COMPRESSIVE VIDEO SAMPLING FROM A UNION OF DATA-DRIVEN SUBSPACES

Yong Li, Hongkai Xiong, Xinwei Ye*

Department of Electronic Engineering
Shanghai Jiao Tong University
Shanghai 200240, P.R. China

ABSTRACT

Recently, compressive sampling (CS) is an active research field of signal processing. To further decrease the necessary measurements and get more efficient recovery of a signal x, recent approaches assume that x lives in a union of subspaces (UoS). Unlike previous approaches, this paper proposes a novel method to sample and recover an unknown signal from a union of data-driven subspaces (UoDS). Instead of a fix set of supports, this UoDS is learned from classified signal series which are uniquely formed by block matching. The basis of these data-driven subspaces is regularized after dimensionality reduction by principal component extraction. A corresponding recovery solution with provable performance guarantees is also given, which takes full advantage of block-sparsity structure and improves the recovery efficiency. In practice, the proposed scheme is fulfilled to sample and recover frames in video sequences. The experimental results demonstrate that the proposed video sampling behaves better performance in sampling and recovery than the classical CS.

Index Terms— Compressive sampling, block matching, PCA, union of subspaces, data-driven, video compression.

1. INTRODUCTION

Compressive sampling (CS) is a prevailing theory in a variety of fields [1]-[3]. It attempts to acquire an unknown sparse signal by randomly projecting the original signal into its measurements whose dimension is much smaller. Let us consider that x is the unknown k-sparse signal, y is the measurements, and Φ is the sensing matrix. It is demonstrated that x can be reconstructed exactly from $m = O(k \log \frac{n}{k})$ random measurements when Φ satisfies certain conditions by solving ℓ_1-norm minimization problem [3]:

$$\min_x \|x\|_1, \text{ subject to } \Phi x = y. \tag{1}$$

where $y \in \mathbb{R}^{m\times 1}$, $\Phi \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^{n\times 1}$, $m \ll n$, $m > k$.

Recently, CS is applied to digital video sampling, compression and recovery [4]-[8]. In [4] and [5], each frame is decomposed into non-overlapped blocks and each block is approximated by a linear combination of blocks in previous frames at the CS decoder side. In [6], a block-based adaptive framework for CS is applied to maximum frame rate video acquisition at the encoder side. In [7], the pseudo-random down-sampling is performed in the 2-D Fourier transformation domain of digital video frames at the encoder side. In [8], a block-based CS framework is proposed where Karhunen-Loève basis is used to recover blocks at the decoder side.

It is worth mentioning that these methods converge on the situation where the signal x lies in a given single subspace. Besides these methods are complicated. To further decrease the necessary measurements and get more efficient recovery of a signal x, some researchers assume that x lives in a union of subspaces [9]-[11]. However, the current work mainly focuses on the theoretical investigation that what kind of sampling operators could be utilized to recover the signals.

In this paper, an explicit sampling strategy on a union of subspaces is provided for digital video sampling and recovery in practice. The underlying union of subspaces is constructed which makes full use of neighboring structures and enhances the sparsity. To be concrete, similar spatial fragments (e.g., blocks) are grouped and vectorized into a data array which is called "group". The 3-D group is changed into 2-D signal spaces which can be considered as a union of data-driven subspaces (UoDS). Noticeably, the UoDS is learned from classified signal series which are uniquely formed by block matching. Further, it is optimized by principal component analysis (PCA), which derives a basis matrix and enhances the sparsity. Besides the proof of stable reconstruction, we apply the proposed strategy to sampling and recovering video sequences where the UoDS is learned from decoded key frames. As a result, the non-overlapped blocks of non-key frames can be sampled and recovered stably and efficiently.

The remainder of the paper is organized as follows. The preliminary knowledge is discussed in Section 2. Section 3 describes the proposed algorithm of constructing a union of data-driven subspaces and demonstrates its stability. Moreover, it is witnessed to sample and recover video signal. The experimental results in Section 4 are evaluated.

2. PRELIMINARIES

Traditional CS considers the problem in form as

$$y = \Phi x \tag{2}$$

where Φ is the sensing matrix. Instead of a single subspace, x is considered to lie in a union of subspaces $\mathcal{U}$ [9]:

$$x \in \mathcal{U} \doteq \bigcup_{\lambda\in\Lambda} \mathcal{S}_\lambda \tag{3}$$

where $\mathcal{S}_\lambda$ are subspaces of Hilbert space $\mathcal{H}$ and Λ is a list of indices. $\mathcal{S}_\lambda$ are based on a fix basis (e.g., a Fourier or wavelet

*The work has been partially supported by the NSFC grants No. U1201255, No. 61271218, No. 61228101.

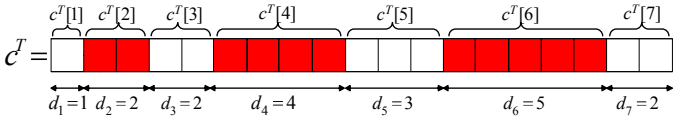

Fig. 1. k-block-sparse vector over $\mathcal{I} = \{d_1, d_2, \ldots, d_7\}$ with $k = 3$

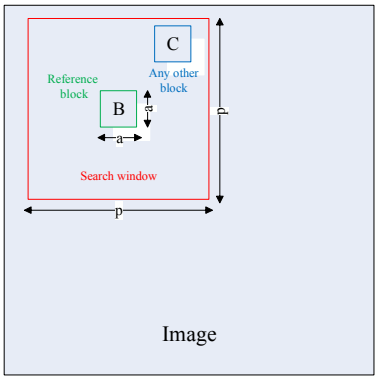

Fig. 2. The reference block, search window in block-matching for a union of data-driven subspaces

basis). Theoretically, $x \in \mathbb{R}^{n\times 1}$ is supposed to have a sparse representation with a given basis Ψ:

$$x = \Psi c \tag{4}$$

where c is a K-sparse vector, $\Psi = [\psi_1, \psi_2, \ldots, \psi_n]$ is an invertible matrix, and ψ_i is a transformation corresponding to a basis for subspace $\mathcal{S}_i$. In this sense, Eq. (2) can be written explicitly with vector c.

$$y_{m\times 1} = \Phi_{m\times n} x_{n\times 1} = \Phi_{m\times n}\Psi_{n\times n} c_{n\times 1} = Ac \tag{5}$$

where $A_{m\times n} = \Phi\Psi$. When taking a block-sparse structure in a union of subspaces into consideration [10], if t is the number of subspaces, $c^T = [c[1]^T \cdots c[t]^T]$ is called k-block-sparse if at most k blocks $c[i]_{d_i\times 1}$ are non-zero, with $n = \sum_{i=1}^{t} d_i$. Fig. 1 shows an example of a k-block-sparse vector with $k = 3$ and $t = 7$.

To guarantee an invertible and stable recovery, some conditions on the sampling operator Φ should be imposed. In this case, the block Restricted Isometric Property (Block-RIP) imposed on A would be defined as:

Definition 1 (*Block-RIP*) *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a given matrix, then A has the block RIP over $\mathcal{I} = \{d_1, d_2, \ldots, d_t\}$ with parameter $\delta_{A,k}$ if for every $c \in \mathbb{R}^n$ that is k-block-sparse over $\mathcal{I}$ such that*

$$(1 - \delta_{A,k})\|c\|_2^2 \le \|Ac\|_2^2 \le (1 + \delta_{A,k})\|c\|_2^2 \tag{6}$$

It was proved that if the sensing matrix A satisfies the block RIP, a convex algorithm which is based on minimizing a mixed ℓ_2/ℓ_1 norm will recover the block sparse vector c [10]. The mixed ℓ_2/ℓ_1 norm is defined as

$$\|c\|_{2,\mathcal{I}} = \sum_{i=1}^{t} \|c[i]\|_2 \tag{7}$$

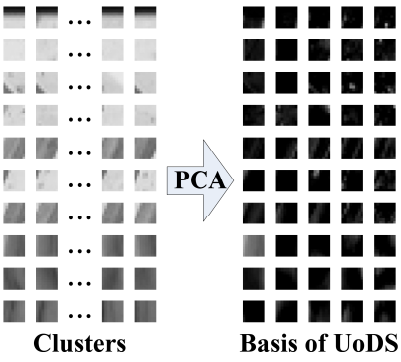

Fig. 3. A union of 10 data-driven subspaces and corresponding basis with 5 extracted features for each subspaces.

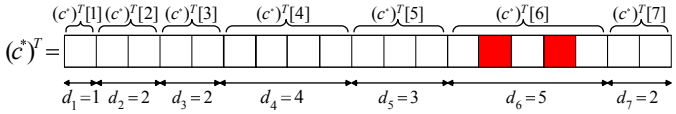

Fig. 4. k-block-sparse vector over $\mathcal{I} = \{d_1, d_2, \ldots, d_7\}$ with $k = 1$

The recovered k-block-sparse vector c can be derived by solving the block-sparse basis pursuit problem

$$\min_c \|c\|_{2,\mathcal{I}} \quad \text{subject to } y = Ac \tag{8}$$

where $\|c[i]\|_2 \ge 0, 1 \le i \le t$.

3. THE PROPOSED SCHEME

3.1. Union of Data-driven Subspaces (UoDS) by Block Matching

As shown in Fig. 2, each block x of size $a \times a$ in an image is clustered by block-matching [12]. The size of a reference block is $a \times a$, and B is denoted as the center of a search window whose size is $p \times p$. The set of similar blocks C in the search window is defined as

$$\mathcal{P}(B) = \{C : d(B, C) \le \tau\} \tag{9}$$

where τ is a distance threshold for d, $d(B,C) = \frac{\|\gamma(B)-\gamma(C)\|_2^2}{k^2}$ is the normalized quadratic distance between blocks, γ is a hard thresholding operator with threshold $\lambda\sigma$, and σ^2 is the variance of zero-mean Gaussian noise. First of all, two dimensional discrete Fourier transform (2-D DFT) is performed on x. Specially, the search window is set as the whole image. By computing the distance between the reference block x_B and the block x_C which is not clustered in the frequency domain, an image is decomposed into t groups (clusters) of blocks according to Eq. (9). All the 2-D blocks are vectorized into vectors whose dimensions are $a^2 \times 1$, and each vectorized group $X_i, i = 1, \ldots, t$, corresponds to a data-driven subspace $\mathcal{S}_i^*$ containing a vectorized block series. Therefore, the set of these matrices $[X_1, \ldots, X_t]$ corresponds to a UoDS denoted as $\mathcal{U}^*$. Apparently, it is dependent on the

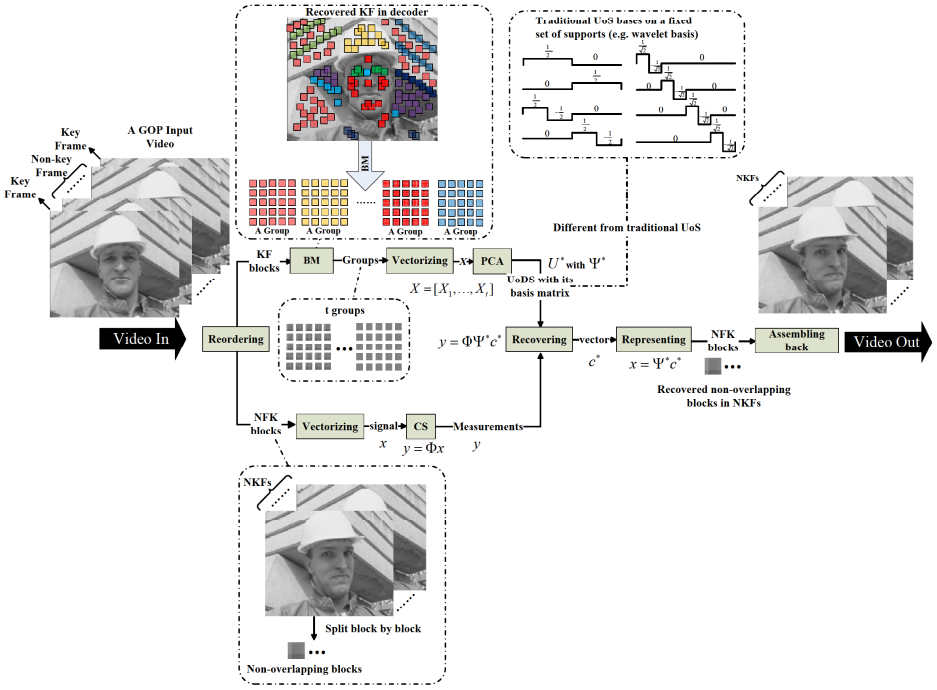

Fig. 5. The proposed compressive video sampling and reconstruction framework from a union of data-driven subspaces

signal source to form the UoDS which makes CS on UoDS more adaptive than CS on UoS. Similar to Eq. (3), each vectorized block x belongs to $\mathcal{S}_i^*$. It also lives in the union of subspaces as

$$x \in \mathcal{U}^* \doteq \bigcup \mathcal{S}_i^* \tag{10}$$

$[X_1, \ldots, X_t]$ can be called a dictionary or training set.

The union of data-driven subspaces would be learned from $[X_1, \ldots, X_t]$ by a linear dimensional reduction where principal component analysis (PCA) is utilized to derive basis matrix Ψ^* in Eq. (4). To be concrete, a typical example is illustrated from Fig. 3. Each cluster $X_i \in \mathbb{R}^{a^2 \times l_i}$ contains l_i vectorized blocks. After the singular value decomposition (SVD) performed on X_i separately, the basis $\psi_i^* \in \mathbb{R}^{a^2 \times d_i}$ of each subspace $\mathcal{S}_i^*$ is obtained where d_i is the dimension of $\mathcal{S}_i^*$. Hence, $\Psi^* = [\psi_1^*, \psi_2^*, \ldots, \psi_t^*]$ is the bases of $\mathcal{U}^*$ where ψ_i^* is an orthonormal matrix. $q = \Sigma_{i=1}^t d_i$ should be larger than m, which makes Eq. (11) underdetermined.

3.2. Stable Reconstruction

Similar to Eq. (5), it can be derived as

$$\begin{aligned} y_{m\times 1} &= \Phi_{m\times a^2} x_{a^2\times 1} \\ &= \Phi_{m\times a^2} \Psi^*_{a^2\times q} c^*_{q\times 1} = A^* c^* \end{aligned} \tag{11}$$

where Φ is an i.i.d. random matrix. Because it is based on a union of data-driven subspaces, c^* is a 1-block-sparse vector which is more sparser than c in Eq. (5). Take Fig. 4 for example, it is obvious that the sparsity of the proposed scheme exists in only one block of c^*. It is different from the sparsity of previous work which would exist in several blocks of c, as shown in Fig.1. As follows, the uniqueness and stability conditions are given for a self-contained description.

Proposition 1 *The k-block sparse vector c^* is unique with the measurements $y = A^*c^*$ if and only if $A^*c^* \neq 0$ for all $c^* \neq 0$ that is $2k$-block sparse.*

Similar to Eq. (18) in [10], if Φ is replaced by A^* and x is replaced by k-block sparse vector c^*, and set $u = c_1^* - c_2^*$, then Proposition 2 can be shown as:

Proposition 2 *The measurement matrix A^* is stable for every $2k$-block sparse vector u if and only if there exists $C_1 > 0$ and $C_2 < \infty$ such that*

$$C_1\|u\|_2^2 \leq \|A^*u\|_2^2 \leq C_2\|u\|_2^2 \tag{12}$$

Proof: First, $A^* = \Phi\Psi^*$ in terms of Eq. (11). The basis ψ_i^* is an orthonormal basis of each subspace from PCA and Φ is an i.i.d. random matrix. According to Proposition 4 and Proposition 5 in [9], we can easily prove Proposition 1 and Proposition 2.

If A^* satisfies the block-RIP condition with $\delta_{2k} \leq \sqrt{2}-1$, the vector c^* of Eq. (8) can be determined according to the convex second-order cone program (SOCP) [10].

3.3. Compressive Video Sampling from UoDS

The proposed compressive video sampling and reconstruction framework from a union of data-driven subspaces is depicted in Fig. 5. Given a video sequence with group of pictures (GOP), it is decomposed into a set of sampled key frames (KFs) and the remaining non-key frames (NKFs). The KFs can be fully sampled first. Through block matching, the decoded KFs would form blocks of different groups (clusters)

Fig. 6. Recovered samples with block size of 8×8, $m = 40$ (i.e. SR=62.5%), UoDS is trained by overlapped blocks from the KF with $d = 10$. The first and third columns are the result of standard CS, while the second and fourth columns are the result of CS based on UoDS.

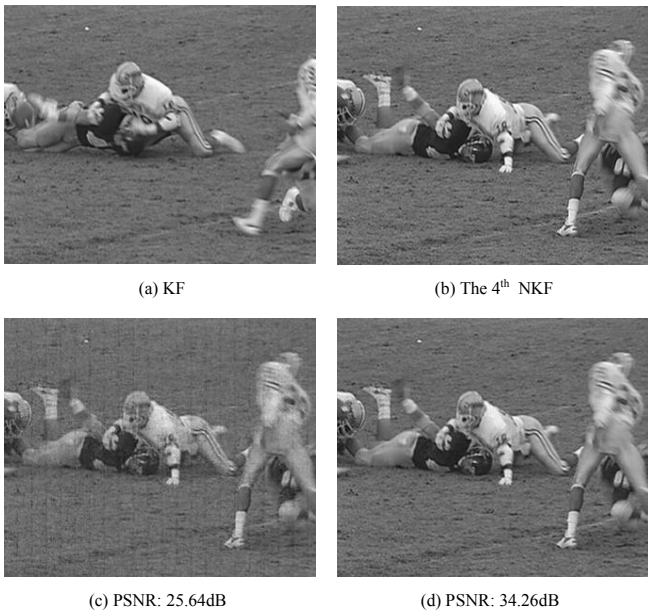

Fig. 7. The experimental results on Football with 16×16 blocks, $m = 160$ (i.e. SR=62.5%) and $d = 10$. (a): the key frame where UoDS is derived from; (b): the 4th non-key frame whose blocks are sampled compressively; (c): standard CS; (d): the proposed scheme.

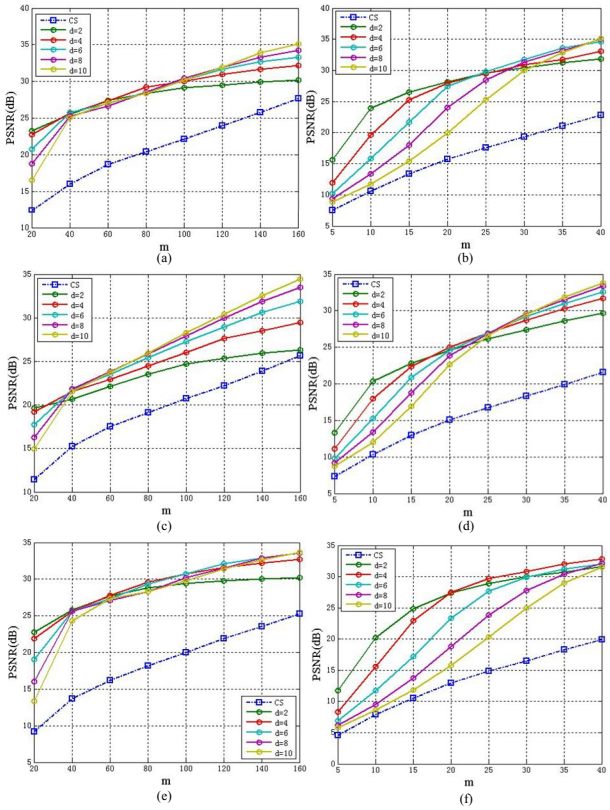

Fig. 8. The R-D performance comparison between the proposed scheme with different dimension d and standard CS. (a),(b): Akiyo; (c),(d): Football; (e),(f): Foreman. The first column is the results of 16×16 patches, while the second column is the results of 8×8 patches. UoDS is trained by overlapped patches from the KF.

and the blocks of each group are vectorized to form a data-driven subspace. All the subspaces are integrated into a union of data-driven subspaces (UoDS) $\mathcal{U}^*$, in turn the basis matrix Ψ^* from $\mathcal{U}^*$.

Consider the spatio-temporal consistency in video sequence, each vectorized non-overlapped block x in NKFs is supposed to live in $\mathcal{U}^*$ and can be represented as a linear combination of a small subset of atoms from Ψ^*. The measurements y of x are derived according to Eq. (2) at the encoder side. Thus, the vector c^* of x can be recovered by

$$\min_{c^*} \|c^*\|_{2,\mathcal{I}} \quad \text{subject to } y = A^* c^* \tag{13}$$

where $A^* = \Phi\Psi^*$. Finally, the recovered vectorized block x in NKFs can be represented by

$$x = \Psi^* c^* \tag{14}$$

and resembled back to form NKFs in the decoder.

4. EXPERIMENT

In experiments, a variety of test sequences (i.e., *foreman, akiyo, bike, tempete, football, and mobile*) are with CIF

Table 1. PSNR (dB) comparison between standard CS and UoDS trained by non-overlapped patches from KF, when block size is 8×8, $m = \{10, 20, \ldots, 60\}$.

Sequences	Model	m					
		10	20	30	40	50	60
Foreman	Standard	7.646	13.041	16.568	19.992	24.060	30.811
	UoDS	7.504	12.406	17.609	22.801	28.786	35.802
Akiyo	Standard	10.615	15.993	19.202	22.656	26.668	33.715
	UoDS	11.087	16.955	22.878	28.440	33.075	38.018
Bike	Standard	11.232	15.266	18.243	21.196	25.120	32.240
	UoDS	11.343	15.813	20.460	25.105	30.309	36.885
Templete	Standard	12.789	16.810	19.554	22.447	26.274	33.307
	UoDS	13.608	18.058	22.010	25.277	28.902	34.207
Football	Standard	10.234	14.982	18.141	21.443	25.627	32.669
	UoDS	9.935	15.360	21.276	26.262	31.563	37.834
Mobile	Standard	7.291	11.788	14.821	17.827	21.626	28.676
	UoDS	7.183	11.378	15.574	19.806	24.197	29.923

(352×288) resolution, YUV 4:2:0 format, and a GOP size of 10 frames. The size of each block is 8×8, 16×16 thereby $n = 64$ and 256 respectively. The sampling matrix is an i.i.d. Gausisian random matrix with zero-mean and unit-variance. Both situation of UoDS trained by overlapped and non-overlapped patches from key frame is considered. Without loss of generality, the first frame of each GOP is set as the key frame and the remaining nine frames as the non-key frames. Take foreman sequence for example. For non-overlapped situation, the key frame is divided into 12 clusters with $\tau = 0.7$; the number of 8×8 blocks in each cluster is no less than 10; hence, if the dimension of each subspace is 10, the size of the basis matrix is 64×120. The size of sampling matrix is $m \times 64$ with $m \in \{5, 10, 15, \ldots, 60\}$. If $m = 20$, the sampling rate (SR) $m/n = 31.25\%$. While for overlapped situation, the key frame is divided into 18 clusters with $\tau = 2$; to guarantee sufficient training, the number of 8×8 blocks in each cluster is no less than 500. For trained by overlapped 16×16 blocks situation, the key frame is divided into 19 clusters each of which also has no less than 500 blocks; the size of the sampling matrix is $m \times 256$ with $m \in \{20, 40, 60, \ldots, 160\}$. Because the union of subspaces is data-driven, the number of clusters in different video sequence is different. The dimension of each subspace d_i is simply the same from $\{2, 4, 6, 8, 10\}$. The DCT basis is used as the sparse basis in the standard CS; the bases of UoDS are learned from each cluster by PCA. All these work above are at the video encoder side. At the video decoder side, the basis pursuit (BP) [2] recovery algorithm is enabled for standard CS; the block-sparse basis pursuit algorithm is enabled for the proposed scheme, where the SPGL1 Matlab solver [1] [13] is used. The experimental environment: MATLAB in a workstation with 3.2-GHz CPU and 12-GB RAM.

Fig. 6 shows the subjective visual quality of the sampled frames with 8×8 overlapped situation, while Fig. 7 shows the subjective visual quality of the sampled frames in football sequence with 16×16 overlapped situation. Fig. 8 and Table. 1 provide the overall and averaged R-D performance. It can be seen that the proposed scheme behaves better than standard CS in general. At the same time, comparing Table.1 with Fig. 8, we can see that the performance of overlapped situation is much better than that of non-overlapped situation. Because the overlapped situation can guarantee sufficient training while non-overlapped situation can not. Performance of 16×16 situation is much better than that of 8×8. Because when n is bigger, the recovery algorithm can reconstruct more accurately c or c^* with the higher probability. In addition, the dimension of each subspace d also affects the performance of the proposed method as show in Fig.8. When d is small at a low SR (e.g. $d = 2, SR = 7.81\%(m = 20, n = 256$ *or* $m = 5, n = 64)$), the PSNR value is bigger than that of larger d.

[1] Available at http://www.cs.ubc.ca/∼mpf/spgl1/

5. CONCLUSIONS

This paper proposes an explicit sampling scheme to recover an unknown signal from a union of data-driven subspaces (UoDS). It investigates neighboring data structures by block matching to learn the union of subspaces from classified signal series. Moreover, it is optimized by principal component analysis (PCA) to derive a basis matrix and enhance the sparsity representation. With the proof of stable reconstruction, the proposed algorithm is fulfilled in digital video acquisition and recovery where the UoDS is learned from sampled key frames. The blocks of non-key frames have been evaluated to get better performance in comparison to the standard compressive sampling.

6. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289-1306, Apr. 2006.

[2] 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, Feb. 2006.

[3] E. J. Candès, T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406-5425, Dec. 2006.

[4] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *Proc. Eusipco*, Aug. 2008, pp. 1-5.

[5] J. Prades-Nebot, Y. Ma, and T. Huang, "Distributed video coding using compressive sampling," in *Proc. Picture Coding Symp.*, May. 2009, pp. 1-4.

[6] Zhaorui Liu, A. Y. Elezzabi, and H. Vicky Zhao, "Maximum Frame Rate Video Acquisition Using Adaptive Compressed

Sensing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1704-1718, Nov. 2011.

[7] Jianwei Ma, Gerlind Plonka, and M. Yousuff Hussaini, "Compressive Video Sampling With Approximate Message Passing Decoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 9, pp. 1354-1364, Sep. 2012.

[8] Ying Liu, Ming Li, and Dimitris A. Pados, "Motion-Aware Decoding of Compressed-Sensed Video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 3, pp. 438-444, Mar. 2013.

[9] Yue M. Lu and Minh N. Do, "A Theory for Sampling Signals From a Union of Subspaces," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2334-2345, Jun. 2008.

[10] Yonina C. Eldar, and Moshe Mishali, "Robust Recovery of Signals From a Structured Union of Subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302-5316, Nov. 2009.

[11] Thomas Blumensath, "Sampling and Reconstructing Signals From a Union of Linear Subspaces," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4660-4671, Jul. 2011.

[12] Marc Lebrun, "An Analysis and Implementation of the BM3D Image Denoising Method," *Image Processing On Line*, Oct. 2012.

[13] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890-912, Nov. 2008.