Tensor Dictionary Learning with representation quantization for Remote Sensing Observation Compression

Anastasia Aidini[1,2], Grigorios Tsagkatakis[1], Panagiotis Tsakalides[1,2]

{aidini, greg, tsakalid}@ics.forth.gr

[1]Institute of Computer Science FORTH, Greece

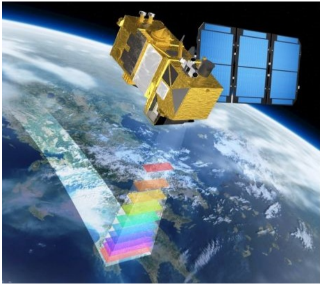
[2]Computer Science Department, UOC

DCC, 2020

Motivation

Remotely sensed images are used for:

- forest monitoring
- disaster evaluation
- land cover estimation



Challenges:

- Increasing spatial, spectral and temporal resolutions of the images
- Increasing storage and transmission requirements
- High dimensional observations modeled as **tensors**
- High spacial, spectral and temporal redundancies

Problem

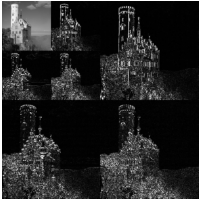
Compression of High-Dimensional Remote Sensing Observations that

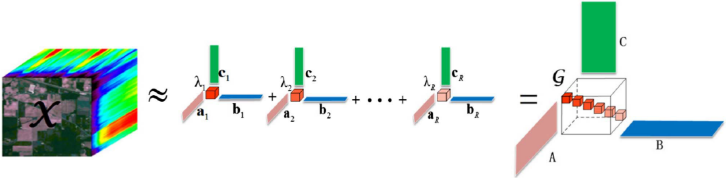
- Includes quantization and coding
- Achieves high compression ratio
- Retains the structure of the data
- Can handle arbitrary high dimensional data structures

Proposed solution:
A novel *tensor dictionary learning* method is used to compress every new sample as a vector of sparse coefficients corresponding to the elements of the learned tensor dictionary, given a set of previous samples.

Related Work

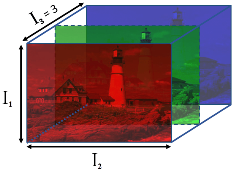
- 3D Wavelet transform on the full data cube.
- A combination of JPEG2000 with Discrete Wavelet Transform or Principal Components Analysis for spectral decorrelation.
- Tensor-based approaches using tensor decompositions by transmitting all the factors of the decomposition.

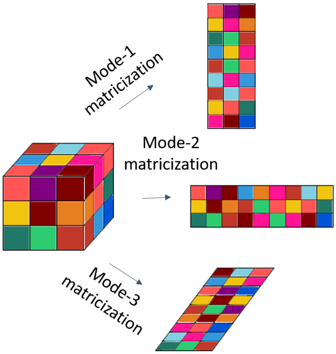




Tensor Based Observation Modeling

A *tensor* $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is a N-way array, a higher-order generalization of vectors and matrices.

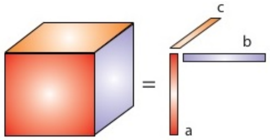




The *mode-n unfolded matrix* $\mathbf{X}_{(n)} \in \mathbb{R}^{I_n \times \prod_{i \neq n} I_i}$ corresponds to a matrix with columns being the vectors obtained by fixing all indices of $\mathcal{X}$ except the n-th index.

Tensor Rank

The outer product of N vectors yields a *rank-1* N-way tensor.



Every tensor can be written as a sum of rank-1 tensors

$$\mathscr{X} \approx \sum_{r=1}^{R} a_r^{(1)} \circ a_r^{(2)} \circ \ldots \circ a_r^{(N)}$$

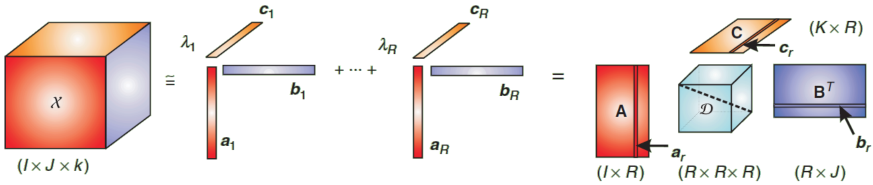
- The *rank* of a N-way tensor $\mathscr{X}$ is the smallest number R of rank-1 tensors needed to synthesize $\mathscr{X}$.
- No straightforward algorithm to determine the rank of a specific given tensor (NP-hard problem).

CP Decomposition

CANDECOMP/PARAFAC (CP) decomposition represents a N-order tensor $\mathscr{X} \in \mathbb{R}^{I_1 \times .. \times I_N}$ as a linear combination of rank-1 tensors in the form

$$\mathscr{X} = \sum_{r=1}^{R} \lambda_r a_r^{(1)} \circ a_r^{(2)} \circ ... \circ a_r^{(N)} = \mathscr{D} \times_1 A^{(1)} \times_2 A^{(2)} \times_3 \cdots \times_N A^{(N)}$$

where $A^{(n)} = [a_1^{(n)} \; a_2^{(n)} \; \cdots \; a_R^{(n)}],\; n = 1, .., N$ are the factor matrices and $\mathscr{D} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_R) \in \mathbb{R}^{R \times \cdots \times R}$ is a diagonal core tensor.



Training Process

1. Learn a *tensor dictionary* $\mathscr{D} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times K}$ of K rank-1 tensors $\mathscr{D}^{(k)} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, $k = 1, .., K$, by minimizing

$$\begin{array}{c} \min_{\mathscr{D}, \mathbf{A}} \frac{1}{2} \|\mathscr{X} - \mathscr{D} \times_{N+1} \mathbf{A}\|_F^2 \\ \text{subject to } \|\mathbf{A}(t,:)\|_0 \leq \lambda, \ \forall t = 1, .., T \end{array},$$

   where $\mathscr{X} = (\mathscr{X}^1, \mathscr{X}^2, ..., \mathscr{X}^T) \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N \times T}$ is a set of T training samples and $\mathbf{A} \in \mathbb{R}^{T \times K}$ contains the corresponding sparse coefficients.

2. Learn a *coding dictionary* which maps a binary number to a set of 2^{bit} symbols by equally partitioning the range of values of the coefficient matrix $\mathbf{A}$.

Proposed Tensor Dictionary Learning Method

Introducing the auxiliary variable $\mathbf{G} \in \mathbb{R}^{T \times K}$, we apply the Alternating Direction Method of Multipliers to solve the reformulated problem

$$\min_{\mathscr{D},\mathbf{A},\mathbf{G}} \frac{1}{2}\|\mathscr{X} - \mathscr{D} \times_{N+1} \mathbf{A}\|_F^2 + \lambda \sum_{t=1}^{T} \|\mathbf{G}(t,:)\|_0$$
$$\text{subject to } \mathbf{G} = \mathbf{A}$$

The augmented Lagrangian function is given by

$$\mathscr{L}(\mathbf{A},\mathbf{G},\mathscr{D},\mathbf{Y}) = \frac{1}{2}\|\mathscr{X} - \mathscr{D} \times_{N+1} \mathbf{A}\|_F^2 + \lambda \sum_{t=1}^{T} \|\mathbf{G}(t,:)\|_0 + < \mathbf{Y}, \mathbf{G} - \mathbf{A} > + \frac{p}{2}\|\mathbf{G} - \mathbf{A}\|_F^2,$$

where $\mathbf{Y} \in \mathbb{R}^{T \times K}$ is the Lagrange multiplier matrix and $p > 0$ denotes the step size parameter.

Tensor Dictionary Learning Algorithm

We solve the problem iteratively by minimizing $\mathscr{L}$ with respect to each variable while keeping the others fixed. At each iteration l we update:

- Sparse coefficient matrix $\mathbf{A}$:

$$\nabla_{\mathbf{A}}\mathscr{L} = 0 \quad \Rightarrow \quad \mathbf{A} \leftarrow (\mathbf{X}_{(N+1)} \cdot \mathbf{D}_{(N+1)}^{T} + \mathbf{Y} + p \cdot \mathbf{G}) \cdot (\mathbf{D}_{(N+1)} \cdot \mathbf{D}_{(N+1)}^{T} + p \cdot \mathbf{I})^{-1}$$

- Auxiliary variable $\mathbf{G}$:

$$\nabla_{\mathbf{G}}\mathscr{L} = 0 \quad \Rightarrow \quad \mathbf{G} \leftarrow H_{\lambda}(\mathbf{A} - \frac{\mathbf{Y}}{p}), \text{ where } H_{\lambda}(x) = \begin{cases} x, & |x| > \lambda \\ 0, & \text{otherwise} \end{cases}$$

- Tensor dictionary $\mathscr{D}$:

$$\nabla_{\mathscr{D}}\mathscr{L} = 0 \quad \Rightarrow \quad \mathscr{D}^{(l)} \leftarrow \mathscr{D}^{(l-1)} + \mathscr{X} \times_{N+1} \mathbf{A}^{-1} \text{ and normalization}$$

- Lagrangian multiplier matrix $\mathbf{Y}$:

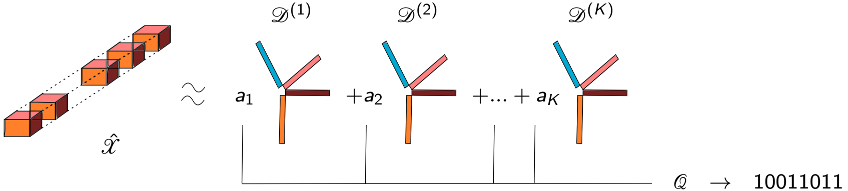
$$\mathbf{Y}^{(l)} \leftarrow \mathbf{Y}^{(l-1)} + p \cdot (\mathbf{G} - \mathbf{A}) \text{ where } p = 0.6 \text{ in our setup}$$

Compression

1. Compress each new sample $\hat{\mathscr{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ as a sparse vector of coefficients $\mathbf{a} \in \mathbb{R}^K$ such that

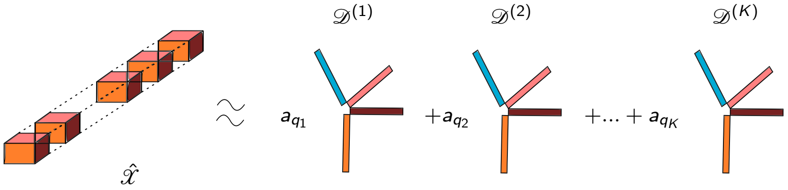
$$\hat{\mathscr{X}} = \mathscr{D} \times_{N+1} \mathbf{a}, \text{ where } \|\mathbf{a}\|_0 \leq \lambda,$$

2. Quantize $\mathbf{a}$ to b bits using a uniform quantizer $\mathscr{Q}$.
3. Encode $\mathscr{Q}(\mathbf{a})$ using Huffman coding and the learned encoding dictionary.



Decompression

1. Decode the transmitted coefficients $\mathbf{a}_q = \mathcal{Q}(\mathbf{a})$ using the learned Huffman dictionary.
2. Decompress the sample as $\hat{\mathcal{X}} \approx \mathcal{D} \times_{N+1} \mathbf{a}_q$.



Experiments

- Data: Time series of satellite derived observations of normalized difference vegetation index (NDVI).
- Sample size: $200 \times 200 \times 7$
  The last dimension indicates the number of days.
- Training samples: 50
- The recovery performance is measured in terms of the Normalized Mean Square Error (NMSE) which is defined as $\mathrm{NMSE} = \frac{\|\mathscr{Y} - \hat{\mathscr{Y}}\|_2^2}{\|\mathscr{Y}\|_2^2}$.

Number of Atoms of the Dictionary

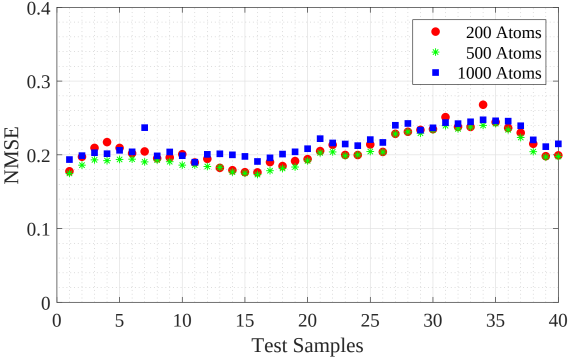


Figure: Reconstruction quality for each test sample and different number of atoms of the dictionary, using 80% sparsity level and 8 bits of quantization.

Sparsity Level

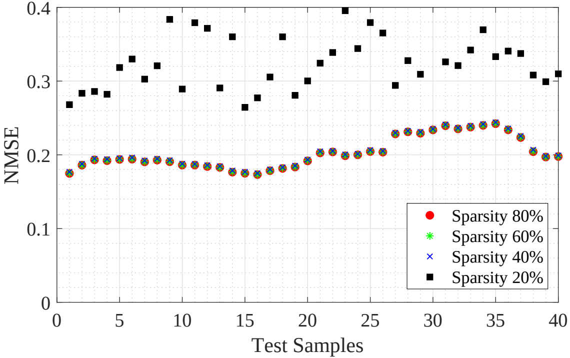


Figure: Reconstruction quality for each test sample and different sparsity levels, using a dictionary with 500 atoms and 8 bits of quantization.

Number of Quantization Bits

Table: Reconstruction error for different samples as a function of quantization bit number.

Number of Bits	NMSE			
	1st Sample	10th Sample	25th Sample	35th Sample
4	0.2915	0.3226	0.3510	0.3936
6	0.1816	0.1941	0.2134	0.2510
8	0.1748	0.1860	0.2044	0.2425

Comparison with State-of-the-art Compression Algorithm

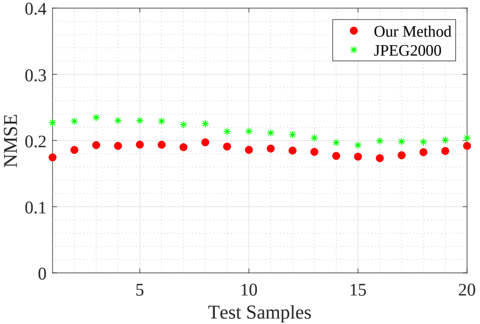


Figure: Reconstruction quality for several test samples, using 0.06 bpppb.

Table: Reconstruction quality for different number of bpppb.

NMSE	bpppb			
	0.20	0.16	0.08	0.03
Our Method	0.1754	0.1751	0.1748	0.1777
JPEG2000	0.1838	0.1929	0.2169	0.2493

Conclusion

- An end-to-end compression algorithm is proposed that includes quantization and coding.
- A novel tensor dictionary learning method based on CP decomposition is presented for compression purposes.
- The proposed scheme can handle arbitrary high dimensions.
- Our method is evaluated on 3D remote sensing observations.

Acknowledgments

This research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under HFRI PhD Fellowship grant no. 1509 and HFRI Faculty grant no. 1725.