-th subspace $S_i$. 
Assuming that the projection of $X$ onto each subspace $S_i$ is $K_i$-sparse, we show that there exists $D$ stable for block sparse vector $u$ with $2K_i$-sparse component $u_i$ for $S_i$.

Theorem 2 (Stable Recovery): Given projection dictionary 
$D = D_1 \otimes \cdots \otimes D_E$, it is stable for arbitrary block-sparse vector $u$ if and only if its $2K_i$-sparse component $u_i$ for $S_i$ satisfies

$$C_i' \|u\|_2^2 \leq \|D_i u_i\|_2^2 \leq C_i'' \|u\|_2^2 \quad (11)$$

Here, $u_i = \theta_1 - \theta_2$ is the difference between two $K_i$ sparse vector $\theta_1$ and $\theta_2$.

Proof: Please refer to Appendix B.

Theorem 2 implies that the multidimensional signal $X$ can 
be recovered when block RIP is maintained for $D_i$ in the $i$-th 
subspace $S_i$. For each subspace $S_i$, its vectorized compo-
nent $V_iX$ can be separately recovered with $\ell_2,1$ optimization, 
when the number of measurements is greater than the sparsity 
$K_i$. Thus, the multidimensional signals $X$ can be recovered by 
progressively combining the vectorized projection from all the 
subspaces $S_i$.

In the fully decomposable CS model, we use a series of under-
determined matrices $\Phi_1, \ldots, \Phi_E$ as the sensing matrices corresponding 
to its subspaces $S_1, \ldots, S_E$. Its measurements enable a 
high compression ratio which comes from the extra compressed 
temporal dimensions. The fully decomposable CS model can 
sufficiently exploit the structured sparsity in multiple subspaces 
to derive irregular measurements to eliminate their correlations.

In comparison to KCS, the fully decomposable CS model would 
linearly increase the computational complexity for sampling, as 
(8) shows that the proposed scheme could progressively sample 
each subspace. The proposed sampling structure retains the feature of traditional CS with the exception of an extra cache to 
store temporary frame snapshot. This cache is utilized for the 
subsequent second-order CS. Once the samples of fully decom-
posable CS are received, an efficient joint recovery will help us 
reconstruct the original 3-D signal.

IV. SENSING MATRIX OPTIMIZATION FOR 
MULTIDIMENSIONAL SIGNALS

In compressive sampling, the reconstruction performance de-
deps on the compressibility of the signal, the choice of the 
reconstruction algorithm, and the incoherence between sensing 
matrix and the sparsifying basis. In this section, we develop an 
algorithm to further optimize the synthetic sensing matrix for 
reconstruction in the fully decomposable CS model.

A. Minimization of Mutual Coherence

The randomly generated sensing matrix can enjoy additional 
benefit from universality, but it cannot perfectly satisfy the in-
coherence requirements. To improve the sampling efficiency 
and reconstruction performance, it is promising to develop an
optimized algorithm to substantially reduce the coherence between sensing and sparsifying matrices. In the fully decomposable CS model, it is imperative to optimize the sensing matrix by considering the large dimensionality of Kronecker product matrix. It is shown that the optimized sensing matrix with smaller mutual coherence with a given basis would achieve efficient sampling and fast convergence.

Recall a related definition of mutual coherence for the projection dictionary $D = \Phi \Psi$.

$$
\mu(D) = \max_{1 \leq i, j \leq N} \frac{|(d_i^T d_j)|}{\|d_i\| \cdot \|d_j\|},
$$

where $d_i$ denotes the $i$-th column of $D$. The mutual coherence measures the worst similarity between columns of $D$. Since pursuit techniques cannot distinguish two closely related columns, the mutual coherence plays an important role on the performance of recovery algorithms. Projection dictionary with small mutual coherence tends to achieve incoherent sampling and indicate a robust reconstruction, while the one with large coherence is more likely to fail. Suppose that the sparse representation for $x_0$ is in the form of $Dx_0$ with $\|x_0\|_0 \leq \frac{1}{2}(1 + \frac{1}{\mu(\Phi \Psi)})$. We have the following results for $\ell_0$ and $\ell_1$ optimization.

1) The vector $x_0$ is necessarily the sparsest one to describe $x_0$ by solving the $\ell_0$ optimization.

$$
\min_\alpha \|\alpha\|_0 \text{ s.t. } x_0 = \Phi \Psi \alpha
$$

2) The Basis Pursuit (BP) algorithm for approximating $x_0$ guarantees an exact solution to the linear programming problem.

$$
\min_\alpha \|\alpha\|_1 \text{ s.t. } x_0 = \Phi \Psi \alpha
$$

Thus, the effect of mutual coherence on the measurements can be evaluated. The two metrics imply that the upper bound of $\|\alpha\|_0$ will be relaxed when $\mu(\Phi \Psi)$ decreases. This relaxed upper bound allows a wider set of candidate signals to reside under the umbrella of successful CS. It benefits the reconstruction in two aspects. First, more accurate recovery can be obtained with the same number of measurements. Second, a fast convergence to the $\ell_1$ convex optimization can be achieved. Therefore, the optimization for sensing matrix can be transferred to minimization of the mutual coherence of projection dictionary $D$.

In practice, to ensure a fair comparison to actual behavior of sparse representations and pursuit algorithm’s performance, we replace the worst-case stand-point with an “average” measure of mutual coherence. To average the significant mutual coherence between columns of $D$, the Gram matrix $G = D^T D$ is considered for the normalized columns of $D$.

Definition 3: The $t$-average mutual coherence of $D$ is

$$
\mu_t(D) = \frac{\sum_{1 \leq i, j \leq k} |g_{ij}| \cdot |g_{ij}|}{\sum_{1 \leq i, j \leq k} ( |g_{ij}| \geq t )},
$$

where $g_{ij}$ is the off-diagonal entry of Gram matrix $G$ that represents the inner product between the normalized columns $d_i / \|d_i\|$ and $d_j / \|d_j\|$.

In [27], Elad developed an algorithm to minimize the $\mu_t(D)$ where off-diagonal entries in $G$ that are larger than a threshold $\gamma$ will be reduced with a shrunk constant and the rank of $G$ would be subsequentially retained. In practice, the value of $\mu_t(D)$ would be reduced iteratively. Obviously, its shortcoming lies in that some small off-diagonal entries in $G$ will be increased in the process of “artificial” rank retaining and will affect the RIP of $D$. To overcome it, a learning method was targeted to simultaneously optimize sensing matrix and sparsifying basis by finding $\Phi$ that makes the corresponding Gram matrix $G$ approximate the identity matrix [29].

$$
\psi^T \Phi^T \Phi \psi \approx I_m.
$$

In Fig. 2, we provide the histograms of the absolute values of off-diagonal entries in $G$ before and after the optimization. Fig. 2(b) shows that Elad’s algorithm [27] preserves the ordering of the absolute entries in $G$ and leads to better distribution, while it creates some artificial large values entries that are not present in the original matrix. These artificial values would completely ruin the worst-case guarantees of the pursuit algorithms. On the contrary, the learning based algorithm [29] improves the distribution of off-diagonal entries by orthogonalizing arbitrary subset of columns in $D$. It makes all the off-diagonal entries in $G$ shrink to zero and the diagonal entries approach one. Fig. 2(c) suggests that [29] provides the best mutual coherence and RIP of $D$, as the histogram decreases monotonically. Consequently, the mutual coherence of corresponding synthetic sensing matrix is optimized for the Kronecker product in the fully decomposable CS model.

B. Separable Optimization for Kronecker Product

The separability of Kronecker product enables us to design a specific algorithm to simplify both the design procedure and the implementation of the synthetic sensing matrix to reduce complexity. In this section, we demonstrate the optimality of the separable minimization for the mutual coherence in the Kronecker product framework. Thus, the optimized synthetic sensing matrix can be obtained by making Kronecker products of the ones optimized separately. Moreover, the optimization process is also separable, which preserves the block feature of
Kronecker product matrix and enables fast low-scale matrix computation.

The mutual coherence of the projection dictionary in fully decomposable CS is formulated as \( \mu(\Phi_\Psi) = \mu(\Phi_1 \Psi_1) \cdots \mu(\Phi_E \Psi_E) \) for the \( E \) subspaces. According to the theorem, we can obtain the optimal mutual coherence \( \mu(\Phi_\Psi) \) by calculating the product of \( E \) optimized mutual coherence \( \mu(\Phi_1 \Psi_1), \ldots, \mu(\Phi_E \Psi_E) \). For each subspace \( \mathcal{S}_i = \Phi_i \), \( \Phi_i \) is optimized for wavelet basis \( \Psi_i \). For arbitrary \( 1 \leq i, j \leq E, \mu(\Phi_i \Psi_j) \) and \( \mu(\Phi_j \Psi_i) \) are independent with each other. Thus, \( \Phi_i \) can be determined separately to minimize \( \mu(\Phi_i \Psi_i) \).

Algorithm 1 is developed to optimize matrix \( \Phi \) by separately minimizing mutual coherence for spatial and temporal sensing matrices based on a fixed sparsifying basis \( \Psi \). Taking the temporal sensing matrix as example, the optimization starts with a randomly generated Gaussian matrix \( \Phi_1 \in \mathbb{R}^{M \times N} \). The optimal sensing matrix \( \Phi_1 \) is obtained under the constraint that the Gram matrix \( G = \Phi_1^T \Phi_1 \) is close to an \( N \times N \) identity matrix. Rewriting (16) by multiplying both side of \( G \) with \( \Psi \) and \( \Psi^T \), we obtain

\[
\Psi \Psi^T \Phi_1^T \Phi_1 \Psi \Psi^T - \Psi \Psi^T \approx 0. \tag{17}
\]

Thus, the optimized sensing matrix \( \Phi_1 \) minimizes \( \| \Psi \Psi^T \Phi_1^T \Phi_1 \Psi \Psi^T - \Psi \Psi^T \|^2_F \) with the fixed basis \( \Psi \). Let us define the eigendecomposition of \( \Psi \Psi^T \) as \( \Psi \Psi^T = Q \Sigma Q^T \), where \( \Sigma = \text{diag}(\lambda_1, \ldots, \lambda_N) \) is the diagonal matrix whose entries are the eigenvalues of \( \Psi \Psi^T \) and \( Q \) is the orthogonal matrix composed of corresponding eigenvectors. Therefore, \( \| \Sigma Q^T \Phi_1^T \Phi_1 Q \Sigma - \Sigma \|^2_F \) can be considered as an equivalent substitution of (17). Substituting \( W \) with \( \Phi Q \), we formulate the equivalent minimization problem for the residual.

\[
\mathbf{W} = \arg \min_{\mathbf{W}} \| \Sigma - \Sigma \mathbf{W}^T \mathbf{W} \Sigma \|^2_F. \tag{18}
\]

Thus, \( \Phi_1 = \Phi Q \) minimizes (18).

Denote \( \mathbf{v}_i = [\lambda_1 \mathbf{w}_{i,1}, \ldots, \lambda_N \mathbf{w}_{i,N}] \) the \( i \)-th row of matrix \( \mathbf{W} \Sigma \). We define \( \mathbf{R} = \Sigma - \sum_{i=1}^M \mathbf{v}_i^T \mathbf{v}_i \) and \( \mathbf{R}_j = \Sigma - \sum_{i:j \neq j} \mathbf{v}_i^T \mathbf{v}_j \). In the optimization, \( \mathbf{v}_j \) is obtained by minimizing the residuals \( \| \mathbf{R}_j - \mathbf{v}_j^T \mathbf{v}_j \|^2_F \). Provided that the eigendecomposition of \( \mathbf{R}_j \) is \( \mathbf{R}_j = \mathbf{U}_j \mathbf{\Delta}_j \mathbf{U}_j^T \), \( \mathbf{v}_j \) is set to \( \sqrt{\delta_{1,j}} \mathbf{u}_{1,j} \) to eliminate the largest item of residuals. Here, \( \delta_{1,j} \) is the largest eigenvalue of \( \mathbf{\Delta}_j \) and \( \mathbf{u}_{1,j} \) is its corresponding eigenvector, \( \mathbf{u}_{1,j} \in \mathbf{U}_j^T \). Therefore, \( \mathbf{W} \) and \( \Phi_i \) can be solved from \( \{ \mathbf{v}_i \} \) and \( \{ \lambda_i \} \). Similarly, we can find the optimal spatial sensing matrix \( \Phi_s \), and consequently, generate the synthetic sensing matrix \( \Phi = \Phi_s \otimes \Phi_t \).

Algorithm 1: Separable Sensing Matrix Optimization.

1. Task: Find the optimal sensing matrix \( \Phi_1 \) that minimizes \( \mu(\Phi_1 \Psi_1) \).
2. Initialization: Sparsifying basis \( \Psi_1 \) and a randomly generated Gaussian matrix \( \Phi_1 \in \mathbb{R}^{M \times N} \).
3. Eigen-decomposition: \( \Psi_1 \Psi_1^T = Q \Sigma Q^T \).
4. Set \( \mathbf{W} = \Phi \mathbf{Q} \).
5. Set \( j = 1 \).
6. Repeat \( M \) times:

   7. Compute \( \mathbf{R}_1 = \mathbf{U}_1 \mathbf{\Delta}_1 \mathbf{U}_1^T \).
   8. Eigen-decomposition: \( \mathbf{R}_1 = \mathbf{U}_1 \mathbf{\Delta}_1 \mathbf{U}_1^T \).
   9. Sort the largest diagonal elements \( \delta_{1,j} \) and its eigenvector \( \mathbf{u}_{1,j} \).
   10. Update \( \mathbf{w}_j \) using \( [\lambda_1 \mathbf{w}_{j,1}, \ldots, \lambda_N \mathbf{w}_{j,N}] = \sqrt{\delta_{1,j}} \mathbf{u}_{1,j} \).
   11. Set \( j = j + 1 \).
   12. Update \( M \) components of \( \mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_M] \).
   13. Compute the optimal \( \Phi_1 = \Phi \mathbf{Q}^T \).

C. Sampling Rate-Distortion Optimization

Since compressive sensing is separably performed on both spatial and temporal dimension, sampling rate-distortion performance is optimized for fully decomposable CS. Considering that Algorithm 1 minimizes \( \| \Psi \Psi^T \Phi_1^T \Phi_1 \Psi \Psi^T - \Psi \Psi^T \|^2_F \) for the corresponding dimension, we introduce Lagrangian cost function \( \mathcal{L}(\Phi_s, \Phi_t) \) for optimized sampling rate allocation.

\[
\mathcal{L}(\Phi_1, \ldots, \Phi_E) = \sum_{i=1}^E \eta_i \| \Psi_i \Phi_i^T \Phi_i \Psi_i - I \|^2_F, \tag{19}
\]

where \( \eta_i \) is the Lagrangian multiplier for the \( i \)-th subspace \( \mathcal{S}_i \). Given the total sampling rate \( r \), we minimize \( \mathcal{L}(\Phi_1, \ldots, \Phi_E) \) based on their partial sampling rates \( r_1, \ldots, r_E \).

**Theorem 3:** Given arbitrary smooth wavelet bases \( \Psi_i \) for the \( i \)-th subspace \( \mathcal{S}_i \), with a high probability, there exists optimal sampling rate allocation \( r_1, \ldots, r_E \), \( 1 \leq i \leq E \) to minimize the upper bound of recovery error.

**Proof:** Please refer to Appendix C.

Theorem 3 implies that there exists an optimal sampling rate allocation \( r_1, \ldots, r_E \) for the multidimensional signals. It shows that the sampling rate allocation is related to the spatial and temporal sparsifying basis, which affects the sparsity \( K_i \), \( 1 \leq i \leq E \) of multidimensional signals.

In practice, the heuristic optimization would be prohibitive to determine sampling rate. To balance the efficiency and recovery performance, a set of \( L \) candidate pairs of spatial and temporal sampling rates \( \{(r_1^1, \ldots, r_E^1), \ldots, (r_1^L, \ldots, r_E^L)\} \) are predetermined for selection. For each pair \( (r_1^j, \ldots, r_E^j) \), the Lagrangian cost function is evaluated for comparison, so that the optimal synthetic sensing matrix can be derived based on the optimized allocation scheme. In Section V, we elaborate the sampling rate allocation scheme for compressive video sampling.

V. APPLICATION INTO VIDEO ACQUISITION

In this section, we employ the proposed FDCS into compressive video sampling. Consider that a video sequence \( \mathcal{F} \) consists of \( T \) consecutive frames \( \mathbf{F}_1, \ldots, \mathbf{F}_T \) with a same resolution of \( N = N_R \times N_C \). Let us denote \( \mathbf{X} \) the \( N \times T \) matrix for the video sequence.

\[
\mathbf{X} = (\text{vec}(\mathbf{F}_1) \quad \text{vec}(\mathbf{F}_2) \cdots \text{vec}(\mathbf{F}_T)) \in \mathbb{R}^{N \times T}. \tag{20}
\]
Here, the $i$-th column vector $\text{vec}(F_i)$ of $X$ is the 1-D vectorization of $N_R \times N_C$ frame $F_i$. In progressive sampling for video acquisition [24], the sensing matrix for $X$ is a block diagonal matrix generated by Kronecker product $\Phi = \Phi_s \otimes I$, where the $M \times N$ diagonal sub-block $\Phi_s$ enables sampling of vectorized frames $\text{vec}(F_1), \ldots, \text{vec}(F_F)$. 

$$
\text{vec}(Y) = \begin{bmatrix}
\Phi_s & 0 & \ldots & 0 \\
0 & \Phi_s & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \Phi_s
\end{bmatrix} \cdot \text{vec}(X), \quad (21)
$$

where $Y = (y_1, y_2, \ldots, y_T)$ is the $M \times T$ matrix with $y_i$ indicating the $M \times 1$ measurements for $F_i$. According to (21), KCS utilizes the identity matrix $I$ to combine measurements from each frame, so that the block diagonal matrix $\Phi$ samples each frame $F_i$ into corresponding measurements $y_i$ with $\Phi_s$. By contrast, fully decomposable CS considers the sparsifying basis defined in (4) to generate $Y$ by jointly optimizing over $\Phi$ and $\Psi$ for spatial and temporal dimension.

FDCS simultaneously performs compressive sampling in spatial and temporal dimension for efficient representation of video sequences. In comparison to KCS, temporal sensing matrix $\Phi_t$ is adopted for structured sparsity along the motion trajectory. Sampling rate allocation scheme is developed to assign sampling rates to spatial and temporal dimensions. Algorithm 2 is adopted to adaptively estimate spatial and temporal sensing matrices and the cost is derived based on the estimated synthetic sensing matrix. Sensing matrix for each dimension is optimized according to Algorithm 1.

Remarkably, the block-diagonal structure is retained in the product $\Phi_s X \Phi_t^T$, which means that the fully decomposable CS model satisfies the progressive sampling for video acquisition. Furthermore, fully decomposable CS can also obtain measurements separately in spatial and temporal domains with existing hardware implementations of CS imaging systems [45], [46] with separable sensing devices. According to (8), the temporal sensing can be implemented similarly to the spatial sensing. Once the 2-D spatial frames are compressively sensed and temporally stored, the DMD (digital micromirror device) is used to linearly combine spatial measurements at corresponding positions like a second-order CS. In Fig. 3, FDCS and KCS are evaluated over various video sequences. The measurements obtained by KCS are redundant on the temporal dimension (horizontally), as they are similar in the same row.

In Fig. 4(a) and 4(b), we provide examples for sampling Akiyo and Foreman sequences with various allocation of spatial and temporal sampling rates. Fig. 4(a) shows the sampling-rate-distortion performance for Akiyo sequence with a resolution of $128 \times 128 \times 128$ with the overall sampling rate $r$ varying in the interval $[0.20, 0.40]$. The temporal sampling rate $r_t$ is fixed at 0.6. It shows that FDCS outperforms KCS and independent CS by a gain of 3–8 dB in PSNR. In Fig. 4(b), we fix the overall sampling rate and compare the recovery performance of the proposed FDCS model with various temporal sampling rates. The Foreman sequence features scene movement, which is reflected in sharp changes in the value of each pixel across frames, so that it requires more temporal measurements. The best recovery is achieved with $r_t = 0.75$.

VI. EXPERIMENTAL RESULTS

To evaluate the proposed fully decomposable CS (FDCS) model in video acquisition, we compare it with the state-of-the-art CS schemes in terms of recovery performance and analyze the influence of sensing matrix optimization. We also discuss the optimal configuration of rate allocation for various sequences.
The Daubechies-8 wavelet was used to form sparsifying basis, and appropriate random matrices were set as the initial temporal and spatial sensing matrices. Referring to the matrix structure, we crop the standard video sequence around the center to have frames of size $128 \times 128$ pixels (approximate QCIF resolution) and $256 \times 256$ pixels (approximate CIF resolution). To appropriately utilize the temporal sparsity within hardware capacity, 128 frames are selected to form a test sequence. In the experiments, the basis pursuit (BP) solver is used to recover the original video signals. All the experiments are executed on a workstation with 3.3-GHz CPU and 12-GB RAM. It is noted that we develop the proposed algorithm over the MATLAB toolbox at http://dsp.rice.edu/kcs.

A. Recovery Performance

Firstly, we evaluate various CS-based sampling schemes on the cropped video cubes, including the independent CS recovery [9], the GTCS [26], the KCS [24] and the proposed fully decomposable CS (FDCS) model. For the independent recovery, CS independently obtains and recovers measurements for each individual frame with the sparsifying basis $\Psi_s$. The 3D-DCT basis is used for GTCS. For KCS and FDCS, the same Kronecker product (hyperbolic) wavelet basis is adopted. It is worth mentioning that KCS only utilizes a spatial-compressed synthetics matrix to reconstruct signals from the entire measurements, while the FDCS takes advantage of optimized holo-compressed sensing matrix $\hat{\Phi} = \hat{\Phi}_s \otimes \hat{\Phi}_t$ to sufficiently exploit the spatio-temporal correlations in video sequences. The total sampling rate (ratio of measurements and pixels) is set as $r = 0.35$. For the FDCS model, the temporal sampling rate $r_t$ is selected as 0.75 to maximally envelop the temporal structures of all test sequences. Figs. 5 and 6 show the reconstructed frames for independent CS, GTCS, KCS, and the proposed FDCS model over test sequences with various resolutions, including Foreman, Akiyo, Mother and Daughter, and Hall. In comparison to KCS, GTCS and CS, the proposed FDCS model achieves best visual quality for recovered frames over all the test sequences. Its improvement on recovery owes to the availability of extra temporal compression, which tends to preserve more details with the same number of measurements.

B. Sensing Matrix Optimization

This section validates the optimized sensing matrix against the randomly generated matrix in FDCS. The sensing matrix optimization algorithm is evaluated in terms of the PSNR of reconstructed frames and time cost. Gaussian random matrix $\Phi$ and optimized sensing matrix $\hat{\Phi}$ are adopted to collect fully decomposable CS measurements under various sampling rates, respectively. Here, the optimized matrix $\hat{\Phi}$ is calculated from $\Phi$ with Algorithm 1. The mutual coherence of the random and optimized matrices is 0.092 and 0.055, respectively. Table II gives the detailed recovery quality and computational complexity of FDCS with both the optimized and the randomly generated sensing matrix. Remarkably, the optimized matrix achieves better recovery performance as well as reduces about 80% of the computational complexity in comparison to the random matrix. This fact coincides with the conclusions in Section IV.

C. Sampling Rate Allocation

Finally, we discuss the sampling rate allocation for the proposed FDCS model. Instead of full sampling along the temporal dimension in KCS, FDCS sufficiently exploits the temporal correlations to further compress the temporal components and achieve higher sampling efficiency. Therefore, it is reasonable to consider rate allocation for the spatial and temporal sampling to obtain optimal overall performance. As mentioned in Section II, the hyperbolic basis outperforms any other basis for its universality in various dimensions, which means that we can also construct appropriate sensing matrix according to the sparsity along both spatial and temporal dimensions.
Table II

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Model</th>
<th>$r=0.20$</th>
<th>$r=0.25$</th>
<th>$r=0.30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR (dB)</td>
<td>Time (S)</td>
<td>PSNR (dB)</td>
<td>Time (S)</td>
</tr>
<tr>
<td>Foreman</td>
<td>Random</td>
<td>25.82</td>
<td>2.9e+04</td>
<td>27.19</td>
</tr>
<tr>
<td></td>
<td>Optimized</td>
<td>25.88</td>
<td>5.5e+03</td>
<td>27.22</td>
</tr>
<tr>
<td>Akiyo</td>
<td>Random</td>
<td>26.60</td>
<td>3.1e+04</td>
<td>28.82</td>
</tr>
<tr>
<td></td>
<td>Optimized</td>
<td>26.65</td>
<td>6.1e+03</td>
<td>28.95</td>
</tr>
<tr>
<td>Hall</td>
<td>Random</td>
<td>22.95</td>
<td>3.1e+04</td>
<td>24.57</td>
</tr>
<tr>
<td></td>
<td>Optimized</td>
<td>22.97</td>
<td>5.2e+03</td>
<td>24.59</td>
</tr>
<tr>
<td>MD</td>
<td>Random</td>
<td>26.56</td>
<td>3.9e+04</td>
<td>30.31</td>
</tr>
<tr>
<td></td>
<td>Optimized</td>
<td>33.38</td>
<td>9.8e+03</td>
<td>35.93</td>
</tr>
</tbody>
</table>

Intuitively, the number of measurements would determine the quality of reconstruction in CS. However, we can notice in Fig. 7(d) and (f) that lower sampling rate could help achieve better reconstruction. To achieve the best performance in FDCS, the allocation of sampling rates for different dimensions should follow the distribution of sparsity, as shown in Figs. 7–10. For example, more measurements are required for the temporal dimension, as there exist sharp changes across frames in Foreman sequence. In Fig. 7, the reconstruction quality in (c)(g) with larger $r_T$ is better than (b)(f) with the same overall sampling rate $r$.

Fig. 7. Recovery performance for fully decomposable CS (FDCS) under various sampling rates. The overall sampling rates for the first and second rows are 0.25 and 0.45, respectively. (a): Original; (e): KCS; (b)(f): FDCS with $r_T = 0.45$; (c)(g): FDCS with $r_T = 0.6$; (d)(h): FDCS with $r_T = 0.75$.

Fig. 8. Recovery performance for fully decomposable CS (FDCS) under various sampling rates. The overall sampling rates for the first and second rows are 0.25 and 0.45, respectively. (a): Original; (e): KCS; (b)(f): FDCS with $r_T = 0.45$; (c)(g): FDCS with $r_T = 0.6$; (d)(h): FDCS with $r_T = 0.75$.

Fig. 9. Recovery performance for fully decomposable CS (FDCS) under various sampling rates. The overall sampling rates for the first and second rows are 0.25 and 0.45, respectively. (a): Original; (e): KCS; (b)(f): FDCS with $r_T = 0.45$; (c)(g): FDCS with $r_T = 0.6$; (d)(h): FDCS with $r_T = 0.75$.

Fig. 10. Recovery performance for fully decomposable CS (FDCS) under various sampling rates. The overall sampling rates for the first and second rows are 0.25 and 0.45, respectively. (a): Original; (e): KCS; (b)(f): FDCS with $r_T = 0.45$; (c)(g): FDCS with $r_T = 0.6$; (d)(h): FDCS with $r_T = 0.75$.
$r_t$ can also guarantee an accurate reconstruction, as shown in Fig. 8.

Figs. 11 and 12 provide the sampling rate-distortion curves obtained by FDCS, GTCS, KCS, and the independent CS for test sequences with various resolutions. These curves are obtained with total sampling rates ranging from 0.2 to 0.45. For fully decomposable CS and GTS, the temporal sampling rates are set to 0.45, 0.6 and 0.75, respectively. To sufficiently exploit temporal correlations, sensing matrix optimization is also employed into fully decomposable CS. Figs. 11 and 12 show that FDCS outperforms the state-of-the-art methods, especially in the regions of low sampling rates ($r < 0.3$). Remarkably, FDCS performs better under the situations that temporal sampling is sufficient. In comparison to the Kronecker product-based method KCS, FDCS achieves a noticeable gain (up to 8 dB) in PSNR. Fig. 13 provides the sampling-rate-distortion curves for test sequences Akiyo, Foreman, Hall, and Mother and Daughter under overall sampling rate 0.4.

VII. CONCLUSION

This paper proposes a novel fully decomposable CS (FDCS) for highly compressed multidimensional signals. With the structured sparsity, FDCS preserves a higher sampling efficiency by further compressing projected signals lived in multiple subspaces. Due to the block feature of Kronecker product, sampling rate allocation is developed to derive the overall sensing matrix for optimized reconstruction performance over the multiple subspaces. With the knowledge that projection matrix with low mutual coherence would lead to lower reconstruction errors, mutual coherence for each subspace is minimized by optimizing its sensing matrix based on the projection dictionary of FDCS. The mutual coherence is proved to be divisible, which makes it possible for obtaining the optimal synthetic sensing matrix from each dimension in a fast low-scale matrix computation. For validation, the proposed FDCS is employed in compressive video sampling. Experiments results show that the proposed scheme substantially improves the reconstruction accuracy and further reduces the necessary number of samples.

APPENDIX A

PROOF OF THEOREM 1

We begin with $E = 2$. In the Kronecker product framework, the mutual coherence for synthetic sensing matrix $\Phi_s \otimes \Psi_t$ and the sparsifying basis $\Psi_s \otimes \Psi_t$ can be decomposed by

$$\mu(\Phi_s \otimes \Phi_t, \Psi_s \otimes \Psi_t) = \mu(\Phi_s, \Psi_s) \cdot \mu(\Phi_t, \Psi_t).$$

(22)

Thus, we consider the mutual coherence for spatial and temporal dimension, respectively.

Without loss of generality, we assume that the corresponding spatial sensing matrices $\Phi_s^{(F)}$ and $\Phi_s^{(K)}$ for the fully decomposable CS and KCS models are the selected sub-matrices of $N \times N$ orthonormal bases $\Phi_s$. Given $M \times N$ and $M_s \times N$ spatial sensing matrix $\Phi_s^{(K)}$ and $\Phi_s^{(F)}$, $\mu(\Phi_s^{(K)}, \Psi_s^{(K)})$ and $\mu(\Phi_s^{(F)}, \Psi_s^{(F)})$ for the fully decomposable CS and KCS model satisfy with probability at least $1 - 5e^{-\tau}$ that [24], [31], [47]

$$\frac{\mu(\Phi_s^{(F)}, \Psi_s)}{\mu(\Phi_s^{(K)}, \Psi_s)} \leq \sqrt{\frac{r \cdot N(N - 1)}{1 - r}} \cdot r_s C K \tau \log (\tau K \log N) \times (\log K)^2.$$

Here, $C$ is a fixed constant. For simplicity, we denote

$$\Delta(K, \tau, N) = \sqrt{N(N - 1)} \cdot K \tau \log (\tau K \log N) (\log K)^2.$$

Similarly, we assume that fully decomposable CS select $M_t$ column vectors from the $T \times T$ orthonormal bases $\Phi_t$ to
generate the temporal sensing matrix $\Phi_i^{(F)}$. Thus, the temporal mutual coherence for fully decomposable CS is derived as
\[
\mu(\tilde{\Phi}_i, \Psi_i) = \max_{1 \leq i, j \leq T} \left| \frac{\langle \tilde{\phi}_i, \psi_j \rangle}{\mu} \right| = \max_{1 \leq i, j \leq T} \left| \sum_{1 \leq k \leq T} \tilde{\phi}_i(i, k) \psi_i(k, j) \right|.
\]
Given $\Psi_i$, we denote $\tilde{\mu} = \max_{1 \leq i, j \leq T} |\psi_i(i, j)|$ the mutual coherence $\mu(I, \Psi_i)$ for KCS. When the elements of $\tilde{\Phi}_i$ obey a normal distribution having mean zero and variance $1/T$, we can find that $\langle \tilde{\phi}_i, \psi_i \rangle \sim \mathcal{N}(0, 1/T)$, as $\psi_i$ is an orthonormal vector. According to [48], we can obtain that
\[
Pr \left\{ \left| \langle \tilde{\phi}_i, \psi_i \rangle \right| \leq \sqrt{\frac{1-r}{r_s}} \cdot \frac{\tilde{\mu}}{\mu(K, \tau, N)} \right\} 
\geq 1 - 2 \exp \left( - \frac{(T-2)(1-r)\tilde{\mu}^2}{2rt_e^2\Delta^2(K, \tau, N)} \right).
\]
As a result, when $T$ grows, fully decomposable CS would yield a smaller mutual coherence in comparison to KCS with a high probability asymptotically approaching $1$.

Subsequently, we extend this result to the cases with $E \geq 2$. Considering that KCS adopts identity sampling matrix $I$ for each dimension with $i = 2, \ldots, E$. We can obtain similar results to $E = 2$. Supposing that fully decomposable CS and KCS leverage sampling matrices $\Phi_i^{(F)}$ and $\Phi_i^{(K)}$ selected from the $N \times N$ orthonormal bases $\Phi_i$, we can find with probability at least $1 - 5e^{-7}$ that
\[
\frac{\mu(\Phi_i^{(F)}, \Psi_i)}{\mu(\Phi_i^{(K)}, \Psi_i)} \leq \sqrt{\frac{\tau N(N-1)}{1-r}} \cdot r_i CK \tau \log (\tau K \log N) \times (\log K)^2,
\]
where $r_i$ is the sampling rate for the $i$-th dimension and the overall sampling rate $r = \prod_{i=1}^E r_i$. Without loss of generality, we suppose the elements of $T_i \times T_i$ orthonormal bases $\tilde{\Phi}_i$ for the $i$-th dimension, $i = 2, \ldots, E$, obey a normal distribution with zero mean and variance $1/T_i$. Thus, we can obtain that
\[
Pr \left\{ \left| \langle \tilde{\phi}_i, \psi_i \rangle \right| \leq \sqrt{\frac{1-r}{r}} \cdot \frac{\tilde{\mu}}{r_i \Delta(K, \tau, N)} \right\} 
\geq 1 - 2 \exp \left( - \frac{(T_i-2)(1-r)\tilde{\mu}^2}{2rt_e^2\Delta^2(K, \tau, N)} \right).
\]
Therefore, provided arbitrary $E \in \mathbb{N}$, with the growth of $T_i, i = 2, \ldots, E$, fully decomposable CS would yield a smaller mutual coherence in comparison to KCS with a high probability asymptotically approaching $1$.

**APPENDIX B**

**PROOF OF THEOREM 2**

Considering $D_i = \Phi_i \Psi_i$ for the $i$-th subspace $S_i$, we rewrite the Kronecker product for the multidimensional signals $X$ of $E$ dimensions.
\[
\text{vec}(Y) = (D_1 \otimes \cdots \otimes D_E) \text{vec}(X).
\]
Since $(A \otimes B) \cdot \text{vec}(X) = \text{vec}(Y)$ is equivalent to $AXB^T = Y$, $Y$ can be obtained by multiplying $X$ with $D_1, \ldots, D_E$ in an arbitrary order. Without loss of generality, the sequential order is $1, \ldots, E$ for proof. For the $i$-th subspace $S_i$, we find for arbitrary $K_i$-sparse vector $u_i$ derived from the $\ell_2,1$ optimization that $C_i \|u_i\|_2^2 \leq \|D_i u_i\|_2^2 \leq C_i \|u_i\|_2^2$. For the $E$ subspaces,
\[
\prod_{i=1}^E C_i \|u_i\|_2^2 \leq \prod_{i=1}^E \|D_i u_i\|_2^2 \leq \prod_{i=1}^E C_i \|u_i\|_2^2.
\]
Thus, there exists $u = u_1 \otimes \cdots \otimes u_E$ satisfying Equation (11).

**APPENDIX C**

**PROOF OF THEOREM 3**

When optimizing the Lagrangian cost function for sampling rate, the upper bound for recover error would be minimized. Given smooth wavelet bases $\Psi_i$, since sensing matrix optimization is separately performed for the $E$ dimension, we can obtain for the $i$-th dimension with $i = 1, \ldots, E$ [24]
\[
\| \hat{x}^{(i)} - x^{(i)} \|_2 \leq C_i \left( \frac{M_i}{\sqrt{N_i} \cdot \mu(\Phi_i, \Psi_i)} \right)^{-\left( \frac{1}{\sqrt{r_i}} + \frac{1}{r_i} \right)}.
\]
Here, we assume that the multidimensional signal $X$ is $K_i$-sparse for the $i$-th subspace $S_i$ in the fully decomposable CS model. Considering that $\|X - \hat{X}\|_2 = \otimes_{i=1}^E \|x^{(i)} - \hat{x}^{(i)}\|_2$, we have
\[
\|X - \hat{X}\|_2 \leq \prod_{i=1}^E C_i \left( \frac{M_i}{\sqrt{N_i} \cdot \mu(\Phi_i, \Psi_i)} \right)^{-\left( \frac{1}{\sqrt{r_i}} + \frac{1}{r_i} \right)}.
\]
Since the sparsity $K = \max(k_1, \ldots, k_E)$, we can obtain for constant $C = \prod_{i=1}^E C_i$,
\[
\|X - \hat{X}\|_2 \leq C \left( \frac{M}{\sqrt{N} \prod_{i=1}^E \mu(\Phi_i, \Psi_i)} \right)^{-\left( \frac{1}{\sqrt{r_i}} + \frac{1}{r_i} \right)}.
\]
Considering mutual coherence for $S_i$, we can find that
\[
\mu(\Phi_i, \Psi_i) \geq \sqrt{\frac{N_i - M_i}{M_i(N_i - 1)}} = \sqrt{\frac{1 - r_i}{r_i(N_i - 1)}}.
\]
Therefore, we can obtain that
\[
\|X - \hat{X}\|_2 \leq C \left( \frac{N \prod_{i=1}^E r_i}{\sqrt{N} \prod_{i=1}^E \mu(\Phi_i, \Psi_i)} \right)^{-\left( \frac{1}{\sqrt{r_i}} + \frac{1}{r_i} \right)} \left( \sum_{i=1}^E \frac{1}{N_i} \right)^2.
\]
When fixing the overall sampling rate by $r = \prod_{i=1}^E r_i$, there exists $r_i, i = 1, \ldots, E$ to minimize the recovery error.

**REFERENCES**


JunnZ (M’07) received the M.S. and the Ph.D. degrees in communication and information system from Shanghai University, Shanghai, China, in 2004 and 2006, respectively.

She is currently a Full Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China. From 2006 to 2016, she was with the School of Communication and Information Engineering, Shanghai University, and became a Full Professor in 2014. From June 2011 to June 2012, she was with the Department of Electrical and Computer Engineering, University of California, San Diego, as a Visiting Professor. She has authored/co-authored more than 80 IEEE journal/conference papers, and 2 book chapters, including 16 IEEE Transactions journal papers. She holds 12 patents and has 10+ under-reviewing patents. She has served on some technical program committees for the IEEE and other international conferences.

Her research interests include multimedia communication, network resource optimization, wireless communication, and network information theory.

Dr. Zou was granted National Science Fund for Outstanding Young Scholar in 2011. She was a recipient of Shanghai Yong Rising Star Scientist Award in 2008, the First Prize of the Shanghai Science and Technology Advance- ment Award in 2011, and the First Prize of the Shanghai Science and Technology Advance- ment Award in 2008.

Hongkai Xiong (M’01–SM’10) received the Ph.D. degree in communication and information system from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 2003.

Since 2003, he has been with the Department of Electronic Engineering, SJTU, where he is currently a full Professor. From December 2007 to December 2008, he was a Research Scholar with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. From 2011 to 2012, he was a Scientist with the Division of Biomedical Informatics, University of California, San Diego, San Diego, CA, USA. Since 2012, he has been a member of Innovative Research Groups, National Natural Science. His research interests include source coding/network information theory, signal processing, computer vision, and machine learning. He has published more than 190 refereed journal/conference papers.

Dr. Xiong was the recipient of the Top 10% Paper Award at the 2016 IEEE Visual Communication and Image Processing, the Best Student Paper Award at the 2014 IEEE Visual Communication and Image Processing, the Best Paper Award at the 2013 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, and the Top 10% Paper Award at the 2011 IEEE International Workshop on Multimedia Signal Processing. In 2016, he was granted Yangtze River Scholar Distinguished Professor from the Ministry of Education of China, and the Youth Science and Technology Innovation Leader from the Ministry of Science and Technology of China. He was also awarded Shanghai Academic Research Leader. In 2014, he was granted National Science Fund for Distinguished Young Scholar and Shanghai Youth Science and Technology Talent as well. In 2013, he was awarded Shanghai Shu Guang Scholar. In 2011, he was the recipient of the First Prize of the Shanghai Technological Innovation Award for Network-Oriented Video Processing and Dissemination: Theory and Technology. In 2010 and 2013, he was the recipient of the SMC-A Excellent Young Faculty Award of Shanghai Jiao Tong University. In 2009, he was awarded New Century Excellent Talents in University, Ministry of Education of China. He served as a TPC member for prestigious conferences, such as ACM Multimedia, ICIP, ICME, and ISCAS.