Efficient Multi-Dimensional Tensor Sparse Coding Using t-linear Combination

Fei Jiang[1], Xiao-Yang Liu[1,2], Hongtao Lu[1], Ruimin Shen[1]
[1]Department of Computer Science and Engineering, Shanghai Jiao Tong Univerisyt, Shanghai, 200240, China
[2]Department of Electrical Engineering, Columbia University, New York, NY10027, USA

Outline

Shortcomings of traditional sparse coding model:

Vector representation => ignore spatial structure => high-dimensional vector => **linear combination** => large-size dictionary => high computational complexity => hard for high-dimensional data application

Advantages of proposed tensor Sparse Coding model:

Tensor Representation => preserve spatial structure => original size tensor => **t-linear combination** => small-size dictionary => low computational complexity => potential application for large-scale MD data

t-linear vs. linear combination

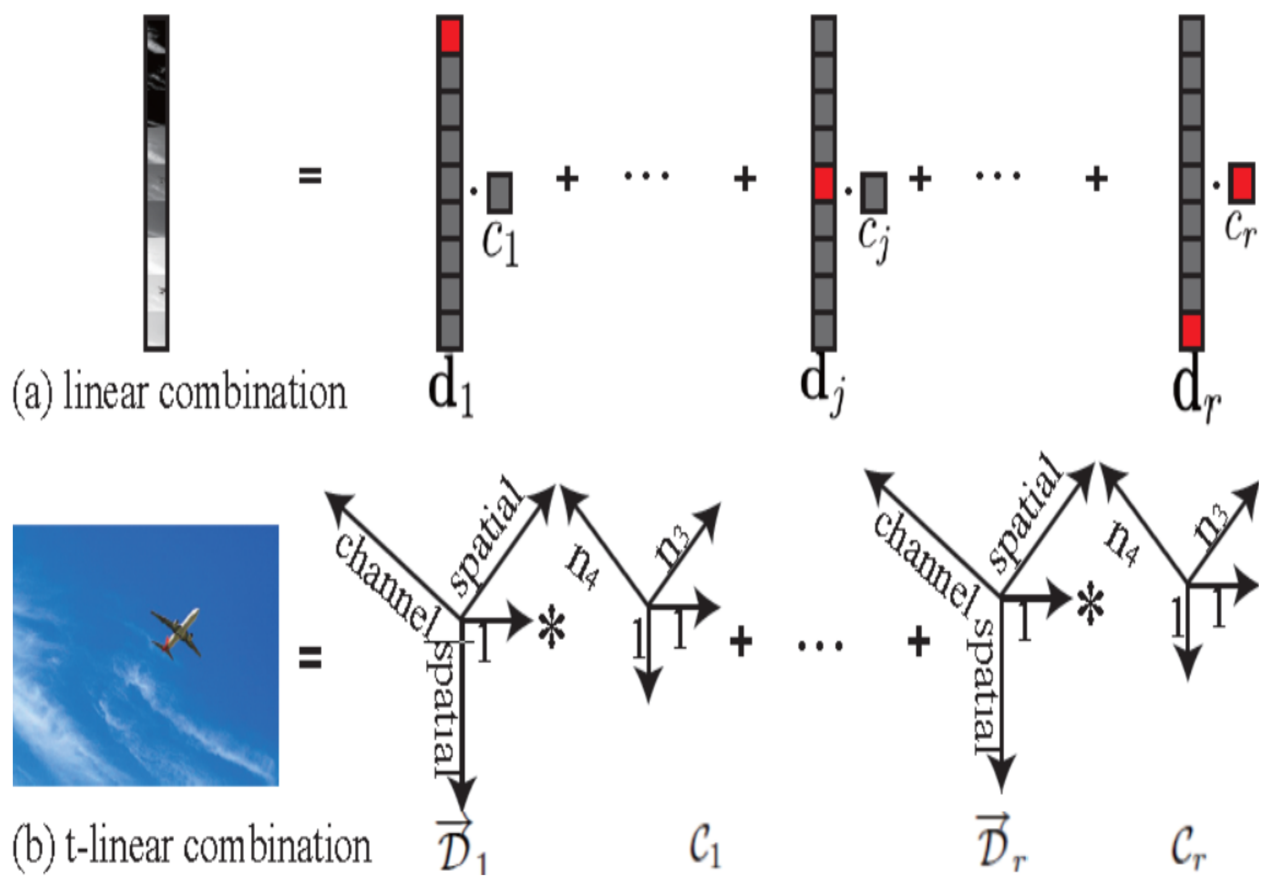


Fig.1 t-linear vs. linear combination. Similar expression but different multiplication operators

Circular multiplication

- Order-3 circular multiplication

$\mathcal{A} \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{B} \in \mathbb{R}^{r \times n_2 \times n_3}$, $\mathcal{C} = \mathcal{A} * \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$

$$\begin{bmatrix} \mathcal{C}^{(1)} \\ \mathcal{C}^{(2)} \\ \vdots \\ \mathcal{C}^{(n_3)} \end{bmatrix} = \begin{bmatrix} \mathcal{A}^{(1)} & \mathcal{A}^{(n_3)} & \cdots & \mathcal{A}^{(2)} \\ \mathcal{A}^{(2)} & \mathcal{A}^{(1)} & \cdots & \mathcal{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{A}^{(n_3)} & \mathcal{A}^{(n_3-1)} & \cdots & \mathcal{A}^{(1)} \end{bmatrix} \cdot \begin{bmatrix} \mathcal{B}^{(1)} \\ \mathcal{B}^{(2)} \\ \vdots \\ \mathcal{B}^{(n_3)} \end{bmatrix}$$

$$\mathrm{unfold}(\mathcal{C}) = \mathrm{circ}(\mathcal{A}) \cdot \mathrm{unfold}(\mathcal{B})$$

- Order-p circular multiplication

$\mathcal{A} \in \mathbb{R}^{n_1 \times r \times n_3 \times \cdots \times n_p}$, $\mathcal{B} \in \mathbb{R}^{r \times n_2 \times n_3 \times \cdots \times n_p}$

$$\mathcal{C} = \mathcal{A} * \mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3 \times \cdots \times n_p}$$

$$\mathrm{unfold}(\mathcal{C}) = \mathrm{circ}(\mathcal{A}) * \mathrm{unfold}(\mathcal{B})$$

- Efficient computation in frequency domain

$$\widehat{\mathcal{C}}(:,:,\ell) = \widehat{\mathcal{A}}(:,:,\ell) \cdot \widehat{\mathcal{B}}(:,:,\ell),\ 1 \le \ell \le n_3 n_4 \cdots n_p$$

Equivalence between t-linear and linear combination

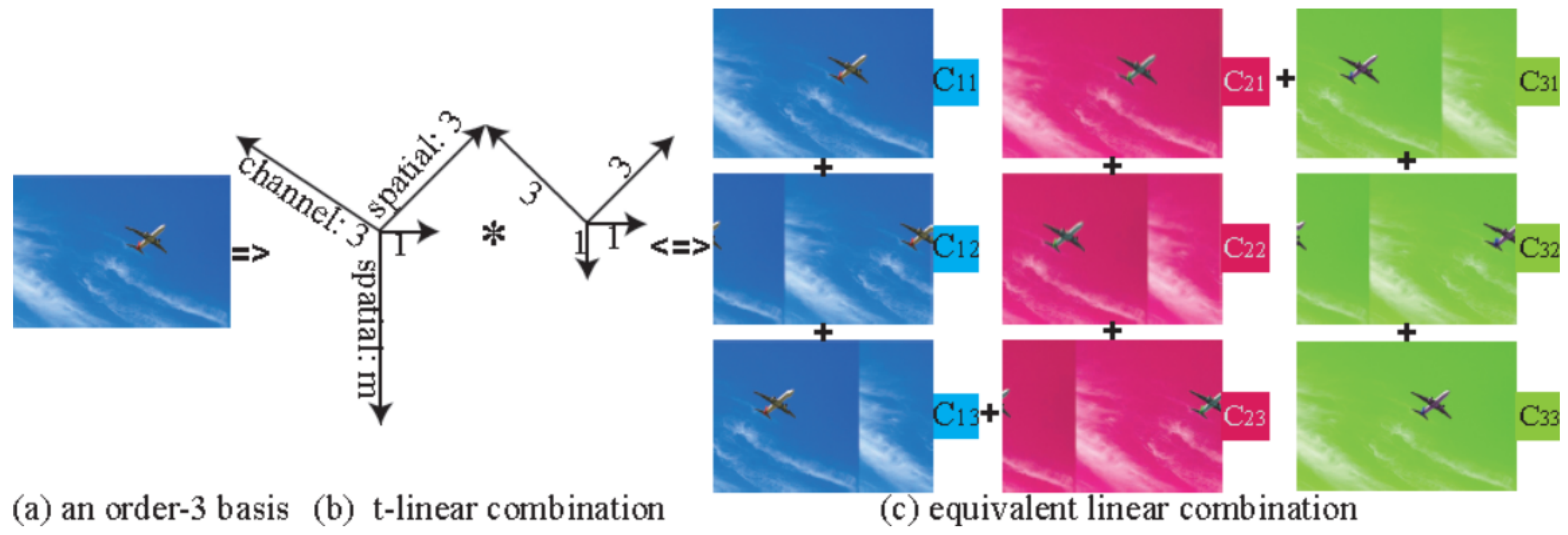


Fig.2 Equivalence between t-linear combination (one basis) and linear combination (9 bases).

- One tensor basis in t-linear combination ⇔ a group of bases in linear combination (small-size dictionary)
- The bases in the group are shifted versions to each other (shifting invariance)
- Rich explanations of coefficients (automatically cluster shifted versions into a group)

Tensor sparse coding model

Given a set of n order-p tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n \times n_2 \times \cdots \times n_p}$, over-completed dictionary $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times n_2 \times \cdots \times n_p}$, $r \ge n_1$ and coefficient $\mathcal{C} \in \mathbb{R}^{r \times n_2 \times \cdots \times n_p}$

$$\begin{aligned} \min_{\mathcal{D},\mathcal{C}} \quad & \tfrac{1}{2}\|\mathcal{X} - \mathcal{D} * \mathcal{C}\|_F^2 + \beta\|\mathcal{C}\|_0 \\ \text{s.t.} \quad & \|\vec{\mathcal{D}}_j\|_F^2 \le 1,\ j \in [r], \end{aligned}$$

Tensor coefficient learning

Tensor-based fast iterative shrinkage thresholding algorithm

$$\begin{aligned} \mathcal{C}^{t+1} &= \arg\min\{f(\mathcal{C}^t) + \langle \nabla f(\mathcal{C}^t), \mathcal{C} - \mathcal{C}^t\rangle + \frac{L}{2}\|\mathcal{C} - \mathcal{C}^t\|_F^2 + \beta g(\mathcal{C})\} \\ &= \mathrm{prox}_{r_t, g}\left(\mathcal{C}^t - r_t \nabla f(\mathcal{C}_t)\right), \end{aligned}$$

Tensor dictionary learning

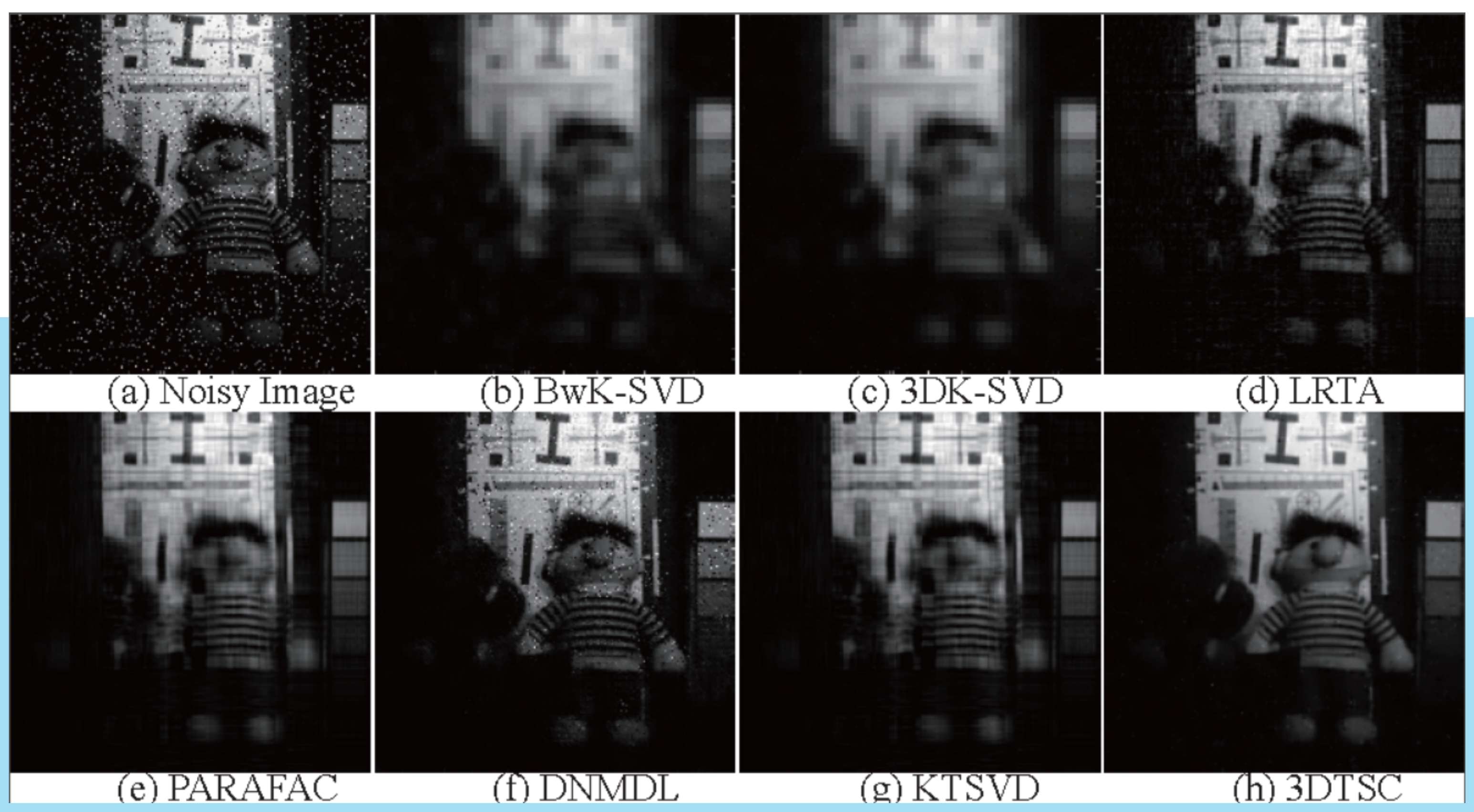
Lagrange-dual algorithm in the frequency domain

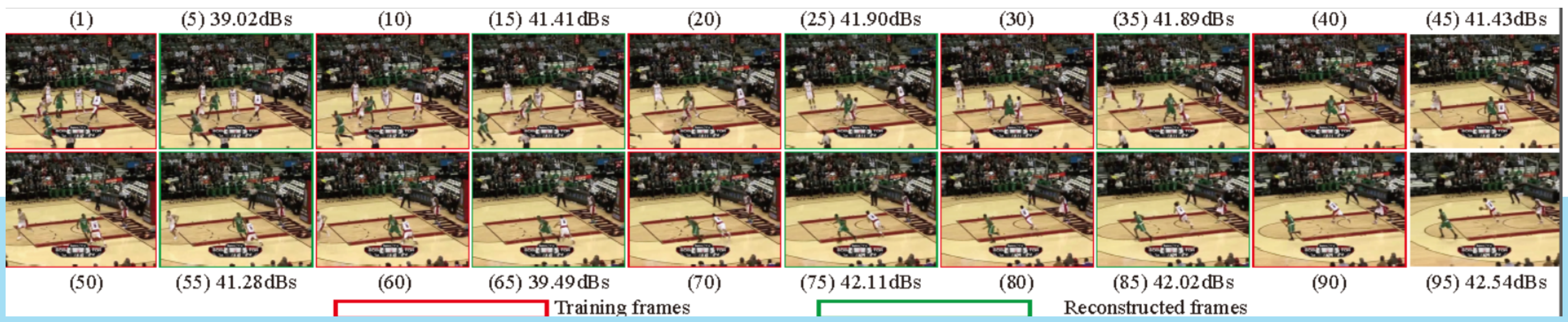
$$\mathcal{L}_{\text{prime}}(\widehat{\mathcal{D}}, \lambda) = \sum_{\ell=1}^{k} \|\widehat{\mathcal{X}}^{(\ell)} - \widehat{\mathcal{D}}^{(\ell)}\widehat{\mathcal{C}}^{(\ell)}\|_F^2 + \sum_{j=1}^{r} \lambda_j \left(\sum_{\ell=1}^{k} \|\widehat{\mathcal{D}}^{(\ell)}(:,j)\|^2 - k\right)$$

$$\mathcal{L}_{\text{dual}}(\lambda) = -\sum_{\ell=1}^{k} \mathrm{Tr}\left(\widehat{\mathcal{D}}^{(\ell)^H} \widehat{\mathcal{X}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)^H}\right) - k\sum_{j=1}^{r} \lambda_j.$$

$$\widehat{\mathcal{D}}^{(\ell)} = \left(\widehat{\mathcal{X}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)^H}\right)\left(\widehat{\mathcal{C}}^{(\ell)} \widehat{\mathcal{C}}^{(\ell)^H} + \mathrm{diag}(\lambda)\right)^{-1}$$

Experimental results (Hyperspectral image denoising & color video reconstruction)





- Compare with recently proposed tensor-based sparse coding algorithms: TenSR(CVPR2016), KTSVD(IJCAI2016), TCSC(ICCV2017), and other tensor-based algorithms: LRTA(TGRSL2008), PARAFAC(TGRS2012), DNMDL(CVPR2014).
- Dataset includes: Columbia MSI dataset (32 scenes, 512x512x31) & basketball video from OTB50 (432x576x3x10).