Signal Processing: Image Communication 55 (2017) 146–156

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Tensor compressed video sensing reconstruction by combination of fractional-order total variation and sparsifying transform

Gao Chen[a], Gang Li[a,*], Jiashu Zhang[b]

[a] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[b] Sichuan Province Key Lab of Signal and Information Processing, Southwest Jiaotong University, Chengdu 610031, China

ARTICLE INFO

Keywords:
Compressed video sensing
Tensor
Fractional-order total variation
Reconstruction
Smoothed L0

ABSTRACT

High reconstructed performance compressed video sensing (CVS) with low computational complexity and memory requirement is very challenging. In order to reconstruct the high quality video frames with low computational complexity, this paper proposes a tensor-based joint sparseness regularization CVS reconstruction model FrTVCST (fractional-order total variation combined with sparsifying transform), in which a high-order tensor fractional-order total variation (FrTV) regularization and a tensor discrete wavelet transform (DWT) L0 norm regularization are combined. Furthermore, an approach for choosing the regularization parameter that controls the influence of the two terms in this joint model is proposed. Afterwards, a tensor gradient projection algorithm extended from smoothed L0 (SL0) is deduced to solve this combined tensor FrTV and DWT joint regularization constrained minimization problem, using a smooth approximation of the L0 norm. Compared with several state-of-the-art CVS reconstruction algorithms, such as the Kronecker compressive sensing (KCS), generalized tensor compressive sensing (GTCS), N-way block orthogonal matching pursuit (N-BOMP), low-rank tensor compressive sensing (LRTCS), extensive experiments with commonly used video data sets show the competitive performance of the proposed algorithm with respect to the peak signal-to-noise ratio (PSNR) and subjective visual quality.

1. Introduction

Compressive sensing or compressed sensing (CS) [1,2] is an emerging framework for data acquisition and reconstruction, which permits us to reconstruct the original sparse or compressible signals from only a small number of linear measurements. As applied to video-compressed video sensing (CVS), due to its obvious advantages of using a sampling frequency much lower than the Nyquist sampling theory and little loss in reconstruction quality, it has wide potential applications in many fields, such as video surveillance [3,4] and commercial video processing cameras [5].

Conventional CS is introduced in a framework of one-dimensional (1D) formulation, when it is applied to three-dimensional (3D) video, it can be easily achieved by treating the three-dimensional (3D) group of frames as a single 1D vector [6]. However, this vector space model destroyed the intrinsic spatial structure of video and imposed a huge computational and memory burden. In order to reduce the computational complexity and memory requirements, a block-based compressive sensing (BCS) framework for video signals was proposed in [7]. In BCS the original video frames were divided into non-overlapping small blocks and each block was sensing measured independently using the same measurement matrix. Besides, in order to exploit the temporal correlation, the BCS technique incorporated reconstruction from a residual arising from motion estimation (ME) and motion compensation (MC). Nevertheless, this method incurred blocky effects and destroyed the intrinsic structure of video data. In order to preserve the intrinsic structure of video data, a Kronecker compressive sensing (KCS) method was introduced in [8] for CS of multidimensional signals. In KCS, the Kronecker sensing matrix and Kronecker sparse dictionary were proved to be satisfied mutual incoherence property (MIP) and restricted isometric property (RIP). However, KCS was not applicable as the vectorization of multidimensional signals involved a very large Kronecker product-based sensing matrix. Another method to preserve the intrinsic structure of video data is to treat the video sequences as high-order tensors and extend the CS theory to tensor CS theory. In [9], a multi-way compressive sensing (MWCS) approach for sparse and low-rank tensors was proposed. MWCS algorithm had high recovery performance, but this performance relied on tensor rank estimation, which is a non-polynomial (NP) hard problem. A generalized tensor compressive sensing (GTCS) method for high-order tensors had been

* Corresponding author.
E-mail address: gangli@mail.tsinghua.edu.cn (G. Li).

http://dx.doi.org/10.1016/j.image.2017.03.021
Received 30 October 2016; Received in revised form 27 March 2017; Accepted 27 March 2017
Available online 30 March 2017

proposed in [10]. GTCS was demonstrated to outperform KCS and MWCS in terms of both reconstruction precision and computational complexity when the size of the signal was small. However, when the size of the signal was large, the recovery precision and speed declined quickly, it was only suitable for small-scale signals. Besides, as was pointed out in [10], GTCS usually required slightly more number of measurements to achieve the same reconstruction accuracy in comparison with KCS.

It is well known that the reconstruction quality of the signals relies heavily on the sparsity level of the original signal. In [11], by exploiting the inherent tree-structure sparsity pattern (TSSP), a sparse representation model for tensors based on Tucker model and the equivalent Kronecker structure was proposed. Its disadvantage was that there was not an efficient reconstruction algorithm for TSSP. In [12], a high-accuracy total variation (TV) sparseness model for CVS reconstruction was proposed. Unfortunately, this method had the disadvantages of high computational complexity and memory requirement. Recently, by using the Kronecker product and block sparsity, a tensor CS reconstruction algorithm named N-BOMP (N-way block OMP) was developed in [13], but this algorithm should assume that the tensor signals to be sampled and reconstructed must have a perfect block sparsity pattern in that the significant coefficients in some transform domain fall within the specified block sparsity pattern while other insignificant coefficients are entirely removed [14]. More recently, a fast low rank tensor CS (LRTCS) method was proposed in [15], which no longer assumed certain sparsity pattern and did not involve iterations, making it suitable for large-scale problems. However, it assumed that the signal to be sensed and recovered must have low multi-linear-rank, leading to redundant sensing, which meant that under the same measurement rate the reconstructed quality was remarkably lower than other CVS algorithms [14].

TV regularization model, proposed by Rudin *et al.* [16], plays a key role in image/video CS reconstruction, this model enforces sparsity for piecewise constant signals and has advantage of preserving edges [12,17,18]. However, it often favors piecewise constant solution and therefore underperforms in smooth transition reconstruction and causes the staircase effect [19]. In order to reduce this staircase effect and provide more smooth transitions in the recovered signal, high order derivatives have been proposed in [12,20–22]. Nevertheless, high order derivatives often suffer from blurring of the edges. In order to improve the ability of texture preservation, a class of non-local fractional-order differential techniques has been widely used in recent years. Unlike the integer order derivative, the fractional order derivative takes more neighboring pixel information into account and is of the non-local property. In [23], a new class of fractional-order anisotropic diffusion equations was proposed for noise removal, in which the fractional derivative values were computed by using the discrete Fourier transform. In [24], based on the Grunwald-Letnikov definition of fractional-order derivative, a class of fractional-order multi-scale variational models for image denoising was proposed and an alternating projection algorithm was developed to solve this denoising model. In [25], a fractional-order TV-L2 model was proposed for image denoising and the solution of this model was provided by using the majorization-minimization (MM) algorithm. Based on these works, it can be concluded that the fractional derivative regularization model performs well for eliminating staircase effect and preserving textures [26].

Natural images and videos are of morphological diversities and are often rich in textures and fine details, only using a certain regularization sparse representation is difficult to effectively represent all of the features. Each specific transform is often limited in that it may be optimal for one kind of features, but may be poor for other kinds of features. In order to have more flexibility to achieve sparser representation, some scholars put forward multi-structural compressed sensing [27–29]. In the framework of multi-structural compressed sensing, multiple structure of signals, such as the multi-sparsity in different domains, piecewise smoothness, low rank, etc, are exploited for the recovery of the signals. It seems that these methods adequately consider the different geometrical characteristics of signals, and hence they can get more effective representation than only using single sparse constraint.

Motivated by these works, this paper proposes a tensor-based CVS reconstruction model FrTVCST (fractional-order total variation combined with sparsifying transform) by introducing the fractional differential regularization model to CVS. According to our best knowledge, it is the first time that fractional operator techniques are applied to solve the CVS reconstruction problem. In our proposed model, by using the tensor calculus, the fractional total variation (FrTV) model [26,30] is generalized to 3D video and a combination of FrTV regularization and tensor discrete wavelet transform (DWT) L_0 norm regularization is incorporated into the sparseness model of the CVS framework. Furthermore, an approach for choosing the regularization parameter that controls the influence of the two terms in the joint sparseness regularization model is proposed. Afterwards, a tensor gradient projection algorithm extended from smoothed L0 (SL0) [31,32] is derived to solve this combined tensor FrTV and DWT constrained minimization problem, using a smooth approximation of the L0 norm. The main advantages of FrTVCST include: (1) videos could not be divided into small blocks or stacked into large vectors (2) Compared with several conventional and state-of-the-art CVS reconstruction algorithms, such as the KCS, GTCS, N-BOMP, LRTCS, our proposed algorithm has better reconstruction performance in terms of the peak signal-to-noise ratio (PSNR) and subjective visual quality.

The rest of this paper is organized as follows. Section 2 briefly introduces tensor notation, definitions and basic results used throughout the paper. It also introduces conventional CS theory and the extended tensor CS theory. In Section 3, our proposed tensor CVS reconstruction algorithm based on FrTV and tensor DWT is described in detail, including the framework of our proposed algorithm, sparseness model and reconstruction algorithm. Experiment results are presented and discussed in Section 4. Finally some conclusions are given in Section 5.

2. Tensor notation and compressive sensing theory

2.1. Tensor notation and definitions

To make our mathematical formulations clear in this paper, we use an overlined boldface uppercase letter to denote a tensor, a bold uppercase letter to denote a matrix, a boldface lowercase letter to denote a vector, e.g. $\overline{\boldsymbol{X}} \in R^{I_1 \times I_2 \times I_3}$ is a 3D tensor of real numbers, $\boldsymbol{Y} \in R^{I_1 \times I_2}$ is a matrix of real numbers, $z \in R^I$ is a vector of real numbers. The i_{th} entry of a vector z is denoted by z_i, the element *(i, j)* of a matrix $\boldsymbol{Y}$ is denoted by $\boldsymbol{Y}(i,j)$ or $\boldsymbol{Y}_{i,j}$. The same notation is used for *N*-dimensional tensors by referring to the element $(i_1, i_2, \cdots i_N)$ as $\overline{\boldsymbol{X}}(i_1, i_2, \cdots i_N)$ or $\overline{\boldsymbol{X}}_{i_1,i_2,\cdots i_N}$.

As a natural generalization, the ℓ_0 norm of a tensor $\overline{\boldsymbol{X}}$ is defined as the number of nonzero components in $\overline{\boldsymbol{X}}$. The ℓ_1 norm of a tensor $\overline{\boldsymbol{X}} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as:

$$\|\overline{\boldsymbol{X}}\|_1 = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} |\overline{\boldsymbol{X}}(i_1, i_2 \cdots i_N)| \tag{1}$$

The Frobenius norm of a tensor $\overline{\boldsymbol{X}}$ is defined as:

$$\|\overline{\boldsymbol{X}}\|_F = \sqrt{\sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} \overline{\boldsymbol{X}}^2_{i_1,i_2 \cdots i_N}} \tag{2}$$

Mode-n unfolding of a tensor $\overline{\boldsymbol{X}} = [\overline{\boldsymbol{X}}_{i_1,i_2,\cdots i_N}] \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is obtained by fixing all indices except i_n. The mode-n unfolding $\boldsymbol{X}_{(n)}$ of $\overline{\boldsymbol{X}}$ arranges the mode-n fibers to be the columns of the resulting $I_n \times (I_1 \bullet \cdots \bullet I_{n-1} \bullet I_{n+1} \bullet \cdots \bullet I_N)$ matrix.

The mode-n product of a tensor $\overline{\boldsymbol{X}} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ and a matrix $\boldsymbol{\Phi} \in R^{J \times I_n}$ is denoted by $\overline{\boldsymbol{Y}} = \overline{\boldsymbol{X}} \times_n \boldsymbol{\Phi}$ and is of size $I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$. Element-wise, the mode-n product

can be written as:

$$\overline{\boldsymbol{Y}}(i_1, \cdots i_{n-1}, j, i_{n+1} \cdots i_N) = \sum_{i_n=1}^{I_n} \overline{\boldsymbol{X}}_{i_1,\ldots,i_n,\ldots,i_N} \boldsymbol{\Phi}_{j,i_n} \tag{3}$$

$\overline{\boldsymbol{Y}} = \overline{\boldsymbol{X}} \times_1 \boldsymbol{\Phi}_1 \times \cdots \times_N \boldsymbol{\Phi}_N$ is equivalent to $\boldsymbol{Y}_{(n)} = \boldsymbol{\Phi}_n \boldsymbol{X}_{(n)} (\boldsymbol{\Phi}_N \otimes \cdots \otimes \boldsymbol{\Phi}_{n+1} \otimes \boldsymbol{\Phi}_{n-1} \otimes \cdots \otimes \boldsymbol{\Phi}_1)^T$

The dot product of two tensors $\overline{\boldsymbol{X}} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ and $\overline{\boldsymbol{Y}} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is denoted by $\overline{\boldsymbol{Z}} = \overline{\boldsymbol{X}}_{\bullet} \times \overline{\boldsymbol{Y}}$. Element-wise, the dot product can be written as:

$$\overline{\boldsymbol{Z}}(i_1, \cdots\cdots i_N) = \overline{\boldsymbol{X}}_{i_1,\ldots,i_N} \overline{\boldsymbol{Y}}_{i_1,\ldots,i_N} \tag{4}$$

2.2. Compressive sensing and tensor compressive sensing

In the standard CS framework, consider an arbitrary discrete-time compressible signal $\boldsymbol{x} \in R^I$ that can be represented as:

$$\boldsymbol{x} = \sum_{i=1}^{I} \psi_i s_i = \boldsymbol{\Psi} s, \quad K = \|\boldsymbol{s}\|_0, \quad K << I \tag{5}$$

where $\boldsymbol{s} = [s_1, \ldots, s_I]$ is the vector of sparse coefficients that represent signal on the basis $\boldsymbol{\Psi}$. Take J linear non-adaptive measurements of $\boldsymbol{x}$ through a linear transformation $\boldsymbol{A}$, *i.e.*:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = A\boldsymbol{\Psi} s = \boldsymbol{A}^{CS} \boldsymbol{s} \tag{6}$$

where $\boldsymbol{y} \in \boldsymbol{R}^J$ is the measurement vector and $\boldsymbol{A} \in \boldsymbol{R}^{J \times I} (J < I)$ is the measurement matrix.

Reconstructing $\boldsymbol{x}$ from $\boldsymbol{y}$ is an ill conditioned problem as the number of measurements $J < I$. However, according to the CS theory, the original signal can be exactly reconstructed from the measurements by solving an optimization problem as long as the measurement matrix $\boldsymbol{A}$ and the representation basis $\boldsymbol{\Psi}$ are sufficiently incoherent and their product satisfies the Restricted Isometry Property (RIP):

$$\min_S \|\boldsymbol{s}\|_0 \quad s.t. \quad A\boldsymbol{\Psi} s = \boldsymbol{y} \tag{7}$$

However, optimization problem (7) is not convex, searching a solution of this problem with the constraint (6) is NP hard. To overcome this difficulty, it was demonstrated that under certain condition, optimization problem (7) can be commonly replaced by solving the linear programming problem as:

$$\min_S \|\boldsymbol{s}\|_1 \quad s.t. \quad A\boldsymbol{\Psi} s = \boldsymbol{y} \tag{8}$$

where $\|\boldsymbol{s}\|_1$ is the L_1 norm of $\boldsymbol{s}$ defined as $\|\boldsymbol{s}\|_1 = \sum_{i=1}^{I} |s_i|$. The current commonly used algorithms for solving problem (7) and (8) include basis pursuit (BP) [33], matching pursuit (MP) [34], orthogonal matching pursuit (OMP) [35], gradient projection (GP) [36] and so on.

In recent years, there has been a grown interest in CS theory for N-dimensional signal application. The most straight forward implementation of CS on N-dimensional signal is to stack this signal into a long vector, this vector space model destroyed the intrinsic spatial structure of the signal and imposed a huge computational and memory burden. To overcome this problem, some researchers extended 1D CS to tensor CS [9,10,15].

Tensor CS provides an efficient approach for representation of the multidimensional signal by offering simultaneous sampling and compression from all tensor modes. In tensor CS, the set of compressive measurements of an N_{th}-order sparse tensor $\overline{\boldsymbol{X}} \in R^{I_1 \times I_2 \times \cdots \times I_N}$ is acquired by multiplying this tensor by sensing matrices in each mode [15]:

$$\overline{\boldsymbol{Y}} = \overline{\boldsymbol{X}} \times_1 \boldsymbol{\Phi}_1 \times \cdots \times_N \boldsymbol{\Phi}_N \tag{9}$$

where $\boldsymbol{\Phi}_i \in R^{J_i \times I_i} (J_i < I_i, i = 1 \cdots N)$ are Gaussian random matrices. Reconstructing $\overline{\boldsymbol{X}}$ from $\overline{\boldsymbol{Y}}$ is realized by solving an optimization problem:

$$\min \|\overline{\boldsymbol{X}}\|_0 \quad s.t. \ \overline{\boldsymbol{Y}} = \overline{\boldsymbol{X}} \times_1 \boldsymbol{\Phi}_1 \times \cdots \times_N \boldsymbol{\Phi}_N \tag{10}$$

Numerous 1D CS reconstruction algorithms, such as BP, OMP can be easily generalized to tensor CS reconstruction [8,13] to solve problem (10).

3. Tensor compressed video sensing reconstruction

3.1. Tensor sparseness model for CVS reconstruction

We consider the following tensor-based CVS model:

$$\overline{\boldsymbol{Y}} = \overline{\boldsymbol{X}} \times_1 \boldsymbol{\Phi}_1 \times_2 \boldsymbol{\Phi}_2 \times_3 \boldsymbol{\Phi}_3 \tag{11}$$

where $\overline{\boldsymbol{X}} \in R^{I_1 \times I_2 \times I_3}$ is an original video sequence, $\boldsymbol{\Phi}_i \in R^{J_i \times I_i} (J_i < I_i, i = 1, 2, 3)$ are Gaussian random matrices, $\overline{\boldsymbol{Y}} \in R^{J_1 \times J_2 \times J_3}$ is the compressive measurement tensor. The sampling ratio is defined as:

$$\gamma = J_1 \times J_2 \times J_3 / I_1 \times I_2 \times I_3 \tag{12}$$

The CVS reconstruction is to reconstruct the original video sequence $\overline{\boldsymbol{X}}$ from the compressive measurement tensor $\overline{\boldsymbol{Y}}$.

TV regularization is one of the most successful sparse representation methods for CS reconstruction [37]. However, TV is based on the assumption that the signal is piecewise smooth. This assumption makes the TV regularization model suffer from oversmoothing, which means that TV cannot perfectly preserve texture and details. To alleviate this side effect of TV, many methods with high-order regularization items were proposed. Nevertheless, a high-order term will cause speckle-like noise. To achieve a good trade off between them, we introduce a fractional-order regularization item into the CVS reconstruction algorithm.

As the generalization of the integer-order derivative, the fractional-order derivative has long history but it has no uniform definition until now, the definition given by Grunwald-Letnikov (G–L) [38] is popular and widely used in image processing. For easy reading, we reproduce the G-L definition of the fractional-order derivative.

Without loss of generality, let $\boldsymbol{f} = [f_1, f_2, \cdots f_I] \in R^I$ denote a one-dimensional discrete time signal, then the first order difference of $\boldsymbol{f}$ is:

$$\nabla \mathbf{f} = [f_1, f_2 - f_1, f_3 - f_2, \ldots, f_{I-1} - f_{I-2}, f_I - f_{I-1}] \tag{13}$$

The second order difference of $\boldsymbol{f}$ is:

$$\nabla^2 \mathbf{f} = [f_1, f_2 - 2f_1, f_3 - 2f_2 + f_1, \ldots, f_{I-1} - 2f_{I-2} + f_{I-3}, f_I - 2f_{I-1} + f_{I-2}] \tag{14}$$

Using the mathematical induction method, the n-order difference of $\boldsymbol{f}$ can be written as:

$$(\nabla^n \mathbf{f})_i = \sum_{t=0}^{T} (-1)^t C_n^t \mathbf{f}_{i-t} = \sum_{t=0}^{T} (-1)^t \frac{n!}{(n-t)!t!} \mathbf{f}_{i-t} \quad T = \min\{i-1, n\} \tag{15}$$

The fractional order derivative of the digital signal $\boldsymbol{f}$ is defined by using an arbitrary real number $\alpha \in \boldsymbol{R}^+$ to replace the integer n and using the generalized binomial coefficients to replace the traditional binomial coefficients:

$$(\nabla^\alpha \mathbf{f})_i = \sum_{t=0}^{i-1} (-1)^t \frac{\Gamma(\alpha+1)}{\Gamma(t+1)\Gamma(\alpha-t+1)} \mathbf{f}_{i-t} \tag{16}$$

where α is the fractional order, $\Gamma(x)$ is the Gamma function.

There are two main reasons why the fractional order derivative performs better than integer-order derivative. First, as shown in Fig. 1, unlike integer-order derivative, fractional-order derivative at a point $\boldsymbol{f}_i$ is not determined by a small neighborhood of $\boldsymbol{f}_i$, in other words, fractional-order derivative is not a local property. This property may enable fractional-order derivative to deal well with texture information

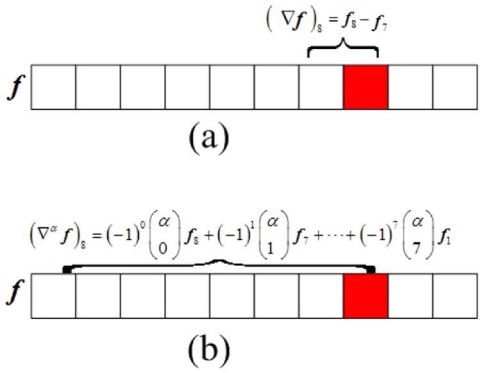

Fig. 1. (a) Integer-order derivative (b) Fractional-order derivative.

Table 1
The proposed FrTVCST reconstruction algorithm.

Algorithm 1: Tensor CVS reconstruction algorithm

1. **Input:** $\overline{Y}$, $\boldsymbol{\Phi}_i$, $\boldsymbol{\Psi}_i (i = 1, 2, 3)$, α, μ_0
2. **Initialization**
1. Let $\overline{X}_0 = \overline{Y} \times_1 \boldsymbol{\Phi_1}^+ \times_2 \boldsymbol{\Phi}_2{}^+ \times_3 \boldsymbol{\Phi}_3{}^+$, $\sigma_1 = 2\max_{i,j,k}(|\overline{X}^0(i, j, k)|)$
2. Choose suitable constants $0 < d < 1$ and L, then assume a decreasing sequence for σ, as $[\sigma_1, d\sigma_1, \ldots, d^L\sigma_1]$
1. For $l = 1, \ldots, L$
1. Let $\sigma = \sigma_l$
2. Minimize $F_{\sigma_l}(\overline{X})$ on the feasible set $\{\overline{X} | \overline{Y} = \overline{X} \times_1 \boldsymbol{\Phi_1} \times_2 \boldsymbol{\Phi}_2 \times_3 \boldsymbol{\Phi}_3\}$
1. Initialization: $\overline{X} = \overline{X}_l$
2. For $t = 1, \cdots T$ (loop T times):
1. Let $\overline{X} \leftarrow \overline{X} - \mu_0\sigma^2\frac{\partial F_\sigma(\overline{X})}{\partial \overline{X}}$
2. Project $\overline{\mathbf{X}}$ onto the feasible:
1. $\overline{X} \leftarrow \overline{X} - \overline{X} \times_1 (\boldsymbol{\Phi_1}^+\boldsymbol{\Phi_1}) \times_2 (\boldsymbol{\Phi}_2{}^+\boldsymbol{\Phi}_2) \times_3 (\boldsymbol{\Phi}_3{}^+\boldsymbol{\Phi}_3) + \overline{Y} \times_1 \boldsymbol{\Phi_1}^+ \times_2 \boldsymbol{\Phi}_2{}^+ \times_3 \boldsymbol{\Phi}_3{}^+$
3. Update the regularization parameter β by using (42):
1. Set $\overline{X}_l = \overline{X}$
1. Output $\overline{\mathbf{X}}_l$ if $l = L$

of images since texture is not local in nature. It can enhance texture details of images in those areas where gray-level values do not change evidently [17,30]. Second, both integer-order and fractional-order differentiations can enhance high-frequency components, but the enhancement of integer-order differentiation is stronger than that of fractional-order differentiation, as shown in [30]. Consequently, fractional-order differentiation introduces relatively low contrast and avoids very large oscillations near edges.

Now we extend this fractional order derivative model to three-dimensional video by using the tensor calculus. Let $\overline{X} \in R^{I_1 \times I_2 \times I_3}$ denote a video sequence with I_3 frames and each frame is represented as a $I_1 \times I_2$ matrix X_k, then the discrete three-dimensional fractional-order gradient $\nabla^\alpha \overline{X}$ can be defined as

$$(\nabla^\alpha \overline{X})_{i,j,k} = \left((\nabla_1^\alpha \overline{X})_{i,j,k}, (\nabla_2^\alpha \overline{X})_{i,j,k}, (\nabla_3^\alpha \overline{X})_{i,j,k}\right) \tag{17}$$

with $1 \le i \le I_1$, $1 \le j \le I_2$, $1 \le k \le I_3$ and

$$\begin{cases} (\nabla_1^\alpha \overline{\mathbf{X}})_{i,j,k} = \sum_{t=0}^{i-1} W_t^\alpha \overline{\mathbf{X}}(i - t, j, k) \\ (\nabla_2^\alpha \overline{\mathbf{X}})_{i,j,k} = \sum_{t=0}^{j-1} W_t^\alpha \overline{\mathbf{X}}(i, j - t, k) \\ (\nabla_3^\alpha \overline{\mathbf{X}})_{i,j,k} = \sum_{t=0}^{k-1} W_t^\alpha \overline{\mathbf{X}}(i, j, k - t) \end{cases} \tag{18}$$

where $W_t^{(\boldsymbol{\alpha})} = (-1)^t\Gamma(\alpha + 1)/\Gamma(t + 1)\Gamma(\alpha - t + 1)$

Using the tensor calculus method, we can rewrite (18) by the following form:

$$\begin{cases} \nabla_1^\alpha \overline{X} = \overline{X} \times_1 \boldsymbol{D}_1 \\ \nabla_2^\alpha \overline{X} = \overline{X} \times_2 \boldsymbol{D}_2 \\ \nabla_3^\alpha \overline{X} = \overline{X} \times_3 \boldsymbol{D}_3 \end{cases} \tag{19}$$

$$\boldsymbol{D}_i = \begin{bmatrix} W_0^{\boldsymbol{\alpha}} & 0 & \cdots & 0 \\ W_1^{\boldsymbol{\alpha}} & W_0^{\boldsymbol{\alpha}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ W_{I_i-1}^{\boldsymbol{\alpha}} & W_{I_i-2}^{\boldsymbol{\alpha}} & \cdots & W_0^{\boldsymbol{\alpha}} \end{bmatrix} (i = 1, 2, 3) \tag{20}$$

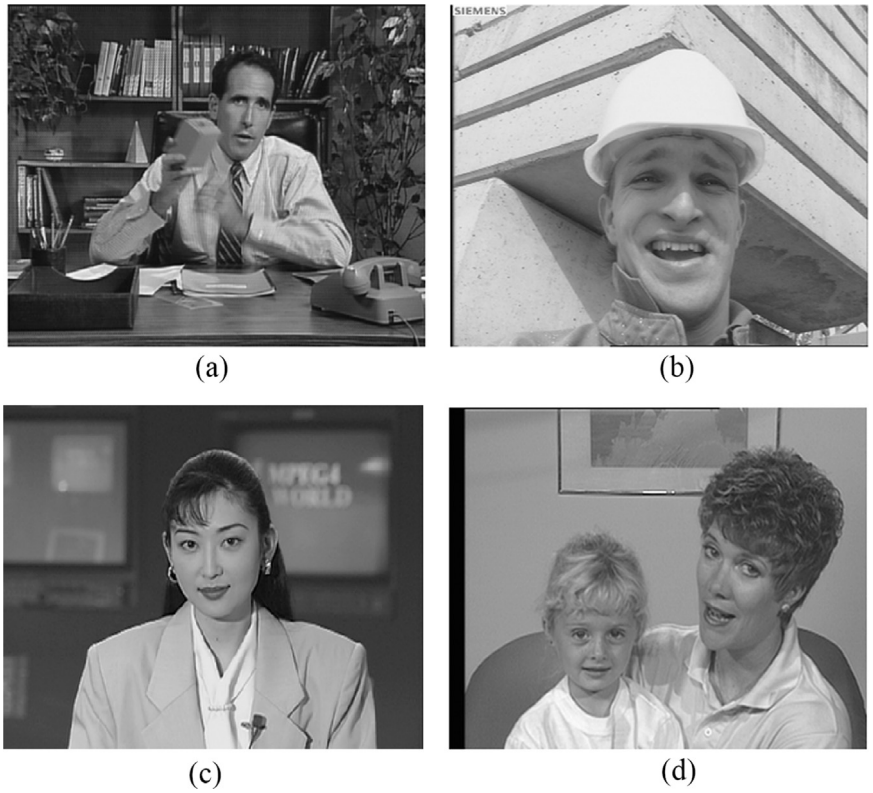

Fig. 2. Test Videos: (a) Salesman (b) Foreman (c) Akiyo (d) Mother and daughter.

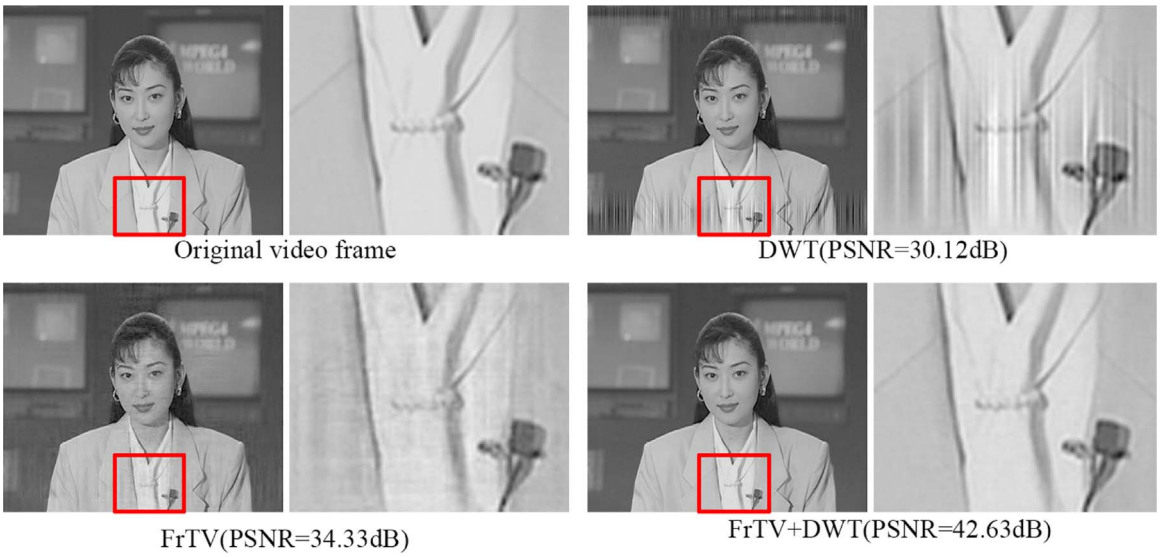

Fig. 3. Comparison of subjective quality (*Akiyo*) using different sparsifying model.

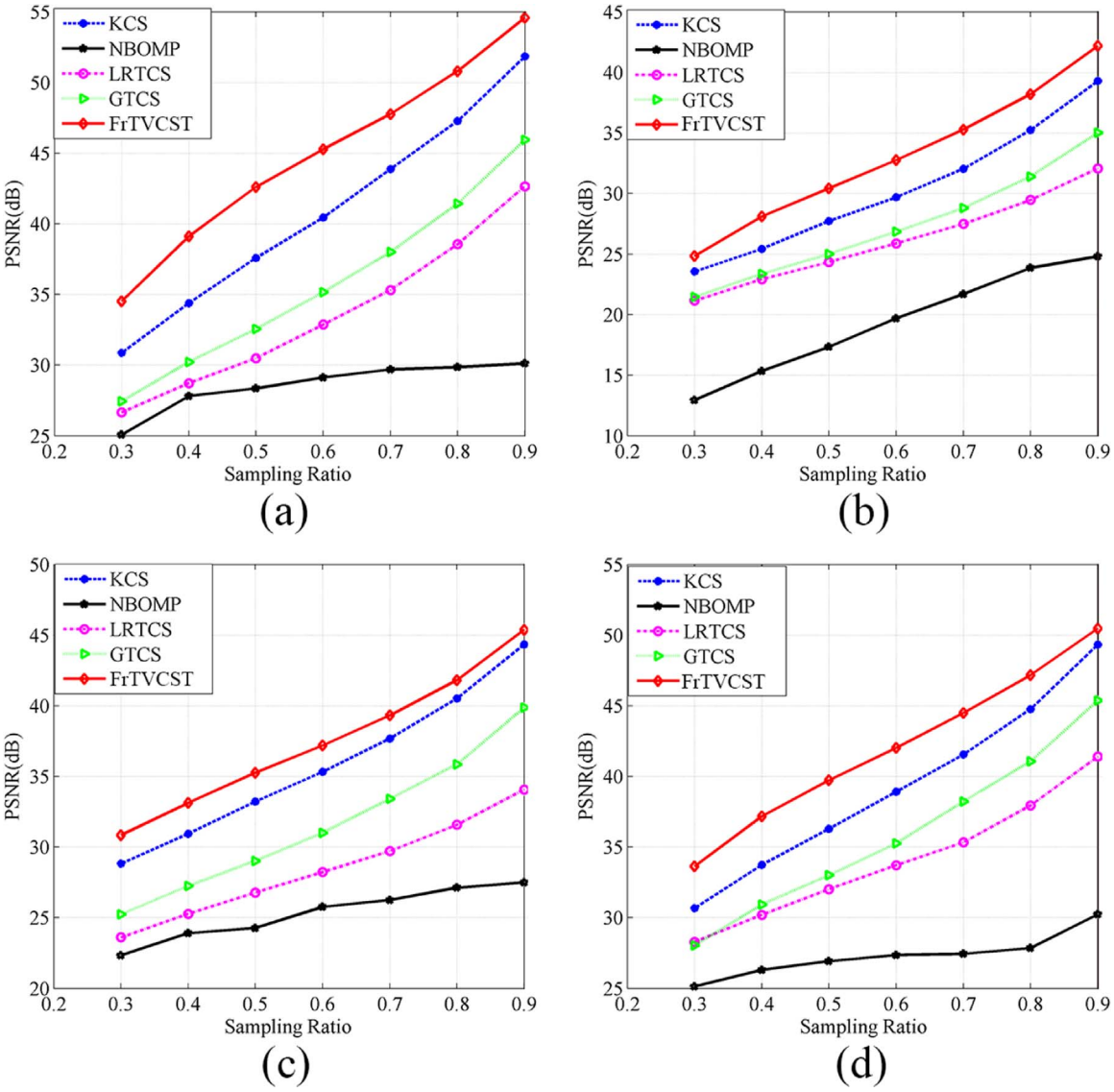

Fig. 4. Comparison of PSNR values at different sampling ratios: (a) Akiyo (b) Salesman (c) Foreman (d) Mother and daughter.

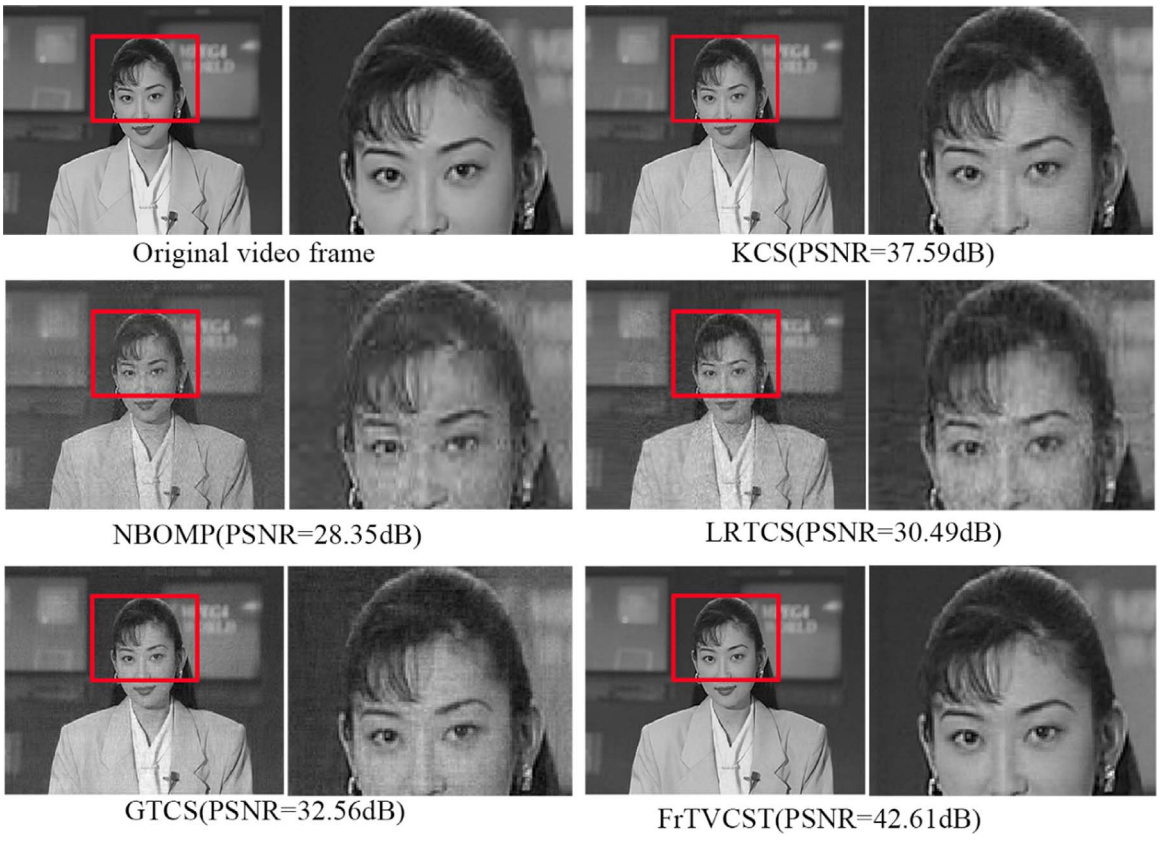

Fig. 5. Comparison of subjective quality (*Akiyo*), the sampling ratio is 50%.

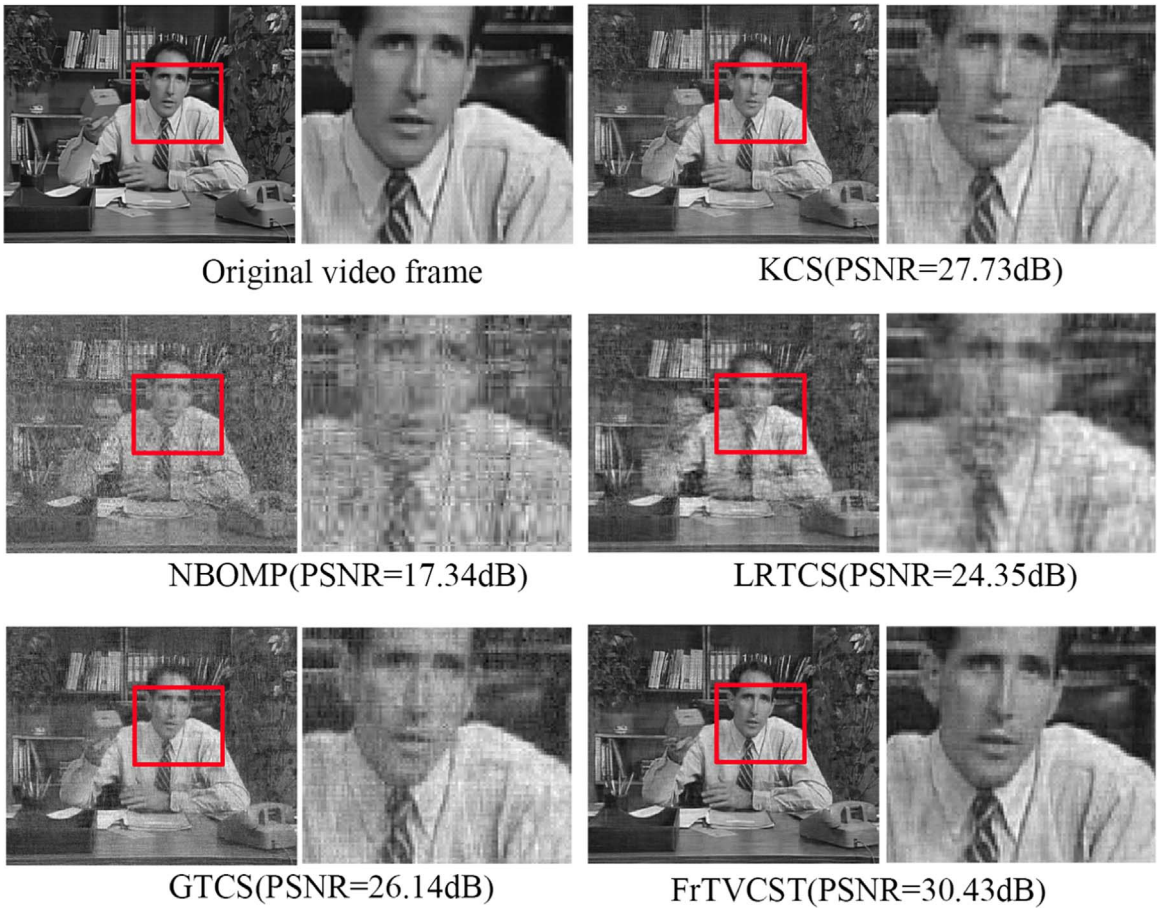

Fig. 6. Comparison of subjective quality (*Salesman*), the sampling ratio is 50%.

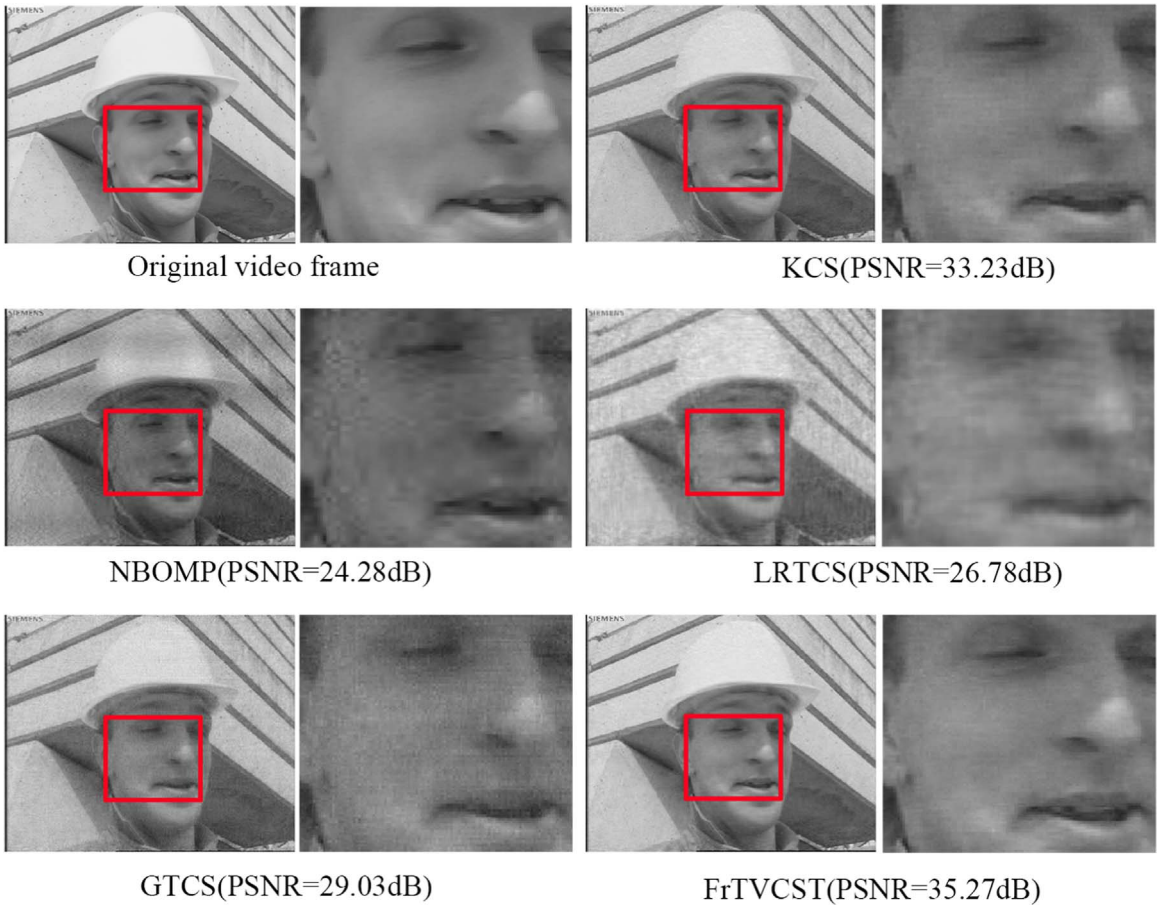

Fig. 7. Comparison of subjective quality(*Foreman*), the sampling ratio is 50%.

Then the FrTV norm of 3D tensor $\overline{X}$ is defined as:

$$\|\overline{X}\|_{FrTV} = \sum_{i=1}^{I_1}\sum_{j=1}^{I_2}\sum_{k=1}^{I_3}\sqrt{\left[(\overline{X}\times_1 \boldsymbol{D}_1)_{i,j,k}\right]^2+\left[(\overline{X}\times_2 \boldsymbol{D}_2)_{i,j,k}\right]^2+\left[(\overline{X}\times_3 \boldsymbol{D}_3)_{i,j,k}\right]^2} \tag{21}$$

Natural images and videos, which include a variety of features, such as rich textures, fine details and strong edges, are typically non-stationary, only using a certain regularization is difficult to effectively represent all of the image features and will result in a poor CS recovery performance. For example, TV model makes the recovered video frame quality sharper, but it loses some details and fine structures. DWT is good at sparsely representing point-like features but fails in sparsely representing the curve-like features. To overcome this shortage, we use multiple structural to sparsely represent the videos. According to the conclusions of our previous work in [17], more sparsifying transforms involved in the CVS model will lead to better reconstruction. However, incorporating more sparsifying transforms will result in a higher recovery time. To achieve a good tradeoff, besides FrTV, a tensor discrete wavelet transform (DWT) L_0 norm regularization is also incorporated into our sparseness model.

Tensor DWT of $\overline{X}$ can be defined by the following formula:

$$\overline{\boldsymbol{\Theta}} = \overline{\boldsymbol{X}}\times_1 \boldsymbol{\Psi}_1\times_2 \boldsymbol{\Psi}_2\times_3 \boldsymbol{\Psi}_3 \tag{22}$$

where $\boldsymbol{\Psi}_1 \in R^{I_1\times I_1}$, $\boldsymbol{\Psi}_2 \in R^{I_2\times I_2}$, $\boldsymbol{\Psi}_3 \in R^{I_3\times I_3}$ are the one-dimensional discrete wavelet transform matrices along the direction of row, column and time of $\overline{\boldsymbol{X}}$, respectively. They can be constructed by the Spot Matlab toolbox [39].

Combining (11), (21) and (22), FrTVCST can be formulated as:

$$\begin{aligned}&\min_{\overline{\mathbf{X}}}\|\overline{\boldsymbol{X}}\times_1 \boldsymbol{\Psi}_1\times_2 \boldsymbol{\Psi}_2\times_3 \boldsymbol{\Psi}_3\|_0\\ &\quad+\beta\textstyle\sum_{i,j,k}\sqrt{(\overline{\boldsymbol{X}}\times_1 \boldsymbol{D}_1)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_2 \boldsymbol{D}_2)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_3 \boldsymbol{D}_3)^2_{i,j,k}}\\ &s.\ t.\quad \overline{\boldsymbol{Y}} = \overline{\boldsymbol{X}}\times_1 \boldsymbol{\Phi}_1\times_2 \boldsymbol{\Phi}_2\times_3 \boldsymbol{\Phi}_3\end{aligned} \tag{23}$$

where β is a positive regularization parameter to control the sparsity in the tensor DWT domain and in the FrTV domain.

3.2. CVS reconstruction algorithm

The minimization problem (23) is a NP hard problem and cannot be solved directly due to the non-convex of the L_0 norm regularization term. In this section, a tensor gradient projection algorithm extended from smoothed L0 (SL0) [31,32] is derived to solve this non-convex problem effectively.

The basic idea of SL0 is using the following continuous Gaussian function [31] to approximate the L_0 norm and changing the non-convex optimization problem to a convex one:

$$\|\boldsymbol{x}\|_0 \approx N - F_\sigma(\boldsymbol{x}) = N - \sum_{i=1}^{N} f_\sigma(x_i) = N - \sum_{i=1}^{N}\exp(-x_i^2/2\sigma^2) \tag{24}$$

where

$$f_\sigma(x_i) = \exp(-x_i^2/2\sigma^2) \approx \begin{cases}1 & if\ \ x_i << \sigma\\ 0 & if\ \ x_i >> \sigma\end{cases} \tag{25}$$

and parameter σ determines how smooth the function $F_\sigma(\boldsymbol{x})$ is: the larger value of σ, the smoother F_σ (but worse approximation to L_0 norm); and the smaller value of σ, the closer behavior of F_σ to L_0 norm

Using this approximate method, problem (23) can be formulated as the following form:

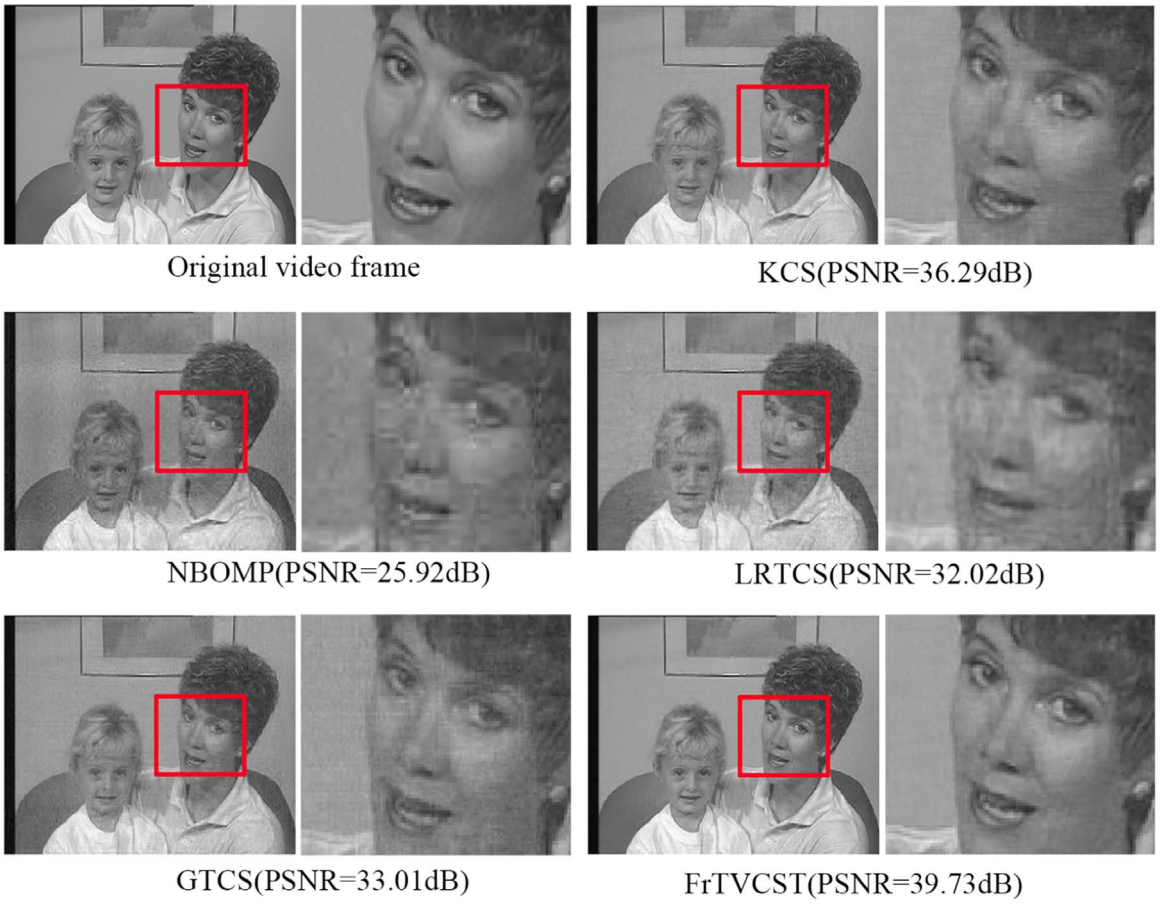

Fig. 8. Comparison of subjective quality (*Mother and daughter*), the sampling ratio is 50%.

Table 2
CPU time (sec.) required to recover the video sequence (*Akiyo*) by the five reconstruction algorithms at different sampling ratios.

Ratio	KCS	NBOMP	LRTCS	GTCS	FrTVCST
0.3	5214	140.9	1.275	2506	**379.6**
0.4	5317	181.9	1.389	2451	**395.4**
0.5	5512	212.4	1.415	2461	**396.7**
0.6	5328	213.7	1.542	2448	**399.9**
0.7	5419	245.0	1.670	2440	**402.1**
0.8	5376	274.0	1.915	2437	**407.7**
0.9	5474	294.0	2.136	2459	**418.6**

$$\begin{aligned}\min_{\mathbf{X}} \quad F_\sigma(\overline{\boldsymbol{X}}) &= f_1(\overline{\boldsymbol{X}}) + f_2(\overline{\boldsymbol{X}}) \\ &= -\sum_{i,j,k} \exp\left(-\frac{[\overline{\boldsymbol{X}} \times_1 \boldsymbol{\Psi}_1 \times_2 \boldsymbol{\Psi}_2 \times_3 \boldsymbol{\Psi}_3]^2_{i,j,k}}{2\sigma^2}\right) \\ &\quad -\beta\sum_{i,j,k} \exp\left(-\frac{(\overline{\boldsymbol{X}} \times_1 \boldsymbol{D}_1)^2_{i,j,k} + (\overline{\boldsymbol{X}} \times_2 \boldsymbol{D}_2)^2_{i,j,k} + (\overline{\boldsymbol{X}} \times_3 \boldsymbol{D}_3)^2_{i,j,k}}{2\sigma^2}\right) \\ s.t. \quad \overline{\boldsymbol{Y}} &= \overline{\boldsymbol{X}} \times_1 \boldsymbol{\Phi}_1 \times_2 \boldsymbol{\Phi}_2 \times_3 \boldsymbol{\Phi}_3\end{aligned} \tag{26}$$

Like SL0, to avoid local optimization, we solve the problem (26) by using a decreasing sequence of σ. Given a decreasing sequence of $[\sigma_1, \sigma_2 \cdots \sigma_L]$, we minimize $F_{\sigma_l}(\overline{\boldsymbol{X}})$ with tensor gradient projection method, then the minimizer of $F_{\sigma_l}(\overline{\boldsymbol{X}})$ is used as a starting point to search the minimizer of $F_{\sigma_{l+1}}(\overline{\boldsymbol{X}})$ for σ_{l+1}. In implementation of this algorithm, the initial value is set as:

$$\overline{\boldsymbol{X}}_0 = \overline{\boldsymbol{Y}} \times_1 \boldsymbol{\Phi}_1^+ \times_2 \boldsymbol{\Phi}_2^+ \times_3 \boldsymbol{\Phi}_3^+ \tag{27}$$

where $\boldsymbol{\Phi}_i^+ = \boldsymbol{\Phi}_i^T(\boldsymbol{\Phi}_i\boldsymbol{\Phi}_i^T)^{-1}$ is the pseudoinverses of $\boldsymbol{\Phi}_i (i = 1, 2, 3)$. The value of σ_l at $l+1$ iteration is $\sigma_{l+1} = d\sigma_l (0 < d < 1)$ and $\sigma_1 = 2\max_{i,j,k}(|\overline{\boldsymbol{X}}^0(i, j, k)|)$. The implementation details of minimizing $F_{\sigma_l}(\overline{\boldsymbol{X}})$ with tensor gradient projection are given below.

1) *Gradient descent:* The derivative of $F_\sigma(\overline{\boldsymbol{X}})$ with respect to the tensor $\overline{\boldsymbol{X}}$ is given by

$$\frac{\partial F_\sigma(\overline{\boldsymbol{X}})}{\partial \overline{\boldsymbol{X}}} = \frac{\partial f_1(\overline{\boldsymbol{X}})}{\partial \overline{\boldsymbol{X}}} + \frac{\partial f_2(\overline{\boldsymbol{X}})}{\partial \overline{\boldsymbol{X}}} \tag{28}$$

$$\frac{\partial f_1(\overline{\boldsymbol{X}})}{\partial \overline{\boldsymbol{X}}} = \frac{1}{\sigma^2}\overline{\boldsymbol{\Delta}} \times_1 \boldsymbol{\Psi}_1^T \times_2 \boldsymbol{\Psi}_2^T \times_3 \boldsymbol{\Psi}_3^T \tag{29}$$

$$\frac{\partial f_2(\overline{\boldsymbol{X}})}{\partial \overline{\boldsymbol{X}}} = \frac{\beta}{\sigma^2}(\overline{\boldsymbol{A}} \times_1 \boldsymbol{D}_1^T + \overline{\boldsymbol{B}} \times_2 \boldsymbol{D}_2^T + \overline{\boldsymbol{C}} \times_3 \boldsymbol{D}_3^T) \tag{30}$$

where:

$$\overline{\boldsymbol{A}} = (\overline{\boldsymbol{X}} \times_1 \boldsymbol{D}_1)_\bullet \times \overline{\boldsymbol{E}} \tag{31}$$

$$\overline{\boldsymbol{B}} = (\overline{\boldsymbol{X}} \times_2 \boldsymbol{D}_2)_\bullet \times \overline{\boldsymbol{E}} \tag{32}$$

$$\overline{\boldsymbol{C}} = (\overline{\boldsymbol{X}} \times_3 \boldsymbol{D}_3)_\bullet \times \overline{\boldsymbol{E}} \tag{33}$$

$$\overline{\boldsymbol{\Delta}}(i, j, k) = \overline{\boldsymbol{X}}(i, j, k)\exp\left(-\frac{[\overline{\boldsymbol{X}} \times_1 \boldsymbol{\Psi}_1 \times_2 \boldsymbol{\Psi}_2 \times_3 \boldsymbol{\Psi}_3]^2_{i,j,k}}{2\sigma^2}\right) \tag{34}$$

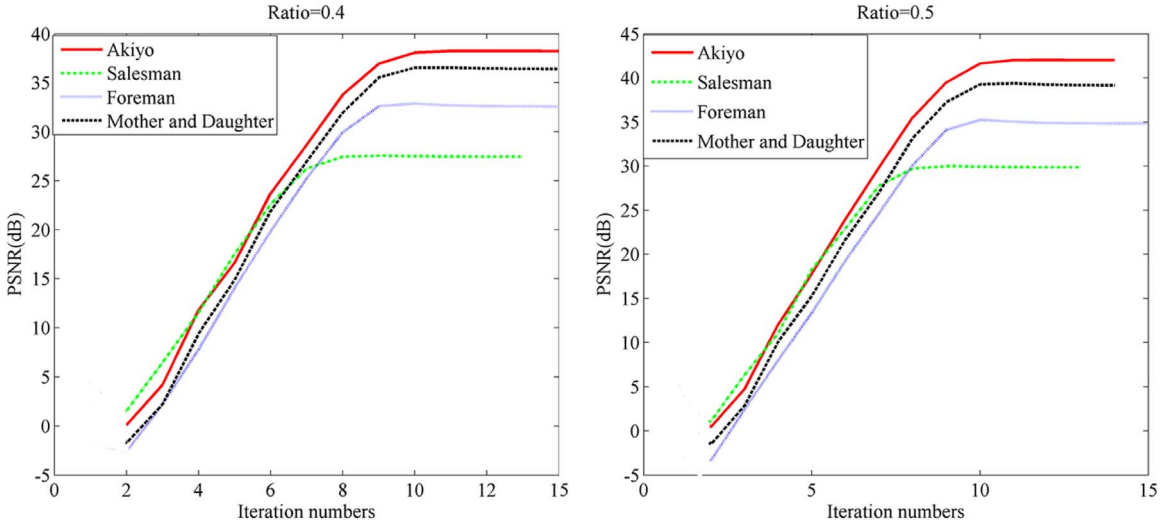

Fig. 9. Convergence of FrTVCST algorithm.

$$\overline{\boldsymbol{E}}(i,j,k)=\exp\left(-\frac{(\overline{\boldsymbol{X}}\times_1\boldsymbol{D}_1)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_2\boldsymbol{D}_2)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_3\boldsymbol{D}_3)^2_{i,j,k}}{2\sigma^2}\right) \tag{35}$$

Then the gradient descent recursion of $\overline{\boldsymbol{X}}$ for $F_\sigma(\overline{\boldsymbol{X}})$ can be written as:

$$\overline{\boldsymbol{X}}'=\overline{\boldsymbol{X}}_n-\mu_l\frac{\partial F_{\sigma_l}(\overline{\boldsymbol{X}}_n)}{\partial\overline{\boldsymbol{X}}_n} \tag{36}$$

where μ_l is the step-size. According to the suggestion of [31], we set $\mu_l=\mu_0\sigma_l^2$, μ_0 is a constant.

2) *Tensor projection:* After the first step, the iterative solution does not lie in the tensor solution space $\{\overline{\boldsymbol{X}}|\overline{\boldsymbol{Y}}=\overline{\boldsymbol{X}}\times_1\boldsymbol{\Phi}_1\times_2\boldsymbol{\Phi}_2\times_3\boldsymbol{\Phi}_3\}$, so the updated tensor is replaced by the projection of the tensor on the solution space.

$$\begin{aligned}\overline{\boldsymbol{X}}_{n+1} &= \overline{\boldsymbol{X}}'-\overline{\boldsymbol{X}}'\times_1(\boldsymbol{\Phi}_1^{+}\boldsymbol{\Phi}_1)\times_2(\boldsymbol{\Phi}_2^{+}\boldsymbol{\Phi}_2)\times_3(\boldsymbol{\Phi}_3^{+}\boldsymbol{\Phi}_3)\\ &\quad+\overline{\boldsymbol{Y}}\times_1\boldsymbol{\Phi}_1^{+}\times_2\boldsymbol{\Phi}_2^{+}\times_3\boldsymbol{\Phi}_3^{+}\end{aligned} \tag{37}$$

3) *Adaptive choice of the regularization parameter*

The regularization parameter β in our proposed reconstruction model (23) controls the tradeoff between the sparsity in the tensor DWT domain and in the FrTV domain. In this section, a strategy that can adaptively choose this regularization parameter is developed. The general idea of this approach is to employ the information available at each iteration step in the reconstruction process to realize the spatial adaptivity of β. There are three important properties for this regularization parameter:

(1) β is proportional to $\|\overline{\boldsymbol{X}}\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3\|_0$, a large $\|\overline{\boldsymbol{X}}\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3\|_0$ signifies that the reconstructed video frames are insufficient sparse in the wavelet domain, so a large regularization parameter should be selected to ensure that the reconstructed video frames are sparse in the transform domain.
(2) β is inversely proportional to $FrTV(\overline{\boldsymbol{X}})$, a large $FrTV(\overline{\boldsymbol{X}})$ signifies that the reconstructed video frames are insufficient smooth and lose some edge and detail information, so a small regularization parameter should be selected to ensure that the reconstructed video frames are sufficient smooth.
(3) β is larger than zero.

To meet the above conditions, β should satisfy:

$$\beta\propto\frac{\|\overline{\boldsymbol{X}}\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3\|_0}{FrTV(\overline{\boldsymbol{X}})} \tag{38}$$

Based on the discussion in [40], among proportional candidate functions, we choose the proportional function as:

$$\beta=\ln\left(\frac{\|\overline{\boldsymbol{X}}\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3\|_0}{FrTV(\overline{\boldsymbol{X}})+\delta}+1\right) \tag{39}$$

where δ is a small positive number used to avoid a zero denominator, assigned 10^{-7} in practice. In order to leverage the merit of SL0, we use the smooth Gaussian function to approximate the L_0 norm and FrTV norm, i.e.:

$$\|\overline{\boldsymbol{X}}\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3\|_0\approx I_1I_2I_3-\sum_{i,j,k}\exp\left(-\frac{[\overline{\boldsymbol{X}}\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3]^2_{i,j,k}}{2\sigma^2}\right) \tag{40}$$

$$\begin{aligned}FrTV(\overline{\boldsymbol{X}}) &= \sum_{i,j,k}\sqrt{(\overline{\boldsymbol{X}}\times_1\boldsymbol{D}_1)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_2\boldsymbol{D}_2)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_3\boldsymbol{D}_3)^2_{i,j,k}}\\ &\approx I_1I_2I_3-\sum_{i,j,k}\exp\left(-\frac{(\overline{\boldsymbol{X}}\times_1\boldsymbol{D}_1)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_2\boldsymbol{D}_2)^2_{i,j,k}+(\overline{\boldsymbol{X}}\times_3\boldsymbol{D}_3)^2_{i,j,k}}{2\sigma^2}\right)\end{aligned} \tag{41}$$

Therefore, we choose the regularization functional as:

$$\beta_{n+1}=\ln\left(\frac{I_1I_2I_3-\sum_{i,j,k}\exp\left(-\frac{[\overline{\boldsymbol{X}}_n\times_1\boldsymbol{\Psi}_1\times_2\boldsymbol{\Psi}_2\times_3\boldsymbol{\Psi}_3]^2_{i,j,k}}{2\sigma^2}\right)}{I_1I_2I_3-\sum_{i,j,k}\exp\left(-\frac{(\overline{\boldsymbol{X}}_n\times_1\boldsymbol{D}_1)^2_{i,j,k}+(\overline{\boldsymbol{X}}_n\times_2\boldsymbol{D}_2)^2_{i,j,k}+(\overline{\boldsymbol{X}}_n\times_3\boldsymbol{D}_3)^2_{i,j,k}}{2\sigma^2}\right)+\delta}+1\right) \tag{42}$$

Based on the above analysis, the procedures of FrTVCST can be summarized in Table 1.

4. Experimental results

In this section, the performance of the proposed tensor CVS reconstruction algorithm is illustrated via comprehensive experiments. Throughout, several popular gray-level videos with sizes $288\times352\times50$ are employed in our experiments, such as *Salesman*, *Foreman*, *Akiyo* and *Mother and daughter*, as shown in Fig. 2. All experiments are performed in MATLAB 2016a environment on a

computer with Intel(R) Core(TM) processor, 8.00 G memory, and under Microsoft Windows 7 operating system. In our experiments, Gaussian random matrices $\boldsymbol{\Phi}_i (i = 1, 2, 3)$ are generated by Matlab command $\boldsymbol{randn(J_i, I_i)}$. DWT matrices $\boldsymbol{\Psi}_i (i = 1, 2, 3)$ are constructed by the Spot Matlab toolbox [39], where the Daubechies wavelet is used. Tensor Toolbox [41] is used to calculate the tensor.

The objective quality of the reconstructed videos was measured in terms of PSNR between the reconstructed videos and original videos, which is defined as:

$$PSNR = 20\log_{10} R/RMSE \tag{43}$$

with

$$RMSE = \sqrt{\sum_{i=1}^{I_1}\sum_{j=1}^{I_2}(X_{ij} - \overline{X}_{ij})^2} \tag{44}$$

where X_{ij} and $\overline{X}_{ij}$ denote the pixel values of the reconstructed video frame and the original video frame, respectively. $R > 0$ is the maximum value of the image gray level range.

4.1. Parameter selection for FrTVCST

Before our experiments, it is worthy to discuss the parameter selection for FrTVCST reconstruction algorithm. μ_0 is the step-size parameter of the gradient projection algorithm, c is the decreasing factor of σ, according to the results of [31], we set $\mu_0 = 2$, $d = 0.5$.α is the fractional order of the FrTV, the larger the parameter α is, the better the textures are preserved, however, if α is too large, then the frequency of textures will be enhanced excessively and become a kind of noise [24]. According to the experimental results of our previous work [17] and the conclusions of [26], fractional order α can be selected between $\alpha = 1.2$ and $\alpha = 2$. In our experiments, we set $\alpha = 1.4$.

4.2. Compared with the reconstruction algorithm only using the single regularization

In this experiment, the quality of the video frames reconstructed by the proposed algorithm using FrTVCST is tested, compared with the reconstruction algorithm only using the single regularization (FrTV or DWT). Reconstructed video frames of *Akiyo* are shown in Fig. 3. From this figure, we can see that only using FrTV constraint makes the recovered video frame quality sharper, but it loses some point-like and fine structures. While only using DWT constraint, the details and fine structures of the recovered video frame are preserved. However, it generates some artifacts. While FrTV and DWT constraints are applied simultaneously, it is able to simultaneously preserve the textures and reconstruct the main geometrical configuration in a video frame, so it has higher PNSR.

4.3. Compared with the other reconstruction algorithms

In this section, the performance of the FrTVCST is demonstrated and compared with the following current state-of-the-art algorithms, which stand out for their reconstruction efficiency:

- KCS [8]: is a CS framework for multidimensional signals, in which the sparsifying bases and measurement matrices of the multidimensional signals are constructed using the Kronecker product matrices. Besides, KCS uses the L_1 -norm minimization solver-SPGL, proposed in [42] to reconstruct the original video frames.
- NBOMP [13]: is a greedy algorithm for multidimensional signal CS reconstruction that is generalized from the classical OMP algorithm for vectors.
- GTCS [10]: is a unified framework for CS of higher order tensors, which preserves the intrinsic structure of tensor data with reduced computational complexity at reconstruction.
- LRTCS [15]: is a non-iterative CS reconstruction algorithm of higher order tensors, which no longer assumes certain sparsity pattern.

The source codes of these algorithms are downloaded online. To make a fair comparison among these algorithms, we choose the empirical parameter settings by trial and error to achieve the best performance in the recovery quality. Besides, the same test conditions are used in all experiments, i.e., the same sensing matrices are applied for KCS, NBOMP, GTCS, LRTCS and the proposed FrTVCST. For each test video, the experiment is repeated one hundred times under different sampling matrix. Then, the average PSNR values are calculated. Fig. 4 shows the comparison results where the average PSNR values of four video sequences reconstructed by five reconstruction algorithms are the function of sampling ratios. It is clearly observed that FrTVCST algorithm performs best for all of the videos at any sampling ratio. The subjective visual quality of the reconstructed video frames is used to further verify the effectiveness of the proposed algorithm. To compare the results visually, Figs. 5–8 display some reconstructed video frames using the five reconstruction algorithms at 50% sampling ratio, better visual comparison can be made by zooming the video frames. We can clearly see that the proposed FrTVCST efficiently reconstructs not only sharp large-scale edges but also fine-scale details. It preserves the video frame structures and textures better than the other reconstruction algorithms. The higher performance of the proposed algorithm is attributed to the efficiently using of the joint regularization model, in which the FrTV regularization and the tensor DWT regularization are included, offering a higher sparsity.

CPU time is used as a rough estimation of computational complexity. Table 2 gives the CPU time (s) required to recover the video sequence (*Akiyo*) by the five reconstruction algorithms at different sampling ratios. From this table we can see that KCS and GTCS algorithms have the highest computational complexity. That is because KCS involved a very large Kronecker product-based sensing matrix and the computational complexity of GTCS is proportional to the size of the video frames. Compared with LRTCS and NBOMP algorithm, FrTVCST has higher complexity. However, it does not rely on any impractical assumptions and restrictions and has higher reconstruction quality.

4.4. Algorithm convergence

SL0 algorithm converges very quickly when applied to certain types of objective functions. A detailed convergence analysis of the SL0 algorithm has been established in [31]. In our algorithm, without perturbing the recovery criteria, a tensor gradient projection algorithm extended from SL0 is derived to deal with the combined FrTV and DWT constrained minimization problem, so our algorithm is also convergent. In this experiment, we provide empirical evidence to illustrate the good convergence of the proposed algorithm. Fig. 9 plots the evolutions of PSNR versus iteration numbers for four test videos with two sampling ratios (ratio = 0.4 and ratio = 0.5). It is observed that with the growth of iteration number, all the PSNR curves increase monotonically and ultimately become flat and stable, exhibiting good convergence property.

5. Conclusion

CVS is remarkably challenging due to the constraints of storage and computation. This paper proposes a joint tensor-based CVS reconstruction model FrTVCST, in which a high-order tensor FrTV regularization and a tensor DWT L0 norm regularization are combined. Furthermore, an approach for choosing the regularization parameter that controls the influence of the two terms in the joint sparseness regularization model is proposed and a tensor gradient projection algorithm extended from smoothed L0 (SL0) is derived to solve this combined tensor FrTV and DWT joint regularization constrained minimization problem. Through experiments, we find that compared with several state-of-the-art tensor-

based CVS reconstruction algorithms, FrTVCST is able to deliver a better performance, it has a higher PSNR and a better subjective visual quality within reasonable recovery time.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 61422110 and 61661130158, and in part by the National Ten Thousand Talent Program of China (Young Top-Notch Talent), and in part by the Tsinghua National Laboratory for Information Science (TNList), and in part by the Tsinghua University Initiative Scientific Research Program.

References

[1] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[2] E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509.
[3] H. Jiang, C. Li, R. Haimi-Cohen, P.A. Wilford, Y. Zhang, Scalable video coding using compressive sensing, Bell Labs Tech. J. 16 (4) (2012) 149–169 (2012).
[4] R.F. Marcia, R.M. Willett, Compressive coded aperture video reconstruction, in: Proceedings of the 16th European Signal Processing Conference, EUSIPCO 2008,. Lausanne, Switzerland, August,pp. 1–5, 2008.
[5] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, R. Baraniuk, Single-pixel imaging via compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 83–91.
[6] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takahar, K. Kelly, R.G. Baraniuk, Compressive imaging for video representation and coding, in: Proceeding of the Picture Coding Symposium, Beijing, China, April, pp.13, 2006.
[7] Sungkwang Mun, J.E. Fowler, Residual reconstruction for block-based compressed sensing of video, in: Proceeding of the Data Compression Conference, Snowbird, USA, March. pp. 183–192, 2011.
[8] M. Duarte, R. Baraniuk, Kronecker compressive sensing, IEEE Trans. Image Process. 21 (2) (2012) 494–504.
[9] N. Sidiropoulos, A. Kyrillidis, Multi-way compressed sensing for sparse low-rank tensors, IEEE Signal Process. Lett. 19 (11) (2012) 757–760.
[10] S. Friedland, Q. Li, D. Schonfeld, Compressive sensing of sparse tensors, IEEE Trans. Image Process. 23 (10) (2014) 4438–4447.
[11] C.S. Lu, W.J. Liang, Fast compressive sensing of high dimensional signals with tree-structure sparsity pattern, in: Proceedings of IEEE Signal and Information Processing (ChinaSIP), Xi'an, China, Jul, pp. 738–742, 2014.
[12] M.S. Hosseini, K.N. Plataniotis, High-accuracy total variation with application to compressed video sensing, IEEE Trans. Image Process. 23 (9) (2014) 4438–4447.
[13] C.F. Caiafa, A. Cichocki, Computing sparse representations of multi-dimensional signals using kronecker bases, Neural Comput. J. 25 (1) (2013) 186–220.
[14] C.S. Lu, H.W. Chen, Compressive image sensing for fast recovery from limited samples: a variation on compressive sensing, Inf. Sci. 325 (2015) 33–47.
[15] C.F. Caiafa, A. Cichocki, Stable, robust and super fast reconstruction of tensors using multi-way projections, IEEE Trans. Signal Process. 63 (3) (2015) 780–793.
[16] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D., Nonlinear Phenom. 60 (1–4) (1992) 259–268.
[17] G. Chen, J.S. Zhang, D.F. Li, Fractional-order total variation combined with sparsifying transforms for compressive sensing sparse image reconstruction, J. Vis. Commun. Image Represent. 38 (2016) 407–422.
[18] G. Chen, J.S. Zhang, D.F. Li, H.X. Chen, Robust Kronecker product video denoising based on fractional-order total variation model, Signal Process. 119 (2016) 1–20.
[19] T. Chan, S. Esedoglu, F. Park, A. Yip, Total variation image restoration: overview and recent developments, Handbook of Mathematical Models in Computer Vision, Springer-Verlag, New York, NY, USA, 2006, pp. 17–31.
[20] T. Chan, A. Marquina, P. Mulet, High-order total variation-based image restoration, SIAM J. Sci. Comput. 22 (2) (2000) 503–516.
[21] K. Bredies, K. Kunisch, T. Pock, Total generalized variation, SIAM J. Imag. Sci. 3 (3) (2010) 492–526.
[22] M. Benning, C. Brune, M. Burger, J. Müller, Higher-order TV methods-Enhancement via Bregman iteration, J. Sci. Comput. 54 (2–3) (2013) 269–310.
[23] J. Bai, X.C. Feng, Fractional-order anisotropic diffusion for image denoising, IEEE Trans. Image Process. 16 (10) (2007) 2492–2502.
[24] J. Zhang, Z.H. Wei, A class of fractional-order multi-scale variational models and alternating projection algorithm for image denoising, Appl. Math. Model 35 (5) (2011) 2516–2528.
[25] D. Chen, S. Sun, C. Zhang, Fractional-order TV-L2 model for image denoising, Cent. Eur. J. Phys. 11 (10) (2013) 1414–1422.
[26] D. Tian, D.Y. Xue, D.H. Wang, A fractional-order adaptive regularization primal-dual algorithm for image denoising, Inf. Sci. 296 (2015) 147–159.
[27] J.Z. Huang, F..Yang, Compressed magnetic resonance imaging based on wavelet sparsity and nonlocal total variation, in: Proceedings of the 9th IEEE International Symposium on Biomedical, Imaging (ISBI'09), pp. 968–971.
[28] Y. Liu, I. Gligorijevic, V. Matic, M. De Vos, S. Van Huffel, Multi-sparse signal recovery for compressive sensing, in: Proceedings 2012 Annual International Conference of Engineering in Medicine and Biology Society (EMBC), Aug, pp. 1053–1056, 2012.
[29] Y. Liu, M. De Vos, I. Gligorijevic, V. Matic, Y. Li, S. Van Huffel, Multi-structural signal recovery for biomedical compressive sensing, IEEE Trans. Biomed. Eng. 60 (10) (2013) 2794–2805.
[30] Z.M. Ren, C.J. Hen, Q.F. Zhang, Fractional order total variation regularization for image super-resolution, Signal Process. 93 (2013) 2408–2421.
[31] H. Mohimani, M. Babaie-Zadeh, Ch Jutten, A fast approach for overcomplete sparse decomposition based on smoothed L0 norm, IEEE Trans. Signal Process. 57 (1) (2009) 289–301.
[32] M.M. Hyder, K. Mahata, An improved smoothed L0 approximation algorithm for sparse representation, IEEE Trans. Signal Process. 58 (4) (2010) 2194–2205.
[33] S.B. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61.
[34] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Signal Process. 41 (2) (1993) 3397–3415.
[35] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inform. Theory 53 (12) (2007) 4655–4666.
[36] M.A.T. Figueiredo, R.D. Nowak R D, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problem, J. Sel. Top. Signal Process.: Spec. Issue Convex Optim. Methods Signal Process. 1 (4) (2007) 586–598.
[37] G. Chen, D.F. Li, J.S. Zhang, Iterative Gradient Projection Algorithm for Two-dimensional Compressive Sensing Sparse Image Reconstruction 104 Signal Process, 2014, pp. 15–26.
[38] I. Podlubny, Fractional differential equations, Academic Press, New York, 1999.
[39] E. Berg, M.P. Friedlander, Spot Toolbox, [Online] available at ⟨http://www.cs.ubc.ca/labs/scl/spot⟩.
[40] H.F. Shen, P.X. Li, L.P. Zhang, A MAP algorithm to super-resolution image reconstruction, in: Proceedings of the Third International Conference on Image and Graphics, pp. 544–547, 2004.
[41] B.W. Bader, T.G. Kolda, Tensor Toolbox Version 2.5, Sandia National Laboratories, Jan [Online]. Available at ⟨http://www.sandia.gov/~tgkolda/TensorToolbox⟩, 2012.
[42] E. van den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912.