Residual domain dictionary learning for compressed sensing video recovery

Yun Song, Gaobo Yang, Hongtao Xie, Dengyong Zhang & Sun Xingming

Multimedia Tools and Applications
An International Journal

ISSN 1380-7501
Volume 76
Number 7

Multimed Tools Appl (2017)
76:10083-10096
DOI 10.1007/s11042-016-3599-4

Multimed Tools Appl (2017) 76:10083–10096
DOI 10.1007/s11042-016-3599-4

Residual domain dictionary learning for compressed sensing video recovery

Yun Song[1,2] · Gaobo Yang[1] · Hongtao Xie[3] · Dengyong Zhang[1,2] · Sun Xingming[4]

Received: 22 May 2015 / Revised: 28 November 2015 / Accepted: 9 May 2016 /
Published online: 16 June 2016

Abstract For compressed sensing (CS) recovery, the reconstruction quality is highly dependent on the sparsity level of the representation for the signal. Motivated by the observation that the temporal residual image is much sparser than its original image, a temporal residual-domain dictionary learning method for CS video recovery is proposed in this paper. The adaptive basis is learned from inter-frame differences by Karhunen–Loeve transform (KLT) to represent the residuals. And a block-based motion estimation/motion compensation (ME/MC) residual reconstruction strategy is incorporated for the CS video recovery. Experimental results on common test sequences at various sampling rates illustrate that the proposed algorithm gains great improvements over existing approaches. For some video sequences, the proposed method outperforms the state-of-art method near 1 dB in terms of peak signal noise rate (PSNR) at some higher sampling rate.

Keywords Compressed sensing · Dictionary learning · Residue-domain · Inter-frame difference · Video recovery

1 Introduction

Compressed sensing (CS) [4, 5] has emerged as a new signal sampling/sensing and recovery paradigm in recent years. The CS theory holds that signals or images can be

✉ Gaobo Yang
yanggaobo@hnu.edu.cn

[1] School of Information Science & Engineering, Hunan University, Changsha 410082, China

[2] School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha 410114, China

[3] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China

[4] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

recovered from far fewer measurements than that suggested by the Nyquist sampling theory, as long as they are sparse in some transform domains. A variety of applications of CS have been developed for signal and image processing. An interesting example is the single-pixel camera which directly acquires random linear projections of images by a digital micromirror device (DMD) to conduct image sampling and compression at the same time [6]. Thereafter, several important frameworks for CS image and video acquirement and compression are presented and numerous approaches are developed to reconstruct the CS-acquired images or videos via minimizing a ℓ_0 or ℓ_1 optimization problem [11, 21].

The quality of the reconstructed signals or images is determined by the number of measurements and the sparsity of the signals in the transform domain. Therefore, the main challenge of recovering signals with high quality is to seek an appropriate sparse representation for signals, i.e. a transform domain in which signals are as sparse as possible. The most commonly sparsifying transforms include discrete cosine transform (DCT) [17], discrete wavelet transform (DWT) [10, 18], finite difference (total variation, TV) [3, 19], etc. However, the image signal is typically non-stationary. There exists no universal transform space in which all parts of the signal are sufficiently sparse. This inspires us to seek adaptive basis schemes to achieve better recovery performance. In an adaptive scheme, an orthogonal basis, also called dictionary here, is usually learned from the reconstructed parts of signal or other side information for each part of the signal. Then, the learned dictionary is adopted as the sparse representation basis to recover the corresponding part of signal. Following this framework, a K-SVD [1] dictionary learning (DL)-based image recovery method was presented in [16]. In [20], the K-SVD dictionary learning and the CS image recovery were incorporated into a joint optimization mode and optimized alternately using the split Bregman iteration-based technique [9]. In [7], this integrated optimization method was extended into CS video recovery. Then, an iterative forward–backward CS-based video decoding method was proposed in [12], in which an adaptive basis was generated by Karhunen–Loeve transform (KLT). Inspired by the fact that the pixels in a block of each frame can be well predicted by nearby blocks in the adjacent frame(s), this method exploited motion estimation (ME) at the decoder side. Specifically, overlapping blocks in the searching window of the adjacent already reconstructed frame(s) were extracted to generate an adaptive basis by KL transform to represent the block to be recovered. In these methods, their dictionaries are trained/learned in the original pixel domain. Recently, a novel approach where the sparse basis is learned from the multi-scale wavelet coefficients was reported in [15]. It claims that the dictionary learned in the sparse transform domain leads to a superior recovery performance to the pixel domain dictionary. Later, a gradient domain dictionary learning method (GradDLRec) was presented for image recovery [13]. The representation dictionary was learned in the spatial gradient domain, which achieves an excellent image recovery performance. This study suggests that the sparser the training samples are, the more representation capacity the learned dictionary will have.

Motivating by the promising representation capacity of transform domain dictionary, we propose a transform-domain dictionary learning method for CS video residual reconstruction in this paper. Specifically, the residual domain dictionary learning is incorporated into the multi-reference frame motion estimation and motion compensation (ME/MC) residual recovery in an iterative manner. In each iteration, an adaptive dictionary is firstly learned from the differences between previously reconstructed frames by KLT. Then, the learned dictionary is exploited as the sparse basis to reconstruct the residual between the initial estimated frame image and its MC prediction for a refined estimation. The main contributions of this paper are summarized as follows: 1) A dictionary learning-based CS residual reconstruction scheme is presented; 2) The adaptive representation basis is proposed to learn in the temporal residual domain from previous adjacent reconstructed frames by KLT. The residual CS video recovery has shown empirically to achieve better performance than the straightforward recovery and the TV minimization recovery [14]. Thus, a residual reconstruction scheme is employed in this paper. But different from [14] where the fixed DCT basis is employed, an adaptive dictionary is applied in our method. The DL process is some like [12] where the dictionary is trained from nearby blocks in the adjacent frame(s) by KLT, so as to exploit the inter-frame similarities and correlations among temporal adjacent frames. However, the dictionary is learned in the residual domain to pursue a sparser representation in this paper. That is, the proposed method exploits the non-local similarity in the pixel domain explicitly by applying MC/ME-based residual CS recovery and further exploits the correlation in the residual domain implicitly by using KLT dictionary learning. The inter-frame differences (temporal residual) between current frame and temporal adjacent reconstructed frame(s) are sparser than the image itself and the infinite differences (gradient). Therefore, since the temporal correlations in the pixel domain and the residual domain are exploited simultaneously, it is very promising that the proposed residual domain dictionary learning can achieve a better CS recovery performance. Moreover, to enhance the sparsity level of the representation, a multi-frame reference ME/MC method is used in this paper. This leads to an even sparser representation of signals than the pixel domain and other transform domains, which naturally does benefits to a superior recovery performance. Experimental results on common test sequences show that the proposed approach presents obvious advantages over state-of-art methods.

The rest of this paper is organized as follows. Section 2 reviews some preliminaries of CS. Then, the proposed method is detailed in Section 3. Experimental results are reported in Section 4. Finally, conclusions are presented in Section 5.

2 Background

Suppose there exists a N dimensional real-valued signal $x \in \Re^N$, projecting the signal onto a sensing basis $\Phi \in \Re^{M \times N}$,i.e.

$$y = \Phi x \tag{1}$$

where $M << N$, we can acquire a M dimensional measurement vector $y \in \Re^M$. CS theory states that, if the signal is sufficiently sparse in a transform basis $\Psi \in \Re^{N \times N}$ incoherent with Φ, then x can be recovered from y by the optimization as (2), even though $M << N$ [4, 5].

$$\min \|s\|_1 \quad s.t. \quad y = \Phi x = \Phi \Psi s \tag{2}$$

where s are the coefficients of x in Ψ space, and $\| \cdot \|_1$ represents the l_1 norm. The ratio $R = M/N$ is defined as the sampling rate.

For images and videos, since the dimensionality of the input signal is very large, a huge amount of memory is demanded for storing the dense sensing matrix and the computing cost is typically high. The block-based compressed sensing (BCS) [12] is proposed to assuage the computational complexity and memory burden. In this framework, a 2D sensing matrix is employed to acquire an image or a video sequence frame by frame. Each image is partitioned into small non-overlapping blocks and each block is acquired independently. Formally, given a same block sensing matrix Φ_B for all the blocks in an image, the whole-image sensing matrix Φ takes on a block-diagonal form as (3) [8, 14].

$$\Phi = \begin{bmatrix} \Phi_B & 0 & \cdots & 0 \\ 0 & \Phi_B & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi_B \end{bmatrix} \tag{3}$$

Thus, an image or a frame of a video sequence is acquired via $y_i = \Phi_B x_i$ block by block, where $x_i \in \Re^{B^2}$ denotes the vectorized block i of the image, and $\Phi_B \in \Re^{M_B \times B^2}$ represents the block sensing matrix. The block size is $B \times B$ and the sampling rate is M_B/B^2.

For the BCS-acquired image, there are various strategies can be applied for its recovery. An image can be reconstructed block by block or be reconstructed as a whole. For block-based recovery, the whole-image sparse transform basis is block-diagonal when a same block-based basis Ψ_B is applied for all the blocks in an image. This paper adopts this framework but applies an adaptive scheme. That is, the block transform basis Ψ_B is adaptive for each block and therefore different from each other, denoted as Ψ_i for block i as illustrated in (4). Here, $N_B = N/B^2$ is the number of blocks in a frame.

$$\Psi = \begin{bmatrix} \Psi_1 & 0 & \cdots & 0 \\ 0 & \Psi_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Psi_{N_B} \end{bmatrix} \tag{4}$$

3 Proposed dictionary learning residual reconstruction method

In this effort, we propose to recover the BCS-acquired video sequence by reconstructing the inter-frame residual iteratively using the dictionary learned adaptive basis. In the proposed method, the adaptive dictionary is firstly learned from the differences between the previously reconstructed frames via KLT. Then, the learned dictionary is applied as the sparse basis to reconstruct the residual image. The residual image is the differences between the initial estimated image and its MC image. The initial estimated image is obtained via a standard straightforward CS recovery method in the first iteration. The multi-frame reference ME is introduced for temporal similar block searching to get the MC image.

The basic idea of the residual recovery comes from the traditional inter-frame coding. In traditional video compression systems, ME/MC is applied at the encoder to make inter-frame predictions so as to exploit the temporal correlation in a sequence. This technology gains a considerable coding performance improvement. Here, we incorporate DL and ME/MC into the CS video recovery process to make a residual reconstruction at the decoding side. Just as the conventional framework, several consecutive frames in a video sequence are treated as a picture of group (GOP). The first frame of the GOP is defined as the key frame and all the others in the GOP are regarded as non-key frames. The key frame is directly reconstructed using a fixed basis, while all the non-key frames are recovered using the dictionary learning-based residual reconstruction. The key frame and the previously reconstructed non-key frames are employed as the reference frames for the current non-key frame reconstruction.

A. KLT adaptive dictionary learning

The dictionary learning adaptive basis scheme is expected to alleviate the drawback of the fixed basis scheme that a given transform domain might not be universally optimal for all parts of the signal. In the proposed method, a block-based residual domain adaptive basis is applied for the residual reconstruction. The residual reconstruction will be discussed in next subsection. The adaptive basis for reconstruction is learned in the residual domain via KLT. KLT is the minimum distortion transform in the sense of the mean square error (MSE) measure and energy compaction. It catches as much energy as possible in as few coefficients as possible. Naturally, it is expected to build a powerful representational dictionary. Another reason for adopting KLT DL method is its simplicity and efficiency. In fact, KLT has been demonstrated to be a powerful and efficient CS dictionary learning method [12].

In the proposed method, the residual domain dictionary is trained from the samples of the inter-frame differences. The initial samples are built from the previously reconstructed reference frames. Specifically, all the blocks with the same size as the current block in the searching widows are firstly extracted as the reference blocks in a sliding manner. The multi-reference frames searching scheme is adopted in this paper. For the current block *i,* the searching window is defined as a square $W \times W$ region centered at the block i in each reference frame. Then, the differences between the inter-frame corresponding blocks are treated as initial atoms for dictionary learning. Before dictionary learning, all the reference frames are extended symmetrically to make the searching windows for the boundary blocks not cross the border.

After that, the KLT basis Ψ_i for the current block i, is generated by the eigenvectors of the correlation matrix of the atoms. The correlation matrix can be estimated by the average of the atoms, i.e.

$$R = \frac{1}{K} \sum_{j=1}^{K} d_j {d_j}^T \tag{5}$$

where R represents the correlation matrix of the atoms, d_j denotes the j-th atom, ${d_j}^T$ denotes its transpose, and K is the number of the atoms which equals $(W - B + 1)^2$. Then, the KLT basis Ψ_i for block i, is formed by the eigenvectors of R. By the singular value decomposition (SVD) of correlation matrix, we can get $R = U\Sigma U^T$. Where U is the eigenvector matrix and U^T is its transpose, while Σ is the diagonal matrix of the eigenvalue. Then, the eigenvector matrix gives the gradient KLT representation basis for the block *i,* i.e. $\Psi_i = U$.

As described above, the adaptive representation basis is generated directly using a KL linear transform and only once SVD decomposition is needed for each block. Therefore, the complexity for the KLT adaptive basis learning is $O(N^3)$ per block. Here, N is the number of training samples. While for K-SVD, learning the adaptive basis is an optimization problem to be solved in an iterative manner. For each iteration of the optimization process in K-SVD, a basis pursuit is needed in the sparse coding stage to get the sparse representation for each training sample. And a subsequent SVD decomposition should be applied to update each column of the dictionary in the codebook optimization stage. The computational complexity of this is naturally far higher than that of KLT.

B. ME/MC based residual reconstruction

For the current reconstructing block i, if its vectorized value is x_i and the value of its motion compensation block is x_i^{mc}, then the residual between them is $x_i^r = x_i - x_i^{mc}$. However, at the decoding side, we only have the measurement y_i of the block, which is a random projection of x_i. We can only estimate the residual in the sensing domain. Suppose we have the MC block x_i^{mc} via ME, then the projection in the sensing domain of the residual y_i^r can be approximated as follows:

$$\begin{aligned} y_i^r &= \Phi_B x_i^r = \Phi_B\left(x_i - x_i^{mc}\right) \\ &\approx \Phi_B x_i - \Phi_B x_i^{mc} = y_i - \Phi_B x_i^{mc} \end{aligned} \tag{6}$$

The MC block x_i^{mc} is the most matching block for the current block in the searching window of the previously constructed reference frames. The searching window is defined as the same as that in the previous subsection. For more accurate estimation, a multi-reference frame ME is employed in the proposed method. Multiple previously constructed frames in the GOP are regarded as the reference frames for searching. The value of the current frame reconstructed in the previous iteration is employed as an initial estimation of the current frame for matching. Note that, ME is not applied in the first iteration since there is no initial estimation for the current frame in the first iteration. A weighted average over the reference frames is employed as the MC frame. Since the residual x_i^r of the current block is sparse in the transform domain Ψ_i for the current block, it can be reconstructed by solving the following optimization problem:

$$\hat{x}_i^r = \min_{x_i^r} \left\|s_i^r\right\|_1 \quad s.t. \quad y_i^r = \Phi_B x_i^r = \Phi_B \Psi_i s_i^r \tag{7}$$

where x_i^r denotes the residual of the current block and s_i^r is its representation coefficients in the adaptive transforming space Ψ_i. Then, the current block can be estimated by (8).

$$x_i = \hat{x}_i^r + x_i^{mc} \tag{8}$$

As discussed above, the proposed residual domain dictionary learning CS video (RDDL-CSV) recovery algorithm for the non-key frames reconstruction can be summarized as Algorithm 1.

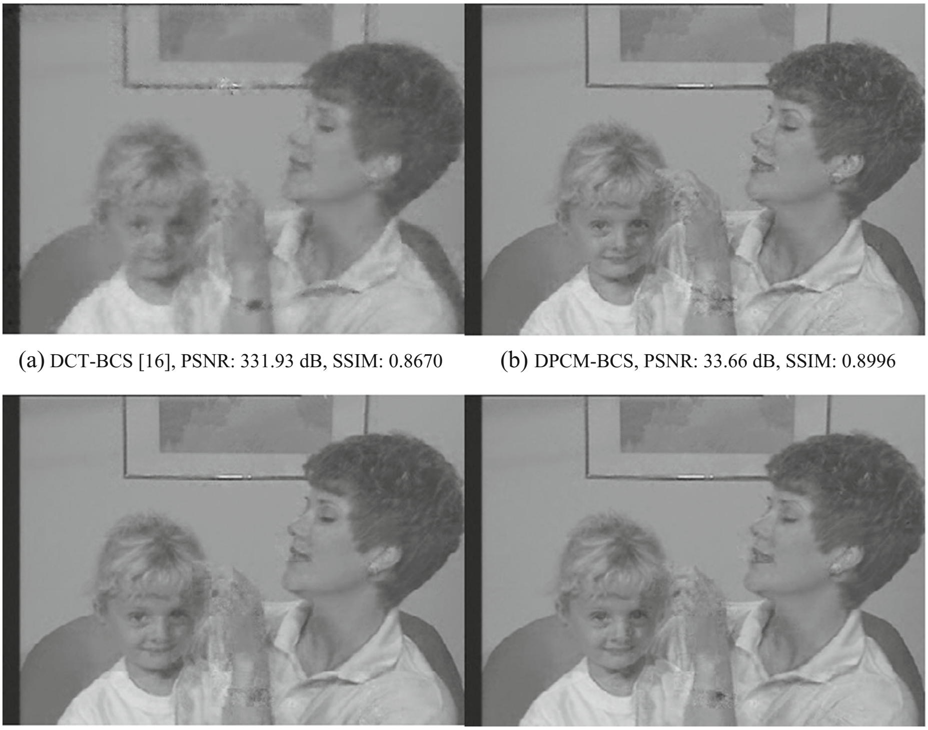
Algorithm 1: Residual domain dictionary learning CS video recovery algorithm

```
Input:  the block sensing matrix $\Phi_B$
        the measurement of the current frame $y$
        the reference frames $x^{ref}$
Output: the reconstructed frame $x$
for k=1:max_iter  do
    for i=1:block_num do
        Learning the adaptive basis $\Psi_i$ for the block $i$ via KLT
        if k==1
            $x_i^{mc} = WeightedAverage(x_i^{ref})$
        else
            Searching $x_i^{mc}$ in the reference frames via ME
        endif
        $y_i^r = y_i - \Phi_B x_i^{mc}$
        Solving $\hat{x}_i^r = \min_{x_i^r} \|s_i^r\|_1 \quad s.t. \quad y_i^r = \Phi_B \Psi_i s_i^r$
        $x_i = \hat{x}_i^r + x_i^{mc}$
    endfor
endfor
```

4 Experimental results

In this section, we evaluate the performance of the proposed method. The 2D DCT straightforward reconstruction DCT-BCS [8], and the MC-BCS-SPL [14] are chosen as benchmarks for performance comparison. In addition, a differential pulse code modulation (DPCM) reconstruction (DPCM-BCS) is implemented for evaluating the gain of the performance of the residual reconstruction. All these methods utilize the block-based smooth projected Landweber (SPL) algorithm [2, 8] for reconstruction, due to its simplicity and high efficiency. The GOP size (the number of frames in a GOP) and the block size are all set to 8 and 16, respectively. The integer pixel full-search motion estimation is applied in both MC-BCS-SPL and the proposed method for simplicity. The size of the searching window is set to 32. Five CIF (352x288) sequences, including Paris, Foreman, Mother and daughter, Tempete, and Mobile are involved for testing. For each sequence, a total of 128 frames are tested. Inspired by the experimental observations in [8], we deploy different sampling rates for the key frame and non-key frames. The sampling rates of the key frame are fixed at 0.3, 0.4, 0.5, 0.6, and 0.7 for the overall sampling rates of 0.05, 0.10, 0.15, 0.20, and 0.25, respectively. Accordingly, lower sampling rates are adopted for the non-key frames encoding.

Figure 1 shows the reconstructed images of the 69th frame of the sequence Mother and daughter recovered using different methods at sampling rate 0.15. Note that, since



(a) DCT-BCS [16], PSNR: 331.93 dB, SSIM: 0.8670

(b) DPCM-BCS, PSNR: 33.66 dB, SSIM: 0.8996

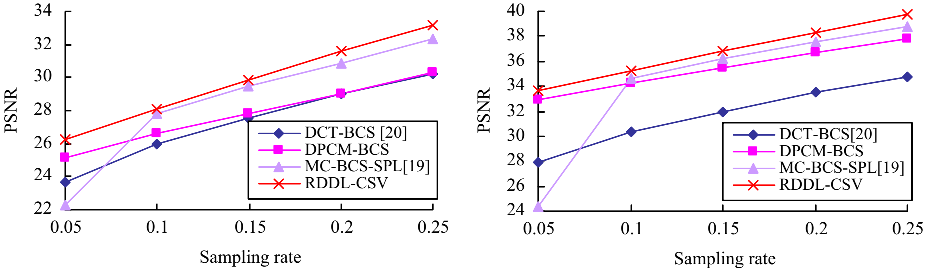
(c) MC-DCT-SPL [17], PSNR: 34.76 dB, SSIM: 0.9131

(d) RDDL-CSV, PSNR: 35.33 dB, SSIM: 0.9173

Fig. 1 Reconstructions of the 69th frame of Mother and daughter at sampling rate 0.15

the quality of reconstruction can vary due to the sensing matrix *Φ*, two same random row-orthogonal sensing matrices for the key frame and non-key frames are used for encoding in all the methods. The distortions of images measured in peak signal noise ratio (PSNR) and structural similarity index measurement (SSIM) are also given in Fig. 1. As illustrated in Fig. 1, the proposed method achieves more competitive performance than other methods both in the objective quality and visual-quality. The reconstruction of the proposed DLRR method has quite well quality whose PSNR is up to 35.33 dB with SSIM 0.9173. It is about 0.6 dB in PSNR or 0.004 in SSIM higher than the reconstruction of MC-DCT-SPL and about 3.5 dB in PSNR or 0.05 in SSIM higher than DCT-BCS which achieves a comparable reconstruction performance to TV as shown in [8].

Figure 2 presents the recovery rate-distortion (R-D) curves for Forman and Mother and daughter using different methods at various sampling rate. Note that, only the reconstruction performance of the grayscale component of the image is evaluated in terms of the average PSNR over all the 128 testing frames. The results indicate that the difference and the residual reconstruction methods outperform the straightforward reconstruction method DCT-BCS-SPL for the temporal correlation in the video sequence is taken into account. Furthermore, MC-BCS-SPL and the proposed method where ME/MC is deployed yield a significant improvement over DPCM-BCS. However, MC-BCS-SPL is observed getting a much lower performance than other methods at the sampling rate



(a) Foreman

(b) Mother and daughter

Fig. 2 R-D curves for Forman and Mother and daughter using different methods

0.05. It is mainly because this method employs the adjacent reconstructed frame as the reference frame for MC and a fixed basis for recovery. At low sampling rates, the quality of the reference frame is relative poor and the fixed basis is not sufficiently sparse for recovery from very few measurements. In contrast, the proposed method demonstrates

Table 1 Reconstruction performance (in PSNR) for various methods at different sampling rates

	Sampling rate	DCT-BCS [8]	DPCM-BCS	MC-BCS -SPL [14]	RDDL-CSV
Paris	0.05	18.25	22.03	18.64	22.38
	0.10	19.84	23.22	23.44	23.63
	0.15	20.79	24.55	25.09	25.09
	0.20	21.87	25.79	26.46	26.50
	0.25	22.68	27.02	27.84	27.96
Foreman	0.05	23.70	25.13	22.25	26.29
	0.10	26.00	26.65	27.85	28.13
	0.15	27.55	27.84	29.45	29.82
	0.20	29.03	29.04	30.90	31.58
	0.25	30.26	30.27	32.37	33.19
Mother and daughter	0.05	27.95	32.87	24.43	33.71
	0.10	30.30	34.20	34.58	35.28
	0.15	31.95	35.51	36.21	36.82
	0.20	33.55	36.68	37.58	38.30
	0.25	34.80	37.83	38.80	39.73
Tempete	0.05	20.02	22.44	20.16	23.23
	0.10	21.53	23.30	23.71	24.13
	0.15	22.62	24.12	24.66	25.00
	0.20	23.59	24.88	25.45	25.83
	0.25	24.36	25.62	26.19	26.71
Mobile	0.05	16.64	17.54	16.35	18.12
	0.10	17.73	18.03	18.60	18.73
	0.15	18.59	18.68	19.40	19.57
	0.20	19.46	19.29	20.10	20.45
	0.25	20.19	19.95	20.66	21.51

desirable performances at various sampling rates. It is natural for the proposed method to yield higher-quality reconstruction.

In addition, it is also observed from Fig. 2 that the proposed method offers a higher PSNR and SSIM gain over other methods at higher sampling rates without considering the sampling rate of 0.05. In fact, the similar results are achieved for other sequences. Detailed reconstruction performance comparative results in term of PSNR on five test sequences are summarized in Table 1. It is indicated in Table 1 that for all test sequences the proposed method achieves a superior reconstruction performance to all other methods. For some sequences (Mother and daughter, Mobile), a PSNR gain of near 1.0 dB over the current leading method MC-BCS-SPL is exhibited at the sampling rate 0.25. Meanwhile, Table 2 lists the reconstruction performance measured in SSIM to evaluate the objective quality of the reconstructed image. It also demonstrates our method outperform compared methods in terms of SSIM of the reconstructed image. That is, the proposed CS recovery method achieves not only the subjective quality but also objective quality. This benefits from the multi-frame reference MC/ME and the adaptive basis representation. Since the non-local similarity and the temporal correlation in the pixel domain and in the residual domain are both exploited, not only the prediction of our approach is more accurate but the

Table 2 Reconstruction performance (in SSIM) for various methods at different sampling rates

	Sampling rate	DCT-BCS [12]	DPCM-BCS	MC-BCS -SPL [15]	RDDL-CSV
Paris	0.05	0.4602	0.6862	0.5647	0.7089
	0.10	0.5611	0.7371	0.7415	0.7629
	0.15	0.6211	0.7865	0.8004	0.8147
	0.20	0.6736	0.8273	0.8436	0.8567
	0.25	0.7140	0.8602	0.8781	0.8892
Foreman	0.05	0.7082	0.7188	0.6068	0.7531
	0.10	0.7698	0.7473	0.7917	0.8027
	0.15	0.8088	0.7742	0.8286	0.8432
	0.20	0.8421	0.8000	0.8565	0.8761
	0.25	0.8670	0.8230	0.8801	0.9007
Mother and daughter	0.05	0.7820	0.8818	0.6779	0.8979
	0.10	0.8351	0.9006	0.9056	0.9174
	0.15	0.8675	0.9176	0.9262	0.9348
	0.20	0.8928	0.9310	0.9401	0.9486
	0.25	0.9118	0.9418	0.9509	0.9588
Tempete	0.05	0.4485	0.6465	0.5400	0.6829
	0.10	0.5553	0.6902	0.7008	0.7288
	0.15	0.6287	0.7300	0.7478	0.7706
	0.20	0.6871	0.7616	0.7806	0.8043
	0.25	0.7311	0.7872	0.8061	0.8331
Mobile	0.05	0.3132	0.3985	0.3253	0.4321
	0.10	0.4001	0.4434	0.4658	0.4876
	0.15	0.4674	0.4907	0.5231	0.5535
	0.20	0.5313	0.5293	0.5665	0.6129
	0.25	0.5823	0.5656	0.6082	0.6749

representation ability of the basis is more powerful than those used in other methods. Therefore, a desirable performance is achieved.

5 Conclusion

In this paper, we propose a dictionary learning-based residual reconstruction method for CS video recovery. In our method, a sparse representation basis is firstly learned from the inter-frame differences (residual domain) by KLT. Then, a multi-frame reference ME/MC-based residual reconstruction is performed employing the learned adaptive basis. Since the non-local similarity in the image and the temporal correlation in the residual domain are extensively exploited to enhance the sparsity, a significant performance gain is achieved over the other CS video straightforward and residual recovery methods. Experimental results show that our method outperforms the current mainstream CS video reconstruction methods in the recovery performance.

Acknowledgments This work was supported in part by the Hunan Province Science and Technology Planning Project (nos. 2014FJ6047 and 2014GK3030), the Science Research Key Project of the Education Department of Hunan Province (nos. 13A107 and 15A007), and the Changsha Science and Technology Planning Project (no. K1403028-11).

References

1. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Sign Proc 54(11):4311–4322
2. Bertero M, Boccacci P (1998) Introduction to inverse problems in imaging. Institute of Physics Publishing, Bristol
3. Bredies K, Kunisch K, Pock T (2010) Total generalized variation. SIAM J Imag Sci 3(3):492–526
4. Candes EJ, Romberg J, Tao T (2006) Stable signal recovery from incomplete and inaccurate measurements. Commun Pure Appl Math 59(8):1207–1223
5. Candes EJ, Wakin MB (2008) An introduction to compressive sampling. IEEE Signal Process Mag 25(2):21–30
6. Duarte MF, Davenport MA, Takhar D (2008) Single-pixel imaging via compressive sampling. IEEE Signal Process Mag 25(24):83–91
7. Eslahi N, Aghagolzadeh A, Andargoli SMH (2015) "Compressive video sensing via dictionary learning and forward prediction." arXiv, preprint arXiv:1508.07640
8. Gan L (2007) "Block compressed sensing of natural images". Proc Int Conf Digit Sign Process, Cardiff, UK 403–406
9. Goldstein T, Osher S (2009) The split Bregman algorithm for L1 regularized problems. SIAM J Imag Sci 2: 323–343
10. Kang LW, Lu CS (2009) "Distributed compressive video sensing,". Proc Int Conf Acoustics, Speech, Sign Process (ICASSP) 1169–1172, Taipei, Taiwan
11. Kwon S, Wang J, Shim B (2014) Multipath matching pursuit. IEEE Trans Inform Theory 50(5): 2986–3001
12. Liu Y, Li M, Pados DA (2013) Motion-aware decoding of compressed-sensed video. IEEE Trans Circ Syst Video Technol 23(3):438–444
13. Liu Q, Wang SS, Ying LL (2013) Adaptive dictionary learning in sparse gradient domain for image recovery. IEEE Trans Image Proc 22(12):4652–4663
14. Mun S, Fowler JE (2011) "Residual reconstruction for block-based compressed sensing of video". Proc Data Compress Conf 183–192, Snowbird, UT
15. Ophir B, Lustig M, Elad M (2011) Multi-scale dictionary learning using wavelets. IEEE J Sel Topics Sign Process 5(5):1014–1024
16. Ravishankar S, Bresler Y (2011) MR image reconstruction from highly undersampled k-space data bydictionary learning. IEEE Trans Med Imaging 30(5):1028–1041

17. Stankovic V, Stankovic L, Cheng S (2008) "Compressive video sampling,". Proc Eur Signal Proc Conf (EUSIPCO)
18. Wakin MB, Laska JN, Duarte MF (2006) "Compressive imaging for video representation and coding". Proc Picture Coding Symp (PCS) 711–716
19. Yue H, Ongie G, Ramani S, Jacob M (2013) Generalized higher degree total variation (HDTV) regularization. IEEE Trans Image Proc 23(6):2423–2435
20. Zhang J, Zhao C, Zhao D, Gao W (2014) Image compressive sensing recovery using adaptively learned sparsifying basis via L0 minimization. Signal Process 103:114–126
21. Zhao YB (2013) RSP-based analysis for sparsest and least ℓ1-norm solutions to underdetermined linear systems. IEEE Trans Signal Proc 61(22):5777–5788

Song Yun is pursuing his PhD from Hunan University, China. He is also an associate professor in the School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha, China. His research interest is video compression, compressive sensing for video processing.

Yang Gaobo is a professor in Hunan University, China. He is also a key member of Hunan Provincial Key Laboratory of Networks and Information Security. He received the Ph.D. degree in Communication and Information System from Shanghai University in 2004. He is the PI of several projects such as Natural Science Foundation of China (NSFC), Special Pro-phase Project on National Basic Research Program of China (973) and program for New Century Excellent Talents (NCET) in university. Currently, his research interests are in the area of image and video signal processing, digital media forensics.

Hongtao Xie received his PhD in Computer Application Technology from the Institute of Computing Technology, Chinese Academy of Sciences, China, in 2012. He is an associate professor in the Institute of Information Engineering, Chinese Academy of Sciences, China. His research interests include multimedia content analysis and retrieval, similarity search, compressive sensing and parallel computing.

Zhang Dengyong is pursuing his PhD from Hunan University, China. He is also a lecture in the School of Computer and Communication Engineering, Changsha University of Science & Technology, Changsha, China. His research interest is video compressive sensing, passive video forensics.

Sun Xingming is a professor in Nanjing University of Information Science and Techology, China. He received his PhD degree from Fudan University, China in 2003. His research interests are in the area of image and text processing, especially image and text security.