IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 8, 2019, accepted June 21, 2019, date of publication June 27, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2925535

Multiscale Sparse Dictionary Learning With Rate Constraint for Seismic Data Compression

XIN TIAN, (Member, IEEE)
Electronic Information School, Wuhan University, Wuhan 430072, China
e-mail: xin.tian@whu.edu.cn

This work was supported in part by the Commission on Innovation and Technology of Shenzhen through the Basic Research (Free Exploration) Project under Grant JCYJ20170303170611174, in part by the Natural Science Foundation of Hubei Province under Grant 2018CFB435, and in part by the Fundamental Research Funds for the Central Universities under Grant 2042018kf1009.

ABSTRACT Seismic survey is one of the most effective tools for oil and gas exploration. To date, there has been an exponential growth in the size of seismic data required for large-scale seismic survey. For transmission and storage purposes, we propose a novel seismic compression method. First, a multiscale sparse dictionary learning model with rate constraint is presented. By combining the advantages of multiscale decomposition and dictionary learning, the seismic data could be effectively represented as a sparse matrix. Rate constraints are used to obtain the sparse coefficients that are properly tailored to the compression objective. To solve the optimization problem, the alternating direction method of multipliers is adopted. Furthermore, a seismic compression scheme based on the learned dictionary is introduced. Finally, public seismic datasets are used to verify the efficiency of different seismic data compression methods. The experimental results indicate that the proposed method achieves the best seismic compression performance, including rate-distortion tradeoff and visual quality.

INDEX TERMS Multiscale dictionary learning, rate constraint, sparse coding, seismic data compression.

I. INTRODUCTION

A seismic survey, which collects a significant amount of seismic data from a field, is a primary strategic tool that is utilized in oil and gas exploration [1]. A typical seismic survey may generate tens of terabytes of raw seismic data on a daily basis [2]. This results in extreme challenges in terms of wireless gathering of the seismic data from the field, given that the bandwidth of a wireless system is inherently limited. Further, once collected, this massive amount of data requires very large storage capacities at data centers. Therefore, data compression is highly desirable to reduce the cost of storage and transmission, especially in the case of wireless seismic signal acquisition. Typical seismic data compression methods can be grouped into two categories: lossless and lossy. Lossless compression [3] is a kind of data compression algorithm that allows the original data to be perfectly reconstructed from the compressed data but results in limited data reduction. This may not be suitable for the compression of a massive amount of seismic data, especially when the seismic signal samples are represented by float or double numbers. Lossy compression methods [4] allow some acceptable information losses but can achieve much higher compression performance. As such, this investigation is based on the lossy scheme as a potential candidate.

Transform-based lossy compression techniques have been the most prevalent approach for many years. The discrete cosine transform (DCT), used for seismic signal compression in [5], achieves a lossy compression gain of approximately three. A two-dimensional seismic adaptive local cosine transform was proposed in [6] to preserve important features. In [7], wavelet packets were used to achieve a higher compression rate and simulation results did not reveal visible artifacts in the reconstructed data. Further, curvelet, which has strong anisotropic selectivity, was applied to explore the directional features of the seismic data [8]. Instead of using off-the-shelf transforms (e.g., DCT, wavelet, and curvelet), the process of learning a data-driven dictionary from the original data, has been shown to be promising in data compression. The authors in [9] reported good image compression performance using recursive least squares dictionary learning algorithm. Furthermore, a compressibility constrained sparse coding formulation is proposed in [10] where low bit-rate image compression was achieved based on the sparsity and

The associate editor coordinating the review of this manuscript and approving it for publication was Chengpeng Hao.

compressibility of the sparse coefficients. In [11], the authors showed that using a double sparsity model to learn a dictionary gives much better compression results for remote sensing images. All these works demonstrate the potential of learned dictionaries for the successful compression of seismic data. By extracting the information from the common memory between the sender and receiver, a memory-assisted seismic signal compression method based on dictionary learning was proposed in [12]. However, online memory will increase the hardware cost, which limits the scope. To explore the resemblance among local seismic traces to facilitate compression for communication, seismic data compression using online double-sparse dictionary learning (SIODL) is proposed. Sparse constraints and a sliding window mechanism are applied to the incremental components of the dictionaries, which improves their compression performance [13]. Recent advances in seismic data compression methods based on dictionary learning have resulted in significant improvement in compression gain [14].

Although sparse dictionary learning can facilitate more efficient representation than a fixed dictionary [15], its performance in data compression is sometimes limited. The optimization function in sparse dictionary learning includes two terms: data fidelity and sparsity constrained. Therefore, the objective is to learn a sparse matrix of coefficients. However, this may not be optimum for compression. To address this issue, a rate-distortion optimal solution for an over-complete signal representation was proposed in [16], [17]. However, each sample is processed independently, which ignores the constraint over the entire input data samples. Thus, the compression performance decreases. Moreover, the dictionary learning algorithm cannot be easily integrated. To design dictionaries that are adapted to the available data resulting in good compression performance, a dictionary learning algorithm for efficient signal compression is introduced in [18]. By minimizing the total rate rather than the sparsity of the coefficients, it outperforms other traditional seismic compression methods. However, the compression performance gap between [18] and other methods (such as DCT) will be small when the measurement noise is included. One reason for this being the fact that its sparsity is not considered in the entire dictionary learning process. This influences its robustness to noise. Delay compensated and entropy constrained dictionary learning (DCECDL) was proposed for seismic data compression in [19]. In this case, the main objective is the reduction of the entropy of the quantized coefficients. Given that this optimization problem is difficult to solve, a searching strategy is adopted to find the solution. Therefore, its performance is highly dependent on the initial point chosen. Although previous dictionary learning based seismic compression methods have demonstrated their individual efficacy, they each have limitations. For example, the initial values may significantly affect the compression performance because the optimization problem in most algorithms is non-convex. This introduces notable instability into the algorithms. It should be noted that multiresolution analysis (MRA) is a powerful tool for data compression. Therefore, multiscale dictionary learning algorithms [20], [21] could potentially be exploited to improve compression efficiency.

To address the aforementioned problems, a multiscale, rate-constrained dictionary learning algorithm for seismic data compression is proposed. The main contributions of this study are as follows: We propose a novel multiscale sparse dictionary learning model with rate constraint. In this model, the training samples are built from the multiscale decomposition like wavelets. Then the dictionary learning problem can be formulated as several sub-dictionary learning problems based on the training samples from the corresponding scales. In this study, dictionary learning in multiscale will be helpful for capturing the characteristics of seismic data and providing more degrees of freedom for compressing the sparse coefficients. Furthermore, not only the sparsity but also the rate is included in the optimization model as prior constraints, the target of which is to make the sparse coefficients more suitable for compression purpose. More specially, the rate is approximately estimated by the coefficients' probability density function, and this function is supposed to be a Gaussian mixture model. Thus, the proposed optimization model is solved, and the alternating direction method of multipliers is adopted in this paper. Finally, a seismic compression scheme based on the multiscale and rate-constrained dictionary learning algorithm is suggested.

The report is divided into multiple sections. Section II presents the multiscale sparse dictionary learning method with rate constraint for seismic data compression. The experimental results are presented in Section III and Section IV summarizes the main conclusions of this investigation. For convenience, some typical notations used in this report are summarized in Table 1.

TABLE 1. Main notations used in this paper.

Notation	Meaning
Lowercase letter	A scalar
Bold lowercase letter	A vector
Bold capital letter	A matrix
$\|\mathbf{W}\|_0$	l_0-norm, Number of nonzero elements in $\mathbf{W}$
$\|\mathbf{W}\|_1$	l_1-norm, $\sum_{i,j} \|\mathbf{W}_{i,j}\|$
$\|\mathbf{W}\|_F$	Frobenious norm, $\sqrt{\sum_{i,j} \mathbf{W}_{i,j}^2}$
$\langle \mathbf{R}, \mathbf{H} \rangle$	Inner product of matrix
$\mathbf{W}^{(b)}$	A matrix $\mathbf{W}$ of the subband b
$\mathbf{W}(i,j)$	The i th and j th element of matrix $\mathbf{W}$
$\mathbf{W}_{(k)}$	A calculated matrix $\mathbf{W}$ in the k iteration
$max(\mathbf{W}, 0)$	A non-negative matrix by changing the negative values of the matrix $\mathbf{W}$ to be zero.
$sign(\mathbf{W})$	A sign matrix of $\mathbf{W}$
$\mid \mathbf{W} \mid$	An absolute matrix by calculating the absolute value of each element in $\mathbf{W}$
$\|\mathbf{w}\|_2$	Euclidean norm of vector, $\sqrt{\sum_i \mathbf{w}_i^2}$
$\mathbf{w}_i$	The i th column vector of matrix $\mathbf{W}$

II. PROPOSED METHOD

A. DICTIONARY LEARNING AND SPARSE CODING

Sparse dictionary learning [22] is a method that aims to identify an adaptive basis (called a dictionary) for a dataset such that each data sample in the dataset can be efficiently estimated by a sparse linear combination of atoms in the learned dictionary. Compared to the off-the-shelf transform, a data-driven dictionary learned directly from the dataset is typically superior in terms of exploring the data regularities and results in efficient sparse coding. It is well-known that the ℓ_0 norm regularized minimization problem [23] in dictionary learning is combinatorial and non-convex, which makes the sparse coding an NP-hard problem. A popular solution to this issue is either to replace the ℓ_0 norm with the ℓ_1 norm [24], which is convex and sparse promoting, or to apply greedy algorithms to obtain approximate solutions for the ℓ_0 norm regularized problem.

During the implementation of dictionary learning, the original data are usually divided into small segments (patches) to reduce the computational burden. The size of the segment (patch) is empirical to capture features of interest. Given the input data set $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_K] \in \mathbb{R}^{m\times K}$ that contains K data segments (patches), dictionary learning techniques are solved for a normalized dictionary $\mathbf{D} \in \mathbb{R}^{m\times n}$ and sparse coefficients of matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_i, \ldots, \mathbf{w}_K] \in \mathbb{R}^{n\times K}$ such that the overall error $\|\mathbf{X}-\mathbf{DW}\|_F^2$ is minimized within a given sparsity level L. This could be formulated as

$$\underset{\mathbf{D},\mathbf{W}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{DW}\|_F^2, \quad \text{s.t. } \|\mathbf{w}_i\|_0 \le L, \quad \text{for } i \in [1, K], \quad (1)$$

where the $\|\mathbf{w}_i\|_0$ counts the number of nonzero elements in $\mathbf{w}_i$. The optimization problem in (1) is typically solved using an alternating procedure involving two stages: sparse coding and dictionary updating. In sparse coding [25], sparse coefficients $\mathbf{w}_i$ are identified given a fixed dictionary $\mathbf{D}$. Some common sparse coding algorithms include orthogonal matching pursuit (OMP), order recursive matching pursuit (ORMP) [26], [27], and partial search (PS) [28]. Additionally, the dictionary updating stage modifies $\mathbf{D}$ for a given $\mathbf{W}$ to further decrease the overall error. With respect to dictionary updating, multiple methods are commonly used such as the method of optimal directions (MOD) [29] and K-SVD [30]. In most dictionary learning algorithms, the objective is to learn a sparse matrix of coefficients. However, this may not be optimum for compression.

B. MULTISCALE SPARSE DICTIONARY LEARNING MODEL WITH RATE CONSTRAINT

For compression, conversion of the raw data into the transform domain is typical [31], [32]. Subsequently, the coefficients tend to be more correlated. The typical transform method used in compression is the multiscale wavelet transform in which the input data could be decomposed into multiscale subbands. In Figure 1, we transform a seismic wave using a multiscale wavelet with a scale of 2. In different subbands, we find high structure similarity across directions. In particular, the lower subbands hold the most significant coefficients that are correlated to the directions of the original data but have a smaller subband. Thus, an approximate version of original seismic data could be reconstructed based only on the coefficients from the lowest subbands. This inspired us to assign different rates for different subbands depending on their importance in the dictionary learning process.

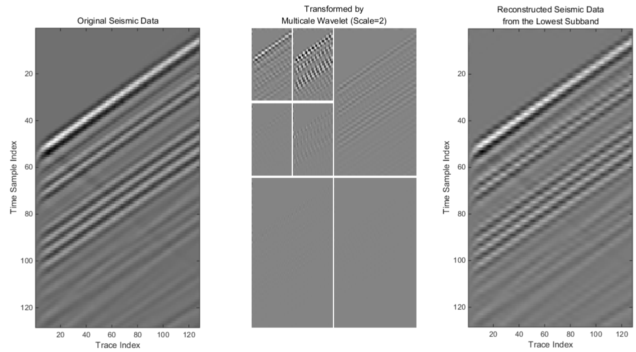


FIGURE 1. Reconstruction of seismic data based on multiscale wavelet.

Let $\mathbf{X}$ denote the original seismic data and $\mathbf{\Phi}$ correspond to a multiscale orthogonal transform, such as the wavelet transform that is primarily used in this study. $\mathbf{\Phi}^+$ is the analysis operator of $\mathbf{\Phi}$. Then, the multiscale transform results are $\mathbf{Z}^{(b)} = (\mathbf{\Phi}^+\mathbf{X})^{(b)}$, where b is the subband index. If the total decomposition level is S, then we have $b = 1, \ldots, 3S + 1$. In dictionary learning algorithms, it is common to split the input data into small data patches [33]. This is helpful in reducing the computational complexity generated by the large training datasets. In our work, each subband data is divided into nonoverlapping patches with the same size. Furthermore, the mean values are removed from these patches for dictionary learning. The results are then organized in vectorized form as $\mathbf{Y}^{(b)} = [\mathbf{y}_1^{(b)}, \ldots, \mathbf{y}_i^{(b)}, \ldots, \mathbf{y}_L^{(b)}], i \in [1, L]$. L is the number of patches in subband b. We, therefore, focus on the following optimization problem solution for dictionary $\mathbf{D}^{(b)}$ and sparse coefficients $\mathbf{W}^{(b)}$ of each subband ($b \in [1, 3S + 1]$):

$$\forall b: \ [\tilde{\mathbf{D}}^{(b)}, \tilde{\mathbf{W}}^{(b)}] = \underset{\mathbf{D}^{(b)},\mathbf{W}^{(b)}}{\operatorname{argmin}} \|\mathbf{Y}^{(b)} - \mathbf{D}^{(b)}\mathbf{W}^{(b)}\|_F^2 + \lambda^{(b)} R(\mathbf{W}^{(b)}) + \alpha^{(b)}\|\mathbf{W}^{(b)}\|_0, \quad (2)$$

where $\|\mathbf{W}^{(b)}\|_0$ is the l_0 norm of the coefficients $\mathbf{W}^{(b)}$. This constraint is important to ensure that the coefficients are as sparse as possible with the distortion controlled by the error between $\mathbf{Y}^{(b)}$ and $\mathbf{D}^{(b)}\mathbf{W}^{(b)}$. $\tilde{\mathbf{D}}^{(b)}$ is the learned dictionary of subband b, and $\tilde{\mathbf{W}}^{(b)}$ are the sparse coefficients of $\mathbf{Y}^{(b)}$ based on the learned dictionary $\tilde{\mathbf{D}}^{(b)}$. $\lambda^{(b)}$ is a parameter used to balance the importance between the reconstruction error and the rate. $\lambda^{(b)}$ is different in different subbands depending on their importance to the reconstruction. For example, the lower subbands could have smaller $\lambda^{(b)}$ value, and $\lambda^{(b)}$ for the higher subbands could be larger. This will be helpful in identifying a balance between the rate and distortion. $\alpha^{(b)}$ is a regularization parameter for the sparsity of the coefficients. $R(\mathbf{W}^{(b)})$ represents the coding rates for the sparse coefficients, which is defined as

$$R(\mathbf{W}^{(b)}) = \sum_{i=1}^{M}\sum_{j=1}^{N} Rate(\mathbf{W}^{(b)}(i,j)). \tag{3}$$

M and N are the number of row and column of the matrix $\mathbf{W}^{(b)}$, respectively. $Rate(\mathbf{W}^{(b)}(i,j))$ is the rate consumed for the coefficient $\mathbf{W}^{(b)}(i,j)$. For compression, the coefficients $\mathbf{W}^{(b)}(i,j)$ are classified into zero elements and nonzero elements. Only the nonzero elements are coded and transmitted. Thus, the rate $Rate(0)$ for the zero elements is approximately 0. According to Shannon's information theory, the rate $Rate(\mathbf{W}^{(b)}_{\neq 0}(i,j))$ of the nonzero elements $\mathbf{W}^{(b)}_{\neq 0}(i,j)$ is computed as

$$Rate(\mathbf{W}^{(b)}_{\neq 0}(i,j)) = -log_2 p(\mathbf{W}^{(b)}_{\neq 0}(i,j)). \tag{4}$$

$p(\mathbf{W}^{(b)}_{\neq 0}(i,j))$ is the probability density function of nonzero elements $\mathbf{W}^{(b)}_{\neq 0}$ $(b \in [1, 3S+1])$. In particular, this probability density function could be approximated by a Gaussian mixture model with the parameters μ and σ^2, and described as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{5}$$

x denotes the sparse coefficient $\mathbf{W}^{(b)}_{\neq 0}(i,j)$. This Gaussian model approximation could be verified based on the following experiment. The histogram of the dictionary-learned sparse coefficients from one seismic data is shown in Figure 2, which demonstrates that our assumption of a Gaussian model for the probability density function of the sparse coefficients is substantially correct.

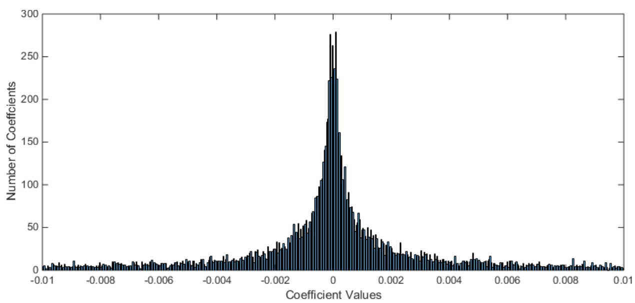


FIGURE 2. Histogram of sparse coefficients from one seismic data.

Then, we reformulate the original optimization problem (2) as

$$\forall b: [\tilde{\mathbf{D}}^{(b)}, \tilde{\mathbf{W}}^{(b)}, \tilde{\mu}, \tilde{\sigma}^2] = \underset{\mathbf{D}^{(b)},\mathbf{W}^{(b)},\mu,\sigma}{\operatorname{argmin}} \|\mathbf{Y}^{(b)} - \mathbf{D}^{(b)}\mathbf{W}^{(b)}\|_F^2 + \lambda^{(b)}R(\mathbf{W}^{(b)}) + \alpha^{(b)}\|\mathbf{W}^{(b)}\|_1, \tag{6}$$

where $\tilde{\mu}$ and $\tilde{\sigma}^2$ are the learned Gaussian model parameters. In (6), the l_0 norm minimization, which is a NP-hard problem, is relaxed to the l_1 norm for optimization purposes. This relaxation has been adopted in many sparse coding algorithms [24], [34].

C. SOLUTION OF THE OPTIMIZATION

To solve the preceding complicated optimization problem, (6) is decoupled into the following subproblems using the alter-minimization method [35]. Then, each variable is updated alternately as:

(1) Rate-constrained Sparse Coding:

In this subproblem, we will solve $\tilde{\mathbf{W}}^{(b)}$,$\tilde{\mu}$,and $\tilde{\sigma}^2$ with fixed $\tilde{\mathbf{D}}^{(b)}$ using the following equation:

$$\forall b: [\tilde{\mathbf{W}}^{(b)}, \tilde{\mu}, \tilde{\sigma}^2] = \underset{\mathbf{W}^{(b)},\mu,\sigma}{\operatorname{argmin}} \|\mathbf{Y}^{(b)} - \tilde{\mathbf{D}}^{(b)}\mathbf{W}^{(b)}\|_F^2 + \lambda^{(b)}R(\mathbf{W}^{(b)}) + \alpha^{(b)}\|\mathbf{W}^{(b)}\|_1. \tag{7}$$

To solve this optimization problem, the alternating direction method of multipliers (ADMM) [36], [37] is used. Using two introduced auxiliary variables $\mathbf{P}$ and $\mathbf{S}$, (7) is rewritten as

$$\forall b: [\tilde{\mathbf{W}}^{(b)}, \tilde{\mu}, \tilde{\sigma}^2] = \underset{\mathbf{W}^{(b)},\mu,\sigma}{\operatorname{argmin}} \|\mathbf{Y}^{(b)} - \tilde{\mathbf{D}}^{(b)}\mathbf{P}\|_F^2 + \lambda^{(b)}R(\mathbf{S}) + \alpha^{(b)}\|\mathbf{W}^{(b)}\|_1, \quad s.t.\ \mathbf{W}^{(b)} = \mathbf{P},\ \mathbf{W}^{(b)} = \mathbf{S}. \tag{8}$$

The augmented Lagrangian form of (8) is

$$\begin{aligned} L(\mathbf{W}^{(b)}, \mathbf{P}, \mathbf{S}, \mathbf{H}0, \mathbf{H}1, \mu, \sigma) &= \|\mathbf{Y}^{(b)} - \tilde{\mathbf{D}}^{(b)}\mathbf{P}\|_F^2 + \lambda^{(b)}R(\mathbf{S}) + \alpha^{(b)}\|\mathbf{W}^{(b)}\|_1 \\ &+ <\mathbf{W}^{(b)} - \mathbf{P}, \mathbf{H}0> + \frac{\rho}{2}\|\mathbf{W}^{(b)} - \mathbf{P}\|_F^2 \\ &+ <\mathbf{W}^{(b)} - \mathbf{S}, \mathbf{H}1> + \frac{\tau}{2}\|\mathbf{W}^{(b)} - \mathbf{S}\|_F^2, \end{aligned} \tag{9}$$

where ρ and τ are regularization parameters associated with the quadratic penalty function of terms $\|\mathbf{W}^{(b)} - \mathbf{P}\|_F^2$ and $\|\mathbf{W}^{(b)} - \mathbf{S}\|_F^2$. $\mathbf{H}0$ and $\mathbf{H}1$ are the Lagrange multipliers associated with the constraints $\mathbf{W}^{(b)} = \mathbf{P}$ and $\mathbf{W}^{(b)} = \mathbf{S}$. The main idea of the augmented Lagrangian method is to find a saddle point of $L(\mathbf{W}^{(b)}, \mathbf{P}, \mathbf{S}, \mathbf{H}0, \mathbf{H}1)$, which is also the solution of (8). Then, $\mathbf{W}^{(b)}$, $\mathbf{P}$, $\mathbf{S}$, $\mathbf{H}0$ and $\mathbf{H}1$ can be solved alternately. We now investigate these subproblems individually.

a) $\mathbf{W}^{(b)}$-Subproblem:

This is equal to the following:

$$\mathbf{W}^{(b)}_{(k+1)} = \underset{\mathbf{W}^{(b)}}{\operatorname{argmin}}\, \alpha^{(b)}\|\mathbf{W}^{(b)}\|_1 + \frac{\rho}{2}\|\mathbf{W}^{(b)} - \mathbf{P}_{(k)} + \mathbf{H}0_{(k)}/\rho\|_F^2 + \frac{\tau}{2}\|\mathbf{W}^{(b)} - \mathbf{S}_{(k)} + \mathbf{H}1_{(k)}/\tau\|_F^2. \tag{10}$$

A shrinkage method could be used to solve (10) as

$$\mathbf{W}^{(b)}_{(k+1)} = max\{|\mathbf{U}| - \alpha^{(b)}/(\rho+\tau), 0\} \cdot sign(\mathbf{U}), \tag{11}$$

where $\mathbf{U} = (\rho\mathbf{P}_{(k)} + \tau\mathbf{S}_{(k)} - \mathbf{H}0_{(k)} - \mathbf{H}1_{(k)})/(\rho+\tau)$.

b) $\mathbf{P}$-Subproblem:

This can be written as follows:

$$\mathbf{P}_{(k+1)} = \underset{\mathbf{P}}{\operatorname{argmin}} \|\mathbf{Y}^{(b)} - \tilde{\mathbf{D}}^{(b)}\mathbf{P}\|_F^2 + \frac{\rho}{2}\|\mathbf{W}^{(b)}_{(k+1)} - \mathbf{P} + \mathbf{H}0_{(k)}/\rho\|_F^2. \tag{12}$$

This generalized problem has an optimal solution that can be solved explicitly using the following formula:

$$\mathbf{P}_{(k+1)} = (\tilde{\mathbf{D}}^{(b)^T}\tilde{\mathbf{D}}^{(b)} + \rho\mathbf{I})^{-1} \times (\tilde{\mathbf{D}}^{(b)^T}\mathbf{Y}^{(b)} + \rho\mathbf{W}^{(b)}_{(k+1)} + \mathbf{H}0_{(k)}). \quad (13)$$

$\tilde{\mathbf{D}}^{(b)^T}$ is the transpose matrix of $\tilde{\mathbf{D}}^{(b)}$, and I is the identity matrix.

c) **S**-Subproblem:

We have:

$$\mathbf{S}_{(k+1)} = \underset{\mathbf{S}}{\operatorname{argmin}}\, \lambda^{(b)}R(\mathbf{S}) + \frac{\tau}{2}\|\mathbf{W}^{(b)}_{(k+1)} - \mathbf{S} + \mathbf{H}1_{(k)}/\tau\|_F^2. \quad (14)$$

Based on the gradient decent algorithm, the solution of (14) is given as

$$\mathbf{S}_{(k+1)}(i,j) = \begin{cases} Tp, & s.t.\ f(Tp) \le f(0) \\ 0, & s.t.\ others, \end{cases} \quad (15)$$

where

$$f(x) = \lambda^{(b)}R(x) + \frac{\tau}{2}(\mathbf{W}^{(b)}_{(k+1)}(i,j) - x + \mathbf{H}1_{(k)}(i,j)/\tau)^2, \quad (16)$$

and

$$Tp = \frac{\sigma^2_{(k)}(\mathbf{W}^{(b)}_{(k+1)}(i,j) + \mathbf{H}1_{(k+1)}(i,j)/\tau) + \lambda^{(b)}\mu_{(k)}\log_2 e}{\sigma^2_{(k)} + \lambda^{(b)}log_2 e}. \quad (17)$$

d) Gaussian model parameter updating:

The coefficients $\mathbf{S}_{(k+1)}$ are collected to update the Gaussian model parameters $\mu_{(k+1)}$ and $\sigma^2_{(k+1)}$ based on the Expectation-Maximization (EM) algorithm [38].

e) **H**0 and **H**1-Subproblem:

$\mathbf{H}0_{(k+1)}$ and $\mathbf{H}1_{(k+1)}$ could be updated directly as follows:

$$\begin{cases} \mathbf{H}0_{(k+1)} = \mathbf{H}0_{(k)} - \rho(\mathbf{W}^{(b)}_{(k+1)} - \mathbf{P}_{(k+1)}) \\ \mathbf{H}1_{(k+1)} = \mathbf{H}1_{(k)} - \tau(\mathbf{W}^{(b)}_{(k+1)} - \mathbf{S}_{(k+1)}). \end{cases} \quad (18)$$

(2) Dictionary Updating:

When other parameters are fixed, the dictionary $\tilde{\mathbf{D}}$ could be updated by solving the following equations:

$$\tilde{\mathbf{D}}^{(b)} = \underset{\mathbf{D}^{(b)}}{\operatorname{argmin}}\, \|\tilde{\mathbf{Y}}^{(b)} - \mathbf{D}^{(b)}\tilde{\mathbf{W}}^{(b)}\|_F^2. \quad (19)$$

Similar to K-SVD, the dictionary atoms are updated individually as

$$\tilde{\mathbf{D}}^{(b)} = \underset{\mathbf{D}^{(b)}}{\operatorname{argmin}}\, \|\mathbf{E}_k - \mathbf{d}_k\mathbf{w}^r_k\|_F^2, \quad (20)$$

where $\mathbf{d}_i$ is the ith column of matrix $\mathbf{D}^{(b)}$, and $\mathbf{w}^r_i$ is the ith row of the matrix $\tilde{\mathbf{W}}^{(b)}$. $\mathbf{E}_l$ is the residual matrix that is computed as $\mathbf{E}_l = \mathbf{Y}^{(b)} - \sum_{i\neq l}\mathbf{d}_i\mathbf{w}^r_i$. Then $\mathbf{d}_l$ is solved by approximating $\mathbf{E}_l$ with a rank-1 matrix using Singular Value Decomposition (SVD) [39], [40].

D. ALGORITHM AND IMPLEMENTATION

Based on the aforementioned inference, the procedure involved in the proposed algorithm is summarized in Algorithm 1 and Algorithm 2. In the following part, we explain how to use the proposed approach for seismic data compression.

Algorithm 1 Sparse Coding Algorithm With Rate Constraint

Input: $\mathbf{Y}^{(b)}$ (input patches from different subbands), I (number of iterations), $\tilde{\mathbf{D}}^{(b)}$ (dictionary), $\lambda^{(b)}$, $\alpha^{(b)}$, ρ and τ.
Initialization: $\mathbf{W}^{(b)}_{(0)}$, $\mathbf{P}_{(0)}$, $\mathbf{S}_{(0)}$, $\mu_{(0)}$, $\sigma^2_{(0)}$, $\mathbf{H}0_{(0)} = \mathbf{0}$, and $\mathbf{H}1_{(0)} = \mathbf{0}$
. **for** $k \leftarrow 1$ **to** I *or not convergence* **do**
  Compute $\mathbf{W}^{(b)}_{(k)}$ by (11);
  Compute $\mathbf{P}_{(k)}$ by (13);
  Compute $\mathbf{S}_{(k)}$ by (15);
  Update $\mu_{(k)}$ and $\sigma^2_{(k)}$ based on EM Algorithm;
  Compute $\mathbf{H}0_{(k)}$ and $\mathbf{H}1_{(k)}$ by (18);
end
Output: $\tilde{\mathbf{W}}^{(b)} = \mathbf{W}^{(b)}_{(I)}$

Algorithm 2 Multiscale Sparse Dictionary Learning With Rate Constraint

Input: $\mathbf{X}$ (original seismic data), $\mathbf{\Phi}$ (multiscale orthogonal transform), $\lambda^{(b)}$, $\alpha^{(b)}$ ($b \in [1, 3S+1]$), ρ and τ (other parameters), T (number of iterations)
Initialization: $\mathbf{W}^{(b)}_{(0)}$, $\mathbf{P}_{(0)}$, $\mathbf{S}_{(0)}$, $\mathbf{H}0_{(0)} = \mathbf{0}$, $\mathbf{H}1_{(0)} = \mathbf{0}$, $\mu_{(0)}$ and $\sigma^2_{(0)}$
Compute the multiscale transform results for each subband as $\mathbf{Z}^{(b)} = (\mathbf{\Phi}^{\mathcal{C}}\mathbf{X})^{(b)}$;
Generate the vectorized forms of $\mathbf{Z}^{(b)}$ (mean values removed) as $\mathbf{Y}^{(b)}$;
for $t \leftarrow 1$ **to** T *or not convergence* **do**
  for $b =$ *1:3S+1* **do**
    Update $\tilde{\mathbf{W}}^{(b)}$ by Algorithm 1;
    Update $\tilde{\mathbf{D}}^{(b)}$ based on (20);
  end
end
Output: $\tilde{\mathbf{D}} = [\tilde{\mathbf{D}}^{(1)}, \ldots, \tilde{\mathbf{D}}^{(b)}, \ldots, \tilde{\mathbf{D}}^{(3S+1)}]$

A diagram based on the proposed multiscale and rate constrained dictionary learning based seismic compression scheme (MRDL) is shown in Figure 3. It includes three components: encoding, decoding, and offline training. In the encoding part, the original seismic data is transformed by the multiscale transform, and wavelet is adopted in this investigation. Furthermore, the transformed seismic data are partitioned into small nonoverlapped patches. The mean values of the different patches, denoted as DC values, and the residual values (the mean value removed from the patches), denoted

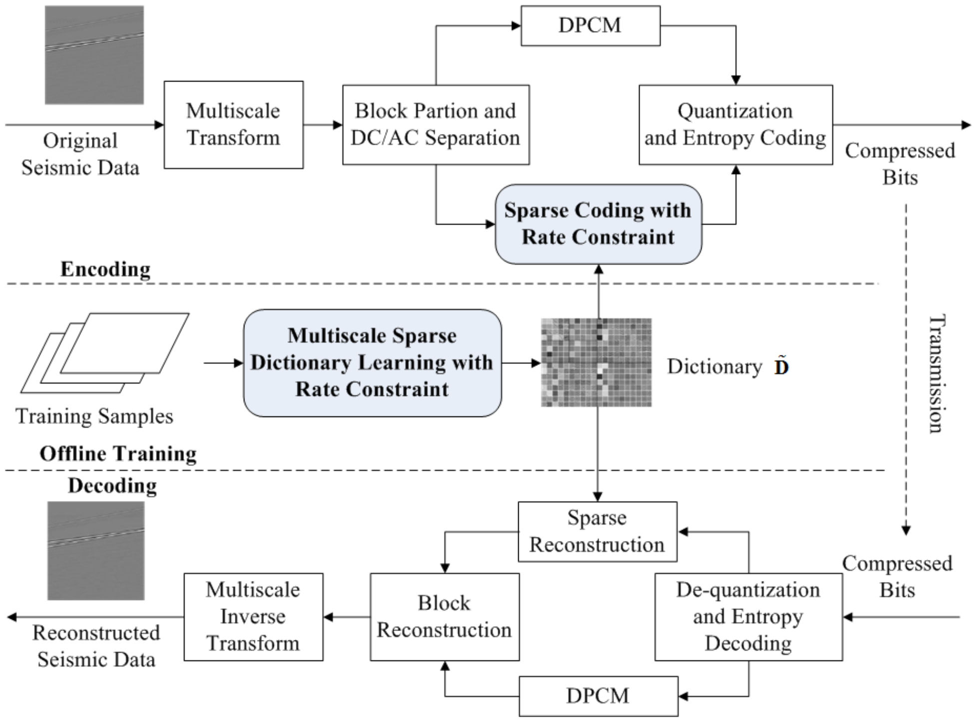


FIGURE 3. The diagram of the proposed seismic data compression method.

as AC values, are encoded separately. The DC values are quantized and coded using differential pulse coded modulation (DPCM) and Huffman coding. The AC values are sparsely represented by an offline trained multiscale dictionary $\tilde{\mathbf{D}} = [\tilde{\mathbf{D}}^{(1)}, \ldots, \tilde{\mathbf{D}}^{(b)}, \ldots, \tilde{\mathbf{D}}^{(3S+1)}]$ based on the proposed sparse coding algorithm with rate constraint. The sparse coefficients are initially scalar-quantized. The values of quantized nonzero coefficients (QNZC) and their positions are separately coded. The positions are considered as a binary sequence where 0 indicates a zero in the sparse coefficients matrix and 1 indicates a nonzero value. These positions are encoded using an adaptive arithmetic coder. Similar to our previous work [12], QNZC values are coded using a near optimum code-book based on the fitted probability model. In the decoding part, the compressed bits are dequantized and entropy decoded. The AC values are reconstructed from the decoded sparse coefficients using the offline trained dictionary $\tilde{\mathbf{D}}$. As such, the seismic data is reconstructed using the multiscale inverse transform with the reconstructed DC and AC values of different blocks. In the offline training part, the training samples are the AC removed values of the partitioned patches. Based on the proposed multiscale sparse dictionary learning algorithm with rate constraint, an offline trained dictionary $\tilde{\mathbf{D}}$ is generated. The total compressed bits include the bits consumed for the DC values and AC values. Other control information including the packet header is not contained because their contribution to the rate is negligible.

III. EXPERIMENTAL RESULTS

In this section, two aspects of the experiments are considered to evaluate the performance of the proposed method.

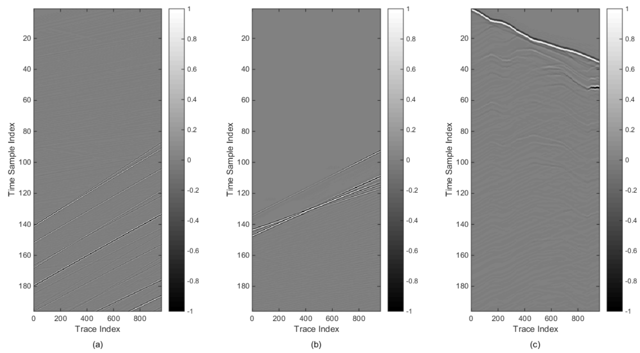


FIGURE 4. Testing of seismic data from (a) 2004BP dataset, (b) 2007BP dataset, and (c) Model94 dataset.

To analyze the properties of the proposed dictionary learning algorithms, the performance of different dictionary learning-based approaches are analyzed. Furthermore, we evaluate the rate-distortion performance of the proposed seismic data compression scheme on a variety of seismic data, thereby comparing it against other typical seismic data compression methods.

A. EXPERIMENTS SETUP

In the experiments, the 2004BP, 2007BP and Model94 datasets of public seismic data [41] are employed for comparison of the different seismic data compression methods. In each dataset, we choose seismic data from 960 adjacent sensors. For each sensor, 192 time samples are selected. Thus, each seismic data from dataset 2004BP, 2007BP and Model94 has a resolution of 192 × 960, which is shown in Figure 4. The dictionary is supposed to be a general dictionary that is learned from the seismic data from all datasets. Other fixed parameters in the dictionary learning are $\rho = 50$ and $\tau = 50$. Similar to other dictionary algorithms, $\alpha^{(b)}$ is the parameter used to control the sparsity of the coefficients. In the quantization step, each QNZC is quantized into 10 bits. The peak signal to noise ratio (PSNR) and the structural similarity index (SSIM) [42], which are two quality metrics used for reconstruction performance comparison, are exploited and defined as follows:

$$\mathrm{PSNR} = 10\log_{10}\frac{J \times MAX(\mathbf{s})^2}{\|\mathbf{s}-\tilde{\mathbf{s}}\|_2^2}, \tag{21}$$

where $\mathbf{s}$ and $\tilde{\mathbf{s}}$ denote the original and reconstructed seismic data, respectively. J is the number of elements in $\mathbf{s}$. $MAX(\mathbf{s}) = 1$ is the maximum absolute value of the original seismic data in this report.

$$\mathrm{SSIM} = \frac{(2\mu_{\mathbf{s}}\mu_{\tilde{\mathbf{s}}} + c_1)(2\sigma_{\mathbf{s}\tilde{\mathbf{s}}} + c_2)}{(\mu_{\mathbf{s}}^2 + \mu_{\tilde{\mathbf{s}}}^2 + c_1)(\sigma_{\mathbf{s}}^2 + \sigma_{\tilde{\mathbf{s}}} + c_2)}, \tag{22}$$

where $\mu_{\mathbf{s}}$ and $\mu_{\tilde{\mathbf{s}}}$ are the average of $\mathbf{s}$ and $\tilde{\mathbf{s}}$. $\sigma_{\mathbf{s}}$ and $\sigma_{\tilde{\mathbf{s}}}$ denote the variance of $\mathbf{s}$ and $\tilde{\mathbf{s}}$. $\sigma_{\mathbf{s}\tilde{\mathbf{s}}}$ is the covariance of $\mathbf{s}$ and $\tilde{\mathbf{s}}$. c_1 and c_2 are two constants.

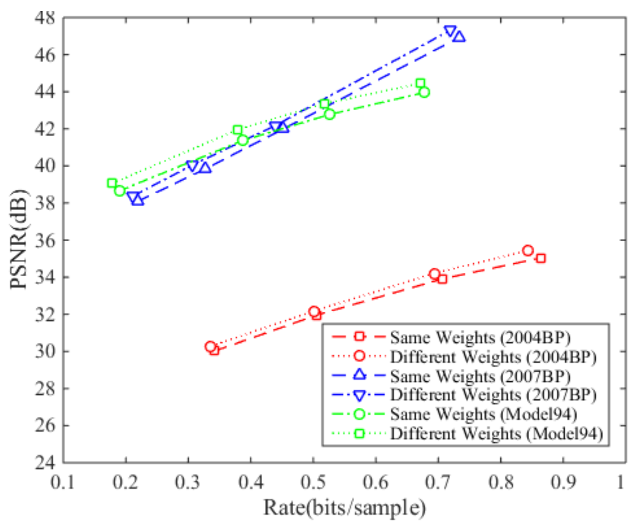


FIGURE 5. Performance comparison for different λ strategies.

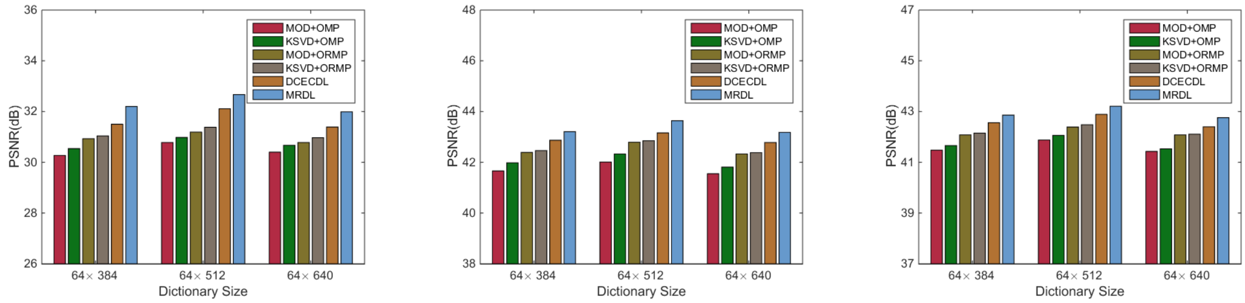
B. COMPARISONS OF DIFFERENT DICTIONARY LEARNING ALGORITHMS FOR SEISMIC DATA COMPRESSION

Initially, the flexibility of the proposed method for rate control is analyzed. We test two different groups of regularization parameters $\lambda^{(b)}$ as shown in Figure 5. In one group, all subbands have the same weights as $[\lambda^{(1)}, \ldots, \lambda^{(7)},] = [2 \times 10^{-7}, \ldots, 2 \times 10^{-7}]$. In the other group, different weights are assigned based on the subbands' importance in the reconstruction process, the values of which are $[\lambda^{(1)}, \ldots, \lambda^{(4)}, \lambda^{(5)}, \ldots, \lambda^{(7)},] = [10^{-7}, \ldots, 10^{-7}, 5 \times 10^{-7}, \ldots, 5 \times 10^{-7}]$. The rate is changed by $\alpha^{(b)}$, which adopts the same values in all subbands for simplicity. The compression performance is mainly evaluated by PSNR (The difference of SSIMs is not significant in this experiment). From Figure 5, it can be determined that parameters with different weights can achieve better rate-distortion curves for all the three test datasets. The main reason is that the lower subbands are more important compared to the higher subbands in the reconstruction process. Therefore, a small λ for the lower subbands means that the reconstruction error of these subbands could be small, and a large λ for the higher subbands implies that fewer rates will be required for the higher subbands. Therefore, a better rate-distortion balance can be built using this strategy. However, we do not discuss the procedure for identifying the best $\lambda^{(b)}$ for different subbands. In the following experiments, we use regularization parameters with different weights.

Secondly, different dictionary learning algorithms for seismic data compression are compared. The influence of the dictionary size is also discussed in the following experiment. The dictionary learning algorithms include MOD, KSVD, and DCECDL. We use OMP and ORMP for sparse coding in MOD and KSVD. The number of iterations was 30 in this experiment. We construct the initial dictionary by randomly selecting atoms from the training dataset. The dictionary size in this comparison is 64×384, 64×512 and 64×640. In the aforementioned dictionary learning methods, the sparse coefficients are generated using sparse coding by changing the sparsity. Furthermore, these coefficients are quantized and coded using the adaptive arithmetic coding algorithm. We choose a rate of 0.5 bits/sample as an example. The comparison results are shown in Figure 6. From Figure 6, we determine that the dictionary size influences the compression performance. The results demonstrate that the best dictionary size for compression in this experiment is 64×512. The reason may be as follows: (1) when the dictionary size is increased from 64×384 to 64×512, superior reconstruction quality could be achieved with the same sparsity by introducing more dictionary atoms. Although this increases the size of the sparse coefficient matrix, compared to the improvement in the reconstruction quality, the increase of the rate can be neglected. However, when the dictionary size is increased to 64×640, the improvement in reconstruction quality is not significant. The rate increases because the size of the sparse coefficient matrix is increased when a bigger dictionary is chosen. Actually, it is very difficult to find an optimal dictionary size for compression, hence not discussed in this report. In addition, using the same dictionary learning algorithm such as MOD or KSVD, ORMP performs OMP. Similarly, KSVD outperforms MOD while the same sparse coding algorithm is adopted. By incorporating the rate constraint into the dictionary learning process, DCECDL and MRDL perform better than MOD and KSVD. MRDL can produce the best compression performance using a more efficient rate controlled strategy with multiscale decomposition.

C. COMPARISONS OF DIFFERENT SEISMIC DATA COMPRESSION METHODS

Firstly, the compression performance of different seismic data compression methods is compared. In the wavelet-based



(a) 2004BP dataset (b) 2007BP dataset (c) Model94 dataset

FIGURE 6. **Compression performance comparisons of different dictionary learning algorithms for different datasets.**

seismic data compression method (denoted as Wavelet), the Cohen-Daubechies-Feauveau 9/7 wavelet is adopted as the transform basis. The decomposition level is $S = 2$. In the traditional wavelet-based compression method for images such as SPIHT and JPEG2000, bit plane coding algorithms are adopted to code the wavelet coefficients. However, these coding algorithms are always complicated. In this report, we use a simple coding strategy for the wavelet-based method. The coefficients are initially quantized. Then the significant coefficients are determined according to a given threshold value. Both the significant coefficients and their positions are additionally coded using an arithmetic coding algorithm before transmission. By changing the threshold value, the number of significant coefficients can be changed. Consequently, both the rate and distortion can be adjusted. The curvelet and contourlet, which also have the properties of multiresolution, are widely used in denoising of seismic signals because of their superior performance in the representation of geometric structures. Therefore, we replace the wavelet transform in the wavelet-based seismic compression method by curvelet and contourlet. Based on these modifications, the curvelet-based and contourlet-based seismic data compression methods (denoted as Curvelet and Contourlet) are developed. For better comparisons, the sparse coding method is further introduced to select the significant coefficients $\mathbf{x}$ from the seismic data $\mathbf{y}$ for the curvelet-based seismic data compression method, which is denoted as

$$\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1, \text{ s.t. } \|\mathbf{\Phi x} - \mathbf{y}\|_2^2 \leq \varepsilon^2. \tag{23}$$

$\mathbf{\Phi}$ denotes the curvelet transform and $\tilde{\mathbf{x}}$ is the sparse coefficient solved from (23) by using a basis pursuit algorithm [43]. The rate and distortion are controlled by changing ε^2. The SIODL is also tested in this paper. A sliding window with the size 64×64 is used to train the online dictionary, and both the sparse coefficients and the increment of dictionary are transmitted. The sparse coefficients are scalar-quantized and coded by an arithmetic coding algorithm for compression. In the MRDL, the multiscale dictionary is trained from the training dataset. Daubechies 9/7 wavelets are chosen as the multiscale orthogonal transform for the decomposition level 2. The number of subbands is then equal to $3S + 1 = 7$.

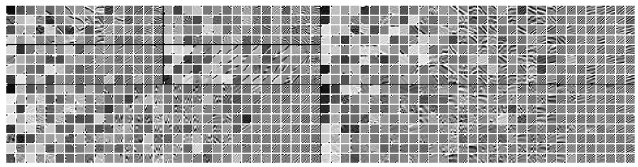


FIGURE 7. **Learned multiscale dictionary.**

Furthermore, the coefficients transformed by the wavelets are partitioned into 8×8 nonoverlapping blocks. Based on the preceding discussion, the dictionary size is 64×512 and each dictionary is trained with 10 iterations. The learned multiscale dictionary in the following experiment is shown in Figure 7.

Figure 8 demonstrates the rate-distortion curves for the comparison of different seismic data compression methods (Wavelet, Contourlet, Curvelet, Curvelet+SP, SIODL, and MRDL). From the experimental results, it can be determined that Curvelet outperforms Wavelet and Contourlet in the 2004BP and 2007BP datasets. For the Model94 dataset, the compression performance of Curvelet and Wavelet is similar (Wavelet performs better at a high compression rate). In addition, both of them have better compression performance than Contourlet. The main reason being as follows: (1) contourlet could not capture the contour of the seismic wave very well for these three seismic datasets. Therefore, its reconstruction quality is not ideal. Moreover, its representation is too redundant, which requires the compression of a large number of contourlet coefficients. This increases the compression rates. (2) Although the representation of curvelet is also redundant, its reconstruction quality for most seismic datasets (2004BP and 2007BP) is high. Therefore, it has a relatively high compression performance. Furthermore, the compression performance of Curvelet could be improved by introducing sparse coding (denoted as Curvelet+SP). For example, the PSNRs of Curvelet could be increased by 1.7 dB in the 2004BP dataset when the rate is 0.5 bits/sample. Moreover, the increase could be more than 2.0 dB in the Model94 dataset. Sparse coding is helpful for increasing the compression performance because the most significant coefficients for the reconstruction could be selected. As such, better reconstruction quality could be achieved with the same number of nonzero coefficients.

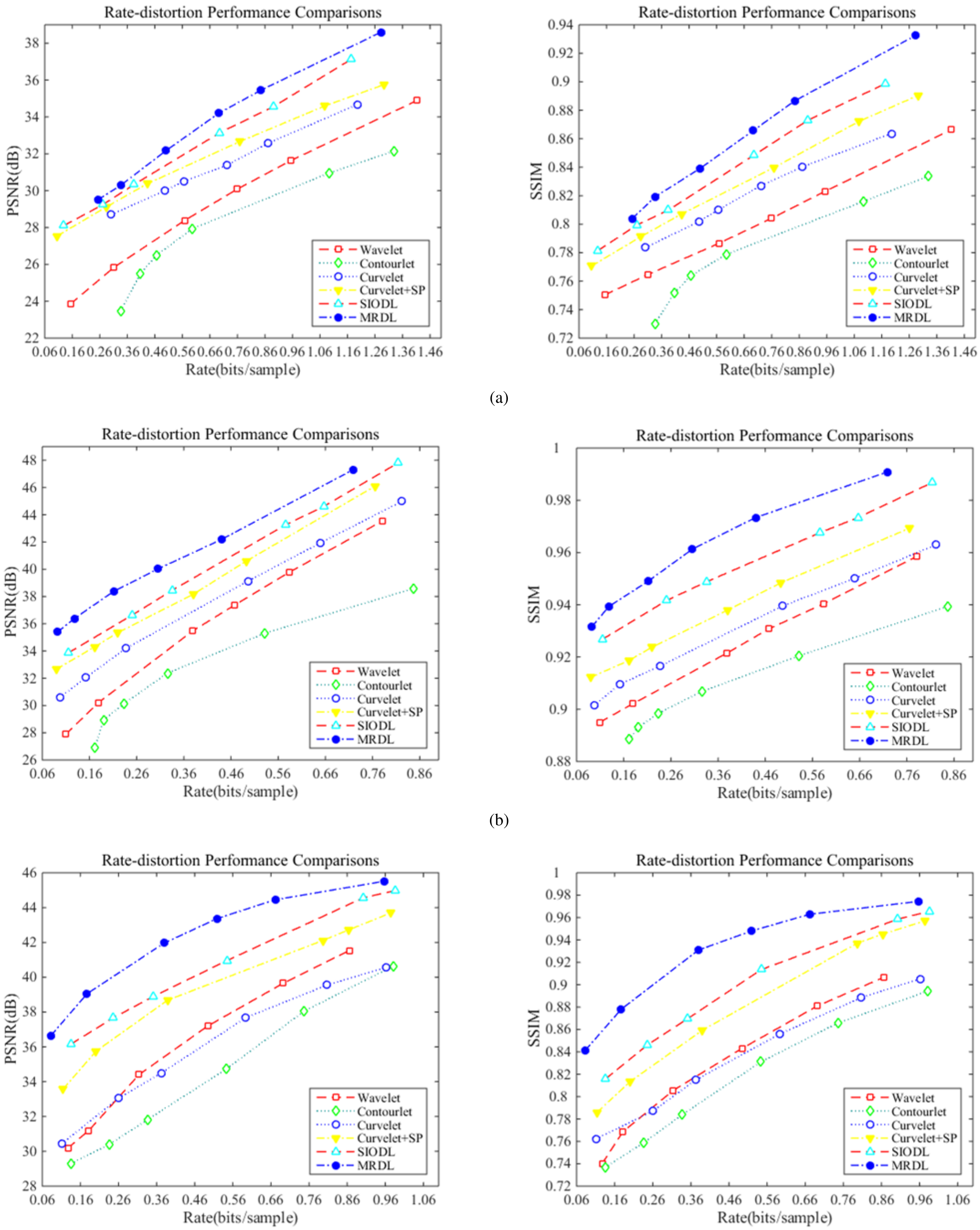


FIGURE 8. **Rate-distortion performance comparisons for different seismic compression methods. (a) 2004BP dataset. (b) 2007BP dataset. (c) Model94 dataset.**

By learning a data-driven dictionary from the data, the compression performance of SIODL and MRDL are better than Wavelet, Contourlet, Curvelet, and Curvelet+SP. The proposed MRDL algorithm performs best in all tested seismic datasets that were evaluated by PSNR and SSIM. For example, compared with Curvelet+SP, the increase of PSNR could be approximately 0.9dB, 1.4dB and 2.6dB in the 2004BP, 2007BP and Model94 dataset, respectively. This could be

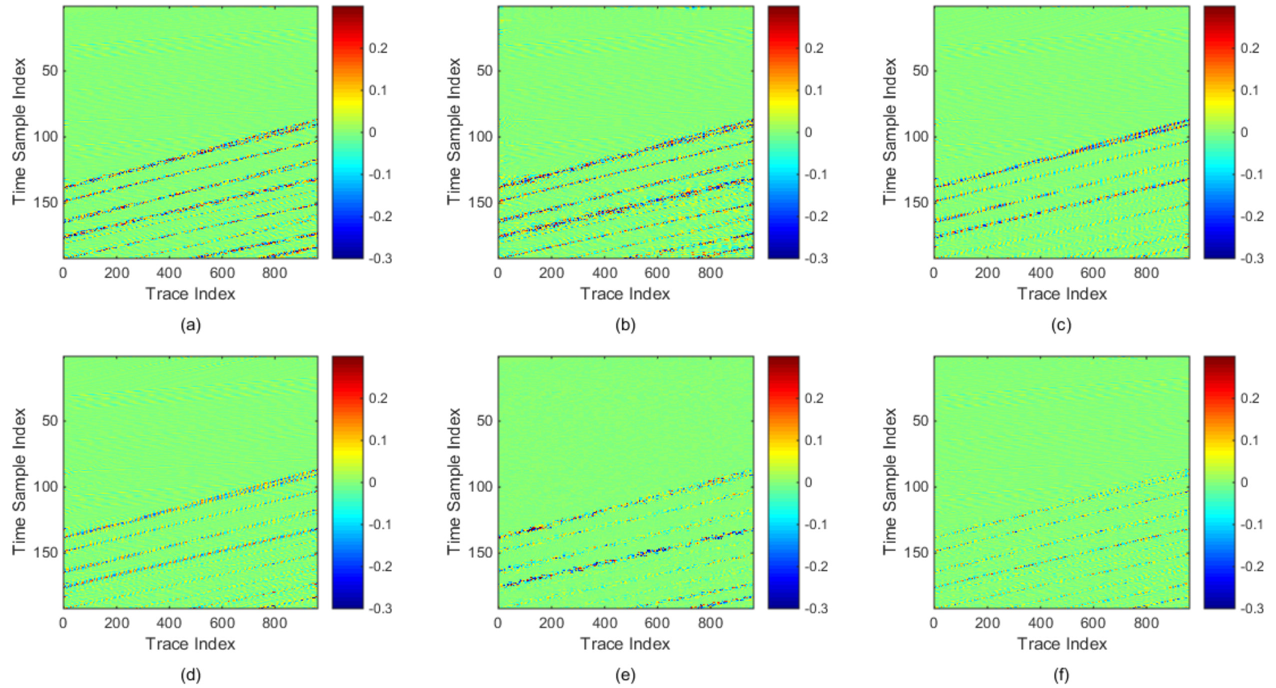


FIGURE 9. Reconstruction errors of 2004BP dataset for different seismic compression methods. (a) Wavelet. (b) Contourlet. (c) Curvelet. (d) Curvelet+SP. (e) SIODL. (f) MRDL.

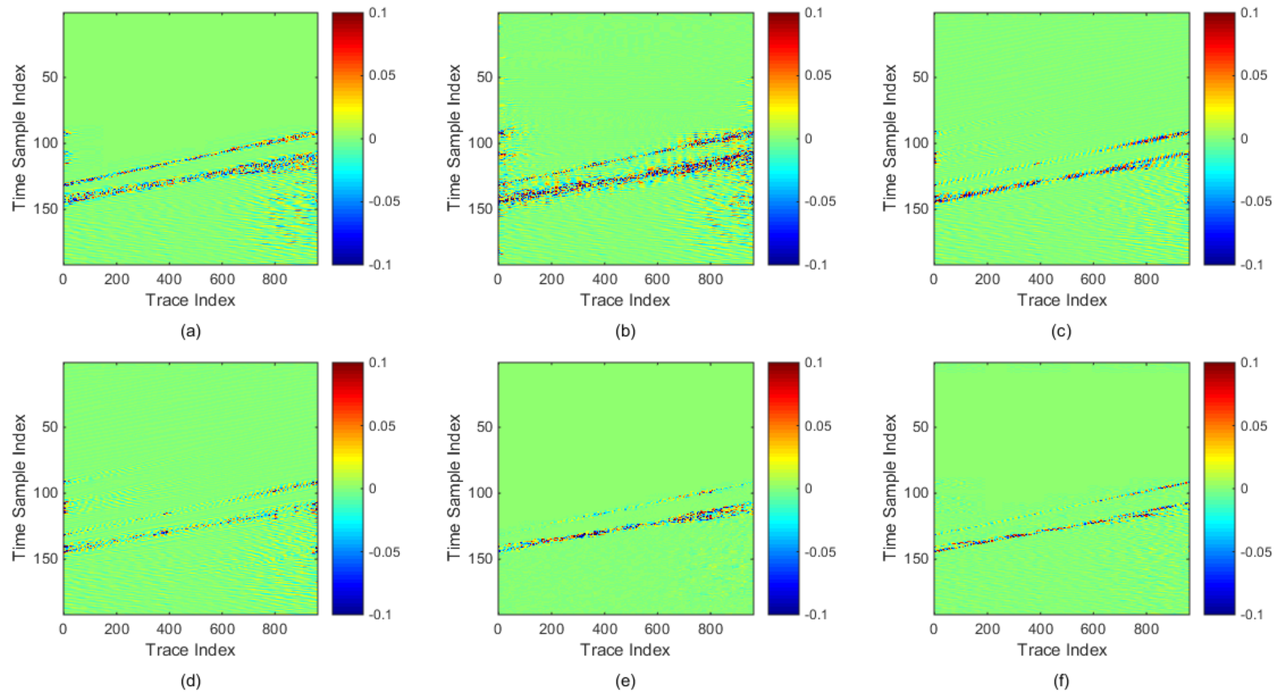


FIGURE 10. Reconstruction errors of 2007BP dataset for different seismic compression methods. (a) Wavelet. (b) Contourlet. (c) Curvelet. (d) Curvelet+SP. (e) SIODL. (f) MRDL.

attributed to the proposed multiscale sparse dictionary learning model with rate constraint. Such strategy could optimize rate-distortion performance better. A similar conclusion could be established based on the evaluation of SSIM.

Next, we evaluate the subjective quality of different seismic compression methods (Wavelet, Contourlet, Curvelet, Curvelet+SP, SIODL, and MRDL) between the original seismic data and the reconstructed seismic data at rate 0.5 bits/sample. The reconstruction errors of different seismic compression methods on 2004BP, 2007BP, and Model94 datasets are presented in Figure 9, 10, and 11. The edges of the seismic data are always important for geological

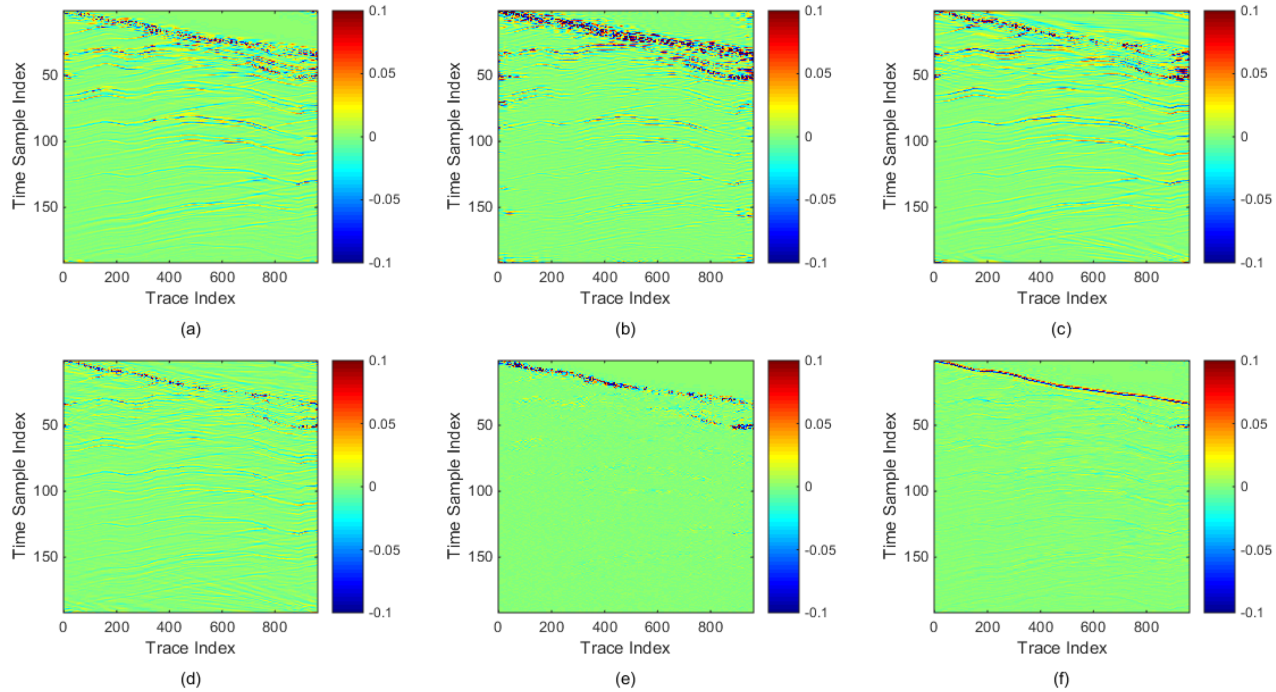


FIGURE 11. **Reconstruction errors of Model94 dataset for different seismic compression methods. (a) Wavelet. (b) Contourlet. (c) Curvelet. (d) Curvelet+SP. (e) SIODL. (f) MRDL.**

analysis and most of the reconstruction errors in the smooth regions are close to 0, therefore, the reconstruction errors in the edge regions are under the most focus in the following paper. Generally speaking, the reconstruction errors of all seismic data compression methods on 2004BP dataset are larger than other test datasets. For example, the maximum absolute value of errors on 2004BP dataset is close to 0.3, but the maximum absolute values of errors on 2007BP and Model94 datasets are no more than 0.1. The reason is that more edges (a lot of tilted lines with large values in Figure 4) exist in 2004BP dataset, and its geometric structures are more complicated than those of 2007BP and Model94 datasets. Therefore, the compression performance of above seismic data compression methods on 2004BP dataset is lower than 2007BP and Model94 dataset. For each dataset, we notice the following: Firstly, the reconstruction errors of Contourlet in most edge regions are larger than other seismic data compression methods. This is consistent with the rate-distortion comparison in Figure 8, where both the PSNR and SSIM values of Contourlet are lowest. Compared with Contourlet, the reconstruction errors can be reduced in Wavelet and Curvelet. The reason is that the edges of seismic data can be better represented by wavelet and curvelet when compared with contourlet. By introducing sparse coding, Curvelet+SP is able to find a set of curvelet basis such that the seismic data is well represented. Therefore, its reconstruction errors are not very obvious. Due to the good fit of the dictionary learning model to the edge structures, SIODL and MRDL will produce the most pleasing visual seismic data. This can be verified from the fact that the reconstruction errors exist in only a small part of the edges in Figure 9, 10 and 11. MRDL has the best reconstruction quality for the reason that its reconstruction errors are not noticeable in most edge regions. For example, most of the tilted lines in 9(f), which denotes the reconstruction errors along the edge of the seismic data, are not apparent. Therefore, we can conclude that MRDL can preserve the best geometric content (edge information) of the seismic data at the same rate compared with other methods, which further verifies its efficiency.

Finally, we add noise to the seismic data and use the noisy data to verify the performance of different seismic data compression methods in more realistic situations. The 2004BP dataset is used in this experiment. The noise is additive white Gaussian noise with a mean of $\mu = 0$ and a standard deviation of $\delta = 0.030$. We only choose Curvelet+SP, SIODL, and MRDL as the seismic compression algorithm because of their superior performance in previous experiments. In Figure 12, the reconstruction errors of Curvelet+SP, SIODL, and MRDL seismic data compression methods under noise situation are presented. In general, the reconstruction errors of all three seismic data compression methods increase when compared with Figure 9, while the rate is also increased to approximately 0.8 bits/sample. This is due to the fact that noise, which can be also regarded as a kind of high frequency signal, will also consume rates for compression and thus degrade the compression efficiency. By introducing sparsity constraint, the seismic data can also be well represented by curvelet or the basis learned from the data itself under noise situation. Thus, the increase of reconstruction errors of the edges is limited. Therefore, all

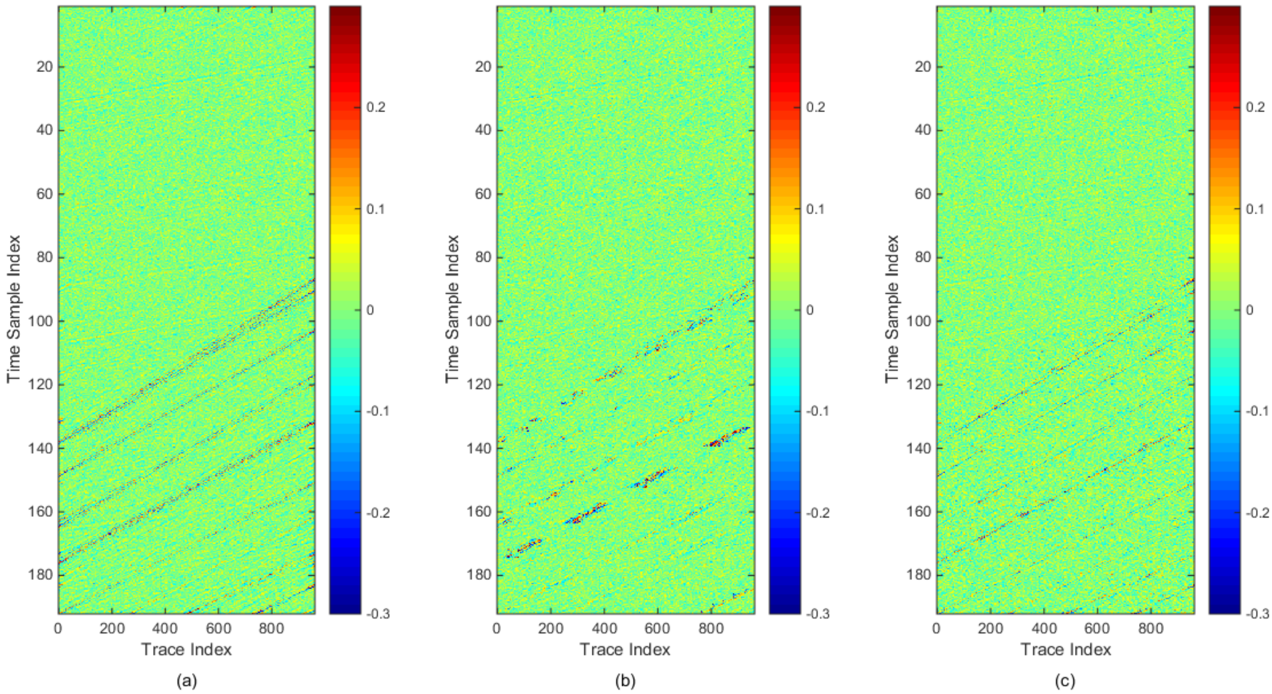


FIGURE 12. **Reconstruction errors of 2004BP dataset for $\delta = 0.030$ using different seismic compression methods. (a) Curvelet+SP. (b) SIODL. (c) MRDL.**

the three methods are robust to the noise. In Figure 12, the reconstruction errors mainly include the reconstruction errors of noise and the reconstruction errors of edges. The reconstruction errors of noise are demonstrated as some random speckle points in the entire seismic, the absolute values of which may be more than 0.1. It could be found that all the three seismic compression methods have similar reconstruction errors of noises. The advantage of proposed MRDL is the ability to recover the edges efficiently. This could be verified from the fact that MRDL has the smallest reconstruction errors of edges compared with Curvelet+SP and SIODL. This implies that the seismic geometrical information could be better reconstructed in MRDL, which is important for seismic applications. Therefore, we can conclude that MRDL is also suitable for compressing seismic data in a noisy situation.

IV. CONCLUSION

In this report, we propose a novel seismic compression method based on multiscale and rate-constrained dictionary learning. To fully capture the characteristics of the seismic data, multiscale decomposition such as wavelet is integrated into the dictionary learning process; therefore, more degrees of freedom could be provided for compression. Furthermore, a rate-constrained term is also included in the optimization model, which renders the optimization more suitable for compression purposes. To address the aforementioned optimization problem, the alternating direction method of multipliers is used. Finally, the experimental results indicate that the proposed compression method results in better compression performance compared to the state-of-the-art dictionary learning and seismic compression methods. We also evaluate its efficiency in a noisy situation. To further improve the compression performance of the proposed method, the process of determining the optimal weights for different subbands should be investigated.

REFERENCES

[1] N. Iqbal, E. Liu, J. H. McClellan, A. Al-Shuhail, S. I. Kaka, and A. Zerguine, "Detection and denoising of microseismic events using time–frequency representation and tensor decomposition," *IEEE Access*, vol. 6, pp. 22993–23006, 2018.

[2] J. Ma, G. Plonka, and H. Chauris, "A new sparse representation of seismic data using adaptive easy-path wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 3, pp. 540–544, Jul. 2010.

[3] Z. Wang, Z. Hou, R. Hu, and J. Xiao, "A lossless recompression approach for video streaming transmission," *IEEE Access*, vol. 7, pp. 35162–35172, 2019.

[4] L. Guo, D. Zhou, J. Zhou, S. Kimura, and S. Goto, "Lossy compression for embedded computer vision systems," *IEEE Access*, vol. 6, pp. 39385–39397, 2018.

[5] A. S. Spanias, S. B. Jonsson, and S. D. Stearns, "Transform methods for seismic data compression," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 407–416, May 1991.

[6] P. Aparna and S. David, "Adaptive local cosine transform for seismic image compression," in *Proc. Int. Conf. Adv. Comput. Commun.*, Dec. 2006, pp. 254–257.

[7] W. Wu, Z. Yang, Q. Qin, and F. Hu, "Adaptive seismic data compression using wavelet packets," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, Jul./Aug. 2006, pp. 787–789.

[8] H. Douma and M. de Hoop, "Leading-order seismic imaging using curvelets," *Geophysics*, vol. 72, no. 6, pp. S231–S248, 2007.

[9] K. Skretting and K. Engan, "Image compression using learned dictionaries by RLS-DLA and compared with K-SVD," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 1517–1520.

[10] M. Xu, S. Li, J. Lu, and W. Zhu, "Compressibility constrained sparse representation with learnt dictionary for low bit-rate image compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1743–1757, Oct. 2014.

[11] X. Zhan, R. Zhang, D. Yin, A. Hu, and W. Hu, "Remote sensing image compression based on double-sparsity dictionary learning and universal trellis coded quantization," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 1665–1669.

[12] X. Tian, A. Abdi, E. Liu, and F. Fekri, "Memory-assisted seismic signal compression based on dictionary learning and sparse coding," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 358–362.

[13] E. Liu, A. Payani, and F. Fekri, "Seismic data compression using online double-sparse dictionary learning schemes," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, p. 449.

[14] A. Payani, A. Abdi, X. Tian, F. Fekri, and M. Mohandes, "Advances in seismic data compression via learning from data: Compression for seismic data acquisition," *IEEE Signal Process. Mag.*, vol. 35, no. 2, pp. 51–61, Mar. 2018.

[15] L. Zhu, E. Liu, and J. H. McClellan, "Seismic data denoising through multiscale and sparsity-promoting dictionary learning," *Geophysics*, vol. 80, no. 6, pp. WD45–WD57, 2015.

[16] T. Ryen, G. M. Schustertt, and A. K. Katsaggelos, "A rate-distortion optimal coding alternative to matching pursuit," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, May 2002, pp. III-2177–III-2180.

[17] T. Ryen, G. M. Schuster, and A. K. Katsaggelos, "A rate-distortion optimal alternative to matching pursuit," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1352–1363, May 2004.

[18] A. Abdi, A. Payani, and F. Fekri, "Learning dictionary for efficient signal compression," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3689–3693.

[19] X. Tian, A. Abdi, E. Liu, and F. Fekri, "Seismic signal compression through delay compensated and entropy constrained dictionary learning," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[20] B. Ophir, M. Lustig, and M. Elad, "Multi-scale dictionary learning using wavelets," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1014–1024, Sep. 2011.

[21] X. Tang, H. Xiong, and X. Jiang, "Multiscale online dictionary learning for quality scalable video coding," in *Proc. Data Compress. Conf.*, Mar. 2014, p. 428.

[22] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Dictionary learning for sparse representation: A novel approach," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1195–1198, Dec. 2013.

[23] M. Hyder and K. Mahata, "An approximate l0 norm minimization algorithm for compressed sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Apr. 2009, pp. 3365–3368.

[24] P. Indyk and M. Ruzic, "Near-optimal sparse recovery in the L1 norm," in *Proc. 49th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2008, pp. 199–207.

[25] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, no. 1, pp. 490–530, May 2015.

[26] J. Adler, B. D. Rao, and K. Kreutz-Delgado, "Comparison of basis selection methods," in *Proc. 13th Asilomar Conf. Signals, Syst. Comput., Conf. Rec.*, vol. 1, Nov. 1996, pp. 252–257.

[27] T. Blumensath, M. E. Davies, and G. Rilling, "Greedy algorithms for compressed sensing," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[28] K. Skretting and J. H. Husøy, "Partial search vector selection for sparse signal representation," in *Proc. NORSIG*, 2003, pp. 1–6.

[29] K. Engan, S. O. Aase, and J. H. Husøy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Mar. 1999, pp. 2443–2446.

[30] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[31] L. Nissenbaum and J. S. Lim, "Globally optimal algorithms for transform selection in multiple-transform signal compression," in *Proc. Data Compress. Conf. (DCC)*, Mar./Apr. 2016, p. 624.

[32] X. Liu, "Composition of DCT and wavelet transform for image compression," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2008, p. 532.

[33] H. Wang, Y. Xia, and Z. Wang, "Dictionary learning-based image compression," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3235–3239.

[34] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, Sep. 2016.

[35] W. Byrne, "Alternating minimization and Boltzmann machine learning," *IEEE Trans. Neural Netw.*, vol. 3, no. 4, pp. 612–620, Jul. 1992.

[36] B. Wahlberg, S. Boyd, M. Annergren, and Y. Wang, "An ADMM algorithm for a class of total variation regularized estimation problems," *IFAC Proc. Vols.*, vol. 45, no. 16, pp. 83–88, Jul. 2012.

[37] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "Optimal parameter selection for the alternating direction method of multipliers (ADMM): Quadratic problems," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 644–658, Mar. 2015.

[38] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

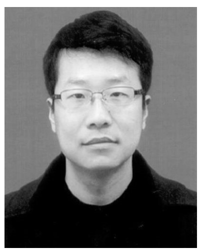
[39] P. Howland and H. Park, "Generalizing discriminant analysis using the generalized singular value decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 995–1006, Aug. 2004.

[40] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognit.*, vol. 46, no. 12, pp. 3519–3532, 2013.

[41] (2004). *BPbenchmark*. [Online]. Available: http://software.seg.org/datasets/2D/

[42] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 2366–2369.

[43] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.



XIN TIAN received the B.S. degree from the Department of Electronic Science and Technology and the Ph.D. degree from the Institute for Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China, in 2004 and 2010, respectively.

From 2015 to 2016, he was a Visiting Faculty Member with the School of Electrical and Computer Engineering, Georgia Institute of Technology. He is currently an Associate Professor with the School of Electronic Information and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, China. He has authored or coauthored more than 60 scientific articles and holds more than ten Chinese patents. His current research interests include image/seismic compression, dictionary learning, sparse coding, and computational imaging.

Dr. Tian's awards include the First Prize for Scientific and Technological Progress in Surveying and Mapping in China and the First Prize for Excellent Achievements of Information Technology in Electric Power Industry of China. He is a Reviewer of the IEEE Transactions on Signal Processing, the IEEE Transactions on Computers, and the IEEE Signal Processing Letters.