2017 Data Compression Conference

Globally Variance-Constrained Sparse Representation for Rate-Distortion Optimized Image Representation

Xiang Zhang†, Siwei Ma†‡, Zhouchen Lin*§, Jian Zhang†, Shiqi Wang⋆, and Wen Gao†‡

†Institute of Digital Media & Cooperative Medianet Innovation Center, Peking University, Beijing, China

‡Peking University Shenzhen Graduate School, Shenzhen, China

*Key Lab. of Machine Perception (MoE), Peking University, Beijing, China

§Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, Shanghai, China

⋆Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore

{x_zhang, swma, zlin, jian.zhang, sqwang, wgao}@pku.edu.cn

Abstract

Sparse representation is efficient to approximately recover signals by a linear composition of a few bases from an over-complete dictionary. However, in the scenario of data compression, its efficiency and popularity are hindered due to the extra overhead for encoding the sparse coefficients. Therefore, how to establish an accurate rate model in sparse coding and dictionary learning becomes meaningful, which has been not fully exploited in the context of sparse representation. According to the *Shannon* entropy inequality, the variance of data source can bound its entropy, thus can reflect the actual coding bits. Therefore, a Globally Variance-Constrained Sparse Representation (GVCSR) model is proposed, where a variance-constrained rate term is introduced to the conventional sparse representation. To solve the non-convex optimization problem, we employ the Alternating Direction Method of Multipliers (ADMM) for sparse coding and dictionary learning, both of which have shown state-of-the-art rate-distortion performance in image representation.

1 Introduction

Transform codings, e.g. Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), have shown great power in de-correlation and have been widely adopted in image compression standards. The basis functions of both DCT and DWT are orthogonal and meanwhile fixed despite the characteristics of the input signals. Such inflexibility may greatly restrict their representation efficiency. In the early 1990s, Olshausen and Field firstly proposed the sparse coding with a learnt dictionary [1] to represent an image, where the bases in sparse model are over-complete and nonorthogonal. It is widely believed that the sparsity property is efficient in dealing with rich, varied and directional information contained in natural scenes [2]. Recent studies further validated the idea that the sparse coding performs in a perceptually meaningful way that mimics the Human Visual System (HVS) on natural images [3–5]. Based on the sparse model, numerous tasks have been successfully achieved, including image denoising [6], restoration [7], quality assessment [8], etc.

In the conventional sparse representation model, the objective function is to minimize the distortion given a constrained sparsity term as follows,

$$(\boldsymbol{D}, \{\boldsymbol{A}_i\}) = \mathop{\arg\min}_{\boldsymbol{D},\{\boldsymbol{A}_i\}} \sum_i \|\boldsymbol{T}_i - \boldsymbol{D}\boldsymbol{A}_i\|_2^2, \text{ s.t. } \|\boldsymbol{A}_i\|_0 \leqslant L, \|\boldsymbol{D}_j\|_2^2 \leqslant 1, \forall j \in \{1, 2, \ldots, M\} \tag{1}$$

2375-0359/17 $31.00 © 2017 IEEE
DOI 10.1109/DCC.2017.63

where $\boldsymbol{D} \in \mathbb{R}^{N \times M}$ is the redundant dictionary with M bases. $\boldsymbol{T}_i \in \mathbb{R}^{N \times 1}$ indicates the training data, and $\boldsymbol{A}_i \in \mathbb{R}^{M \times 1}$ is the corresponding sparse representation vector, whose ℓ_0 norm is constrained by a given sparse level L. Typical algorithms for dictionary learning include the Method of Optimal Directions (MOD) [9], KSVD [10], Online Dictionary Learning (ODL) [11] and Recursive Least Square (RLS) [12]. With respect to the trained dictionary $\boldsymbol{D}$, the sparse decomposition calculates appropriate coefficients $\boldsymbol{A}_i$ for the input signal $\boldsymbol{S}_i$,

$$\boldsymbol{A}_i = \arg\min_{\boldsymbol{A}_i} \|\boldsymbol{S}_i - \boldsymbol{D}\boldsymbol{A}_i\|_2^2, \text{ s.t. } \|\boldsymbol{A}_i\|_0 \leqslant L, \tag{2}$$

which is a subproblem of (1). Several suboptimal solutions have been proposed to solve this, including ℓ_1 convex relaxation [13] and the well-known Matching Pursuit Family (MPF) algorithms [14].

Despite the fact that sparse coding can provide more efficient representation than orthogonal transforms [15], its efficiency in compression tasks is however limited because encoding sparse coefficients would be rather costly. Although the more nonzero coefficients are included from the dictionary, the more costly it usually is to code them, the cost may not be proportional to the number of nonzero coefficients. The Rate-Distortion Optimized Matching Pursuit (RDOMP) approaches [16,17] were proposed to address this issue, where the coding rate was estimated based on the probabilistic model of the coefficients. Despite their performance improvements on image compression, the RDOMPs may also have some limitations. Firstly, such matching pursuit based methods may suffer from the instability in obtaining the sparse coefficients [18]. Secondly, they operate and encode each sample separately, which ignores the data structure information and lacks global constraint over all input samples. This may lead to quite different representations of two similar samples and result in performance decrease [19]. Finally, the RDOMP methods are nontrivial to be incorporated with the dictionary learning algorithm and such inconsistency may also decrease the efficiency.

To overcome these problems, the Globally Variance-Constrained Sparse Representation (GVCSR) model is proposed in this work, where a variance-constraint term is introduced into the objective function in sparse representation. By incorporating the rate model, minimization of the objective function turns out to be a joint rate-distortion optimization problem, which can be efficiently solved by Alternating Direction Method of Multipliers (ADMM) [20] in both sparse coding and dictionary learning. In particular, here "globally" aims to emphasize the way for solving the model, which significantly distinguishes from the separate manner in matching pursuit. Therefore, such optimization based method can effectively utilize the intrinsic data structure and reduce the instability.

The rest of the paper is organized as follows. In Section 2, we present the GVCSR model and the solution to solve it via ADMM. Section 3 evaluates the efficiency of GVCSR. Finally, we conclude this paper in Section 4.

2 Globally Variance-Constrained Sparse Representation

In this section, we firstly present the Globally Variance-Constrained Sparse Representation (GVCSR) model. The effective solutions for sparse coding and dictionary learning are then introduced.

2.1 Rate-Distortion Optimized Sparse Representation

For data compression, the objective function that takes the coding rate into consideration is formulated as follows,

$$\begin{aligned}&\arg\min_{\boldsymbol{A},\boldsymbol{D}}\left\{\frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{A}\|_F^2+\lambda\cdot r\left(\boldsymbol{A}\right)\right\},\\ &\text{s.t. } \|\boldsymbol{A}_i\|_0\leqslant L,\forall i\in\{1,\cdots,K\}\,,\|\boldsymbol{D}_j\|_2^2\leqslant 1,\forall j\in\{1,2,\ldots,M\}\,,\end{aligned} \tag{3}$$

where $\boldsymbol{S}\in\mathbb{R}^{N\times K}$ is the stack of input vectors $\boldsymbol{S_i}$, $\boldsymbol{A}\in\mathbb{R}^{M\times K}$ is the stack of the corresponding sparse coefficient vectors $\boldsymbol{A}_i$, K denotes the number of the input samples. $r(\cdot)$ indicates a function that can represent the coding rate of the coefficients. This formula can be well expressed by the rate-distortion optimization in common image/video compressions [21], where λ controls the relative importance between the rate and distortion. ℓ_0 norm of the coefficients is still critical in order to obtain a sparse approximation. Subsequently, the problem turns to be how to accurately estimate the coding rate and achieve such optimization in sparse coding and dictionary learning.

Based on the *Shannon*'s information theory, the entropy of a data source indicates the average number of bits required to represent it. However, it is difficult to estimate the probability density function of coefficients and formulate the entropy minimization problem. Fortunately, the entropy can be bounded by the variance of data according to the *Shannon* entropy inequality [22],

$$H(\boldsymbol{A})\leqslant\log\left(\sqrt{2\pi eV(\boldsymbol{A})}\right), \tag{4}$$

where $H(\boldsymbol{A})$ and $V(\boldsymbol{A})$ indicate the entropy and the variance of coefficients, respectively. Note that the inequality is tight as the equality holds for Gaussian distributions. Another benefit is that the variance can be feasibly estimated by,

$$V(\boldsymbol{A})=tr\left(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{A}^T\right), \tag{5}$$

where

$$\boldsymbol{Z}=\begin{pmatrix}K-1&\cdots&-1\\ \vdots&\ddots&\vdots\\ -1&\cdots&K-1\end{pmatrix}\in\mathbb{R}^{K\times K}. \tag{6}$$

The diagonal elements in $\boldsymbol{Z}$ equal $K-1$ and others equal to -1.

Due to the *Shannon* entropy inequality and the feasibility of variance estimation, we propose to minimize the variance instead of entropy. The reason lies in that

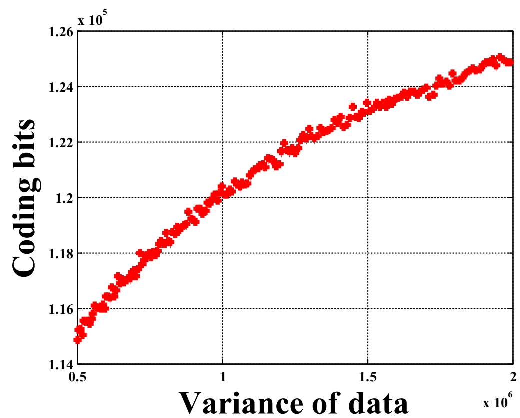


Figure 1: Relationship between the data variance and the corresponding coding bits.

minimizing variance encourages the coefficients to be close to each other, which can be more friendly to compression. To further validate this, we randomly generate data sources and encode them by Huffman coding. The relationship between the variances of the input data and the corresponding coding bits is plotted in Fig. 1, from which one can discern that the variance exhibits a nearly perfect linear relationship to the actual bitrate.

Therefore, the objective function in (3) can be formulated as follows,

$$\begin{aligned}&\arg\min_{\boldsymbol{A},\boldsymbol{D}} \left\{ \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{A}\|_F^2 + \frac{\beta}{2}\, tr\left(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{A}^T\right) \right\}, \\ &\text{s.t. } \|\boldsymbol{A}_i\|_0 \leqslant L, \forall i \in \{1,\cdots,K\}, \|\boldsymbol{D}_j\|_2^2 \leqslant 1, \forall j \in \{1,2,\ldots,M\},\end{aligned} \tag{7}$$

where $\beta/2$ is introduced for computational convenience. Generally speaking, the ℓ_0 norm constraint can be approximately solved by Lagrangian method,

$$\arg\min_{\boldsymbol{A},\boldsymbol{D}} \left\{ \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{A}\|_F^2 + \alpha\|\boldsymbol{A}\|_0 + \frac{\beta}{2}\, tr\left(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{A}^T\right) \right\}, \text{s.t. } \|\boldsymbol{D}_i\|_2^2 \leqslant 1, \forall i \in \{1,2,\ldots,M\}. \tag{8}$$

To solve the non-convex optimization problem effectively, a practical relaxation is to split the problem into two separable parts and update $\boldsymbol{A}$ and $\boldsymbol{D}$ alternately, i.e. the GVCSR based sparse coding and GVCSR based dictionary learning.

2.2 GVCSR based Sparse Coding

The GVCSR based sparse coding given the dictionary $\boldsymbol{D}$ is a subproblem of (8), which can be formulated as follows,

$$\boldsymbol{A} = \arg\min_{\boldsymbol{A}} \left\{ \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{A}\|_F^2 + \alpha\|\boldsymbol{A}\|_0 + \frac{\beta}{2}\, tr\left(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{A}^T\right) \right\}. \tag{9}$$

To solve this, the Alternating Direction Method of Multipliers (ADMM) [20] is employed in this work. Firstly, two auxiliary variables $\boldsymbol{J}$ and $\boldsymbol{G}$ are introduced,

$$\arg\min_{\boldsymbol{A},\boldsymbol{J},\boldsymbol{G}} \left\{ \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{J}\|_F^2 + \alpha\|\boldsymbol{A}\|_0 + \frac{\beta}{2} tr\left(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{G}^T\right) \right\}, \text{s.t. } \boldsymbol{A}=\boldsymbol{J}, \boldsymbol{A}=\boldsymbol{G}, \tag{10}$$

Then the augmented Lagrangian function of (10) can be formulated by,

$$\begin{aligned} \zeta_\mu\left(\boldsymbol{A},\boldsymbol{J},\boldsymbol{G},\boldsymbol{R}\right) = & \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{J}\|_F^2 + \alpha\|\boldsymbol{A}\|_0 + \frac{\beta}{2} tr\left(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{G}^T\right) \\ & + \langle \boldsymbol{A}-\boldsymbol{J}, \boldsymbol{R_0}\rangle + \frac{\mu}{2}\|\boldsymbol{A}-\boldsymbol{J}\|_F^2 + \langle \boldsymbol{A}-\boldsymbol{G}, \boldsymbol{R_1}\rangle + \frac{\mu}{2}\|\boldsymbol{A}-\boldsymbol{G}\|_F^2, \end{aligned} \tag{11}$$

where $\boldsymbol{R} \triangleq [\boldsymbol{R_0}; \boldsymbol{R_1}]$ is the Lagrange multiplier matrix.

The variables $\boldsymbol{A}$, $\boldsymbol{J}$ and $\boldsymbol{G}$ can be alternately updated by minimizing the augmented Lagrangian function ζ with other variables fixed. In this model, each variable can be updated with a closed-form solution. Regarding $\boldsymbol{A}$, it can be updated by,

$$\begin{aligned} \boldsymbol{A} &= \arg\min_{\boldsymbol{A}} \left\{ \alpha\|\boldsymbol{A}\|_0 + \langle \boldsymbol{A}-\boldsymbol{J}, \boldsymbol{R_0}\rangle + \frac{\mu}{2}\|\boldsymbol{A}-\boldsymbol{J}\|_F^2 + \langle \boldsymbol{A}-\boldsymbol{G}, \boldsymbol{R_1}\rangle + \frac{\mu}{2}\|\boldsymbol{A}-\boldsymbol{G}\|_F^2 \right\} \\ &= \arg\min_{\boldsymbol{A}} \left\{ \alpha\|\boldsymbol{A}\|_0 + \frac{\mu}{2}\left\|\boldsymbol{A}-\boldsymbol{J}+\frac{\boldsymbol{R_0}}{\mu}\right\|_F^2 + \frac{\mu}{2}\left\|\boldsymbol{A}-\boldsymbol{G}+\frac{\boldsymbol{R_1}}{\mu}\right\|_F^2 \right\} \\ &= \boldsymbol{H}_{\sqrt{\alpha/\mu}} \left\{ \frac{1}{2}\left(\boldsymbol{J}+\boldsymbol{G}-\frac{\boldsymbol{R_0}+\boldsymbol{R_1}}{\mu}\right) \right\}, \end{aligned} \tag{12}$$

where

$$\boldsymbol{H}_\varepsilon(\boldsymbol{X}) \triangleq \begin{pmatrix} h_\varepsilon(\boldsymbol{X}_{11}) & \cdots & h_\varepsilon(\boldsymbol{X}_{1n}) \\ \vdots & \ddots & \vdots \\ h_\varepsilon(\boldsymbol{X}_{m1}) & \cdots & h_\varepsilon(\boldsymbol{X}_{mn}) \end{pmatrix}. \tag{13}$$

and

$$h_\varepsilon(x) \triangleq \begin{cases} x, & if \quad |x| > \varepsilon \\ 0, & if \quad |x| \leqslant \varepsilon \end{cases} \tag{14}$$

is a hard threshold operator. With respect to $\boldsymbol{J}$ and $\boldsymbol{G}$, we can update them by,

$$\begin{aligned} \boldsymbol{J} &= \arg\min_{\boldsymbol{J}} \left\{ \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{J}\|_F^2 + \langle \boldsymbol{A}-\boldsymbol{J}, \boldsymbol{R_0}\rangle + \frac{\mu}{2}\|\boldsymbol{A}-\boldsymbol{J}\|_F^2 \right\} \\ &= \arg\min_{\boldsymbol{J}} \left\{ \frac{1}{2}\|\boldsymbol{S}-\boldsymbol{D}\boldsymbol{J}\|_F^2 + \frac{\mu}{2}\left\|\boldsymbol{A}-\boldsymbol{J}+\frac{\boldsymbol{R}_0}{\mu}\right\|_F^2 \right\} \\ &= \boldsymbol{V_D}\left(\boldsymbol{\Sigma_D}^T\boldsymbol{\Sigma_D}+\mu\boldsymbol{I}\right)^{-1}\boldsymbol{V_D}^T\left(\boldsymbol{D}^T\boldsymbol{S}+\mu\boldsymbol{A}+\boldsymbol{R_0}\right), \end{aligned} \tag{15}$$

$$\begin{aligned} \boldsymbol{G} &= \arg\min_{\boldsymbol{G}} \left\{ \frac{\beta}{2} tr\left(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{G}^T\right) + \langle \boldsymbol{A}-\boldsymbol{G}, \boldsymbol{R_1}\rangle + \frac{\mu}{2}\|\boldsymbol{A}-\boldsymbol{G}\|_F^2 \right\} \\ &= \arg\min_{\boldsymbol{G}} \left\{ \frac{\beta}{2} tr\left(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{G}^T\right) + \frac{\mu}{2}\left\|\boldsymbol{A}-\boldsymbol{G}+\frac{\boldsymbol{R}_1}{\mu}\right\|_F^2 \right\} \\ &= \beta\left(\mu\boldsymbol{A}+\boldsymbol{R_1}\right)\boldsymbol{V_Z}\left(\boldsymbol{\Sigma_Z}+\mu\boldsymbol{I}\right)^{-1}\boldsymbol{V_Z}^T, \end{aligned} \tag{16}$$

where $\boldsymbol{U_D \Sigma_D V_D^T}$ and $\boldsymbol{U_Z \Sigma_Z V_Z^T}$ are the full Singular Value Decomposition (SVD) of $\boldsymbol{D}$ and $\boldsymbol{Z}$, respectively. Finally, the Lagrangian multiplier $\boldsymbol{R}_0$ and $\boldsymbol{R}_1$ are updated,

$$\boldsymbol{R}_{\boldsymbol{0}}^{j+1} = \boldsymbol{R}_{\boldsymbol{0}}^{j} + \mu^j \left(\boldsymbol{A}^{j+1} - \boldsymbol{J}^{j+1} \right), \tag{17}$$

$$\boldsymbol{R}_{\boldsymbol{1}}^{j+1} = \boldsymbol{R}_{\boldsymbol{1}}^{j} + \mu^j \left(\boldsymbol{A}^{j+1} - \boldsymbol{G}^{j+1} \right), \tag{18}$$

where j indicates the iteration times.

In previous ADMM approach [20], the penalty parameter μ is fixed. To accelerate the convergence, an updating strategy for the penalty parameter is proposed in [23], which can be formulated as follows,

$$\mu^{j+1} = \min \left(\rho \mu^j, \mu_{max} \right), \tag{19}$$

where μ_{max} is an upper bound of the penalty term. $\rho \geqslant 1$ is a constant.

The optimization process of the GVCSR based sparse coding performs iteratively and stops until convergence. In this manner, the globally variance-constrained sparse coding can be achieved. The detailed procedure is presented in Algorithm 1.

Algorithm 1: GVCSR based sparse coding in (9).

Input: Data set $\boldsymbol{S}$, initial dictionary $\boldsymbol{D}$, parameters $\alpha > 0$ and $\beta > 0$.
Output: Sparse representation coefficients $\boldsymbol{A}$.
Initialization:
Set $\boldsymbol{A}^0 = \boldsymbol{J}^0 = \boldsymbol{G}^0 = \boldsymbol{D}^\dagger \boldsymbol{S}, \mu^0 = 1e-2, \mu_{max} = 1e8, \rho = 1.2, \boldsymbol{R_0} = \boldsymbol{R_1} = \boldsymbol{0}, \epsilon = 1e-5$, and $j = 0$;
while *not convergence* **do**
  Fix $\boldsymbol{J}^j$ and $\boldsymbol{G}^j$ to update $\boldsymbol{A}^{j+1}$ by (12);
  Fix $\boldsymbol{A}^{j+1}$ and $\boldsymbol{G}^j$ to update $\boldsymbol{J}^{j+1}$ by (15);
  Fix $\boldsymbol{A}^{j+1}$ and $\boldsymbol{J}^{j+1}$ to update $\boldsymbol{G}^{j+1}$ by (16);
  Update Lagrange multipliers: $\boldsymbol{R}_{\boldsymbol{0}}^{j+1} = \boldsymbol{R}_{\boldsymbol{0}}^{j} + \mu^j \left(\boldsymbol{A}^{j+1} - \boldsymbol{J}^{j+1} \right)$, $\boldsymbol{R}_{\boldsymbol{1}}^{j+1} = \boldsymbol{R}_{\boldsymbol{1}}^{j} + \mu^j \left(\boldsymbol{A}^{j+1} - \boldsymbol{G}^{j+1} \right)$;
  Update penalty parameter μ: $\mu^{j+1} = \min \left(\rho \mu^j, \mu_{max} \right)$;
  $j \leftarrow j + 1$;
  Check convergence: if $\|\boldsymbol{A}^j - \boldsymbol{J}^j\| / \|\boldsymbol{A}^j\| \leqslant \epsilon$ and $\|\boldsymbol{A}^j - \boldsymbol{G}^j\| / \|\boldsymbol{A}^j\| \leqslant \epsilon$ and $\|\boldsymbol{A}^j - \boldsymbol{A}^{j-1}\| / \|\boldsymbol{A}^j\| \leqslant \epsilon$, then stop;
end

2.3 GVCSR based Dictionary Learning

In order to make the dictionary learning consistent with the proposed sparse coding strategy, we further solve the dictionary updating problem based on the overall GVCSR model in (8). The updating process of sparse coefficients $\boldsymbol{A}$ is performed as described in Algorithm 1. After the convergence of $\boldsymbol{A}$, the dictionary $\boldsymbol{D}$ can be updated with the optimized $\boldsymbol{A}$ as follows,

$$\boldsymbol{D} = \arg\min_{\boldsymbol{D}} \left\{ \tfrac{1}{2} \|\boldsymbol{S} - \boldsymbol{D}\boldsymbol{A}\|_F^2 \right\} \text{ s.t. } \|\boldsymbol{D}_i\|_2^2 \leqslant 1, \forall i \in \{1, 2, \ldots, M\}. \tag{20}$$

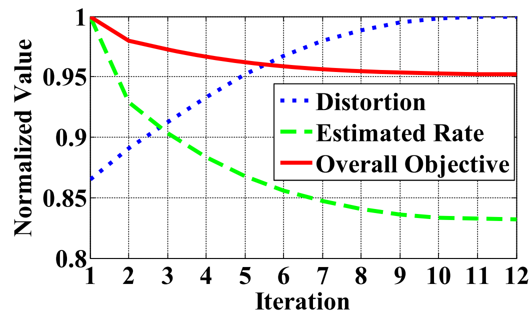


Figure 2: Variations of distortion, estimated rate and the overall objective function value during the ADMM iterations, which are represented by blue dotted line, green dashed line and red solid line, respectively. The horizontal axis denotes the iteration times.

This can be solved via performing SVD on the residuals [10]. Then, the parameters $\boldsymbol{A}$, $\boldsymbol{J}$, $\boldsymbol{G}$ should be further updated according to the new dictionary, while the Lagrangian multipliers $\boldsymbol{R}_0$ and $\boldsymbol{R}_1$ are reset to zero vectors.

3 Experimental Results

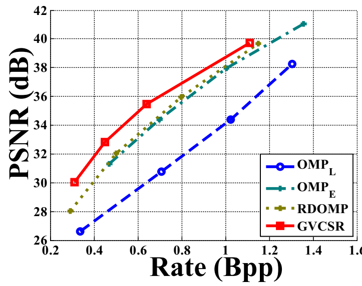
In this section, the performance of the proposed GVCSR model is validated by comparing with other sparse coding and dictionary learning algorithms in image representation. In these experiments, natural images from the public CSIQ dataset [24] are utilized for evaluation. Each image is partitioned into 8×8 non-overlapped blocks for training their dictionary. It is worth mentioning that the rate of sparse coefficients is obtained by performing actual entropy coding during the following comparisons, where the value and position of nonzero coefficients are encoded by run-level method.

Firstly, the changing tendencies of distortion, rate and the overall objective function during the ADMM iterations are shown in Fig. 2. All the values on the three curves are normalized to the same interval for viewing convenience. From this figure we can observe that the proposed method can achieve a better tradeoff in terms of the overall objective function by greatly decreasing the coding rate while keeping the distortions slightly increased. As a result, the coding performance can be improved via the proposed algorithm.

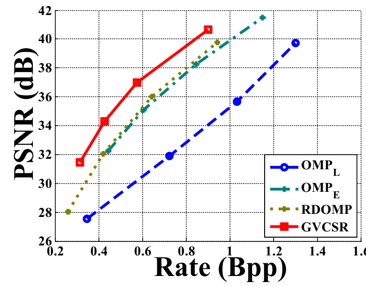
Subsequently, we compare the GVCSR based sparse coding algorithm described in Section 2.2 with other sparse coding approaches. Specifically, three popular sparse decomposition methods are used for comparison, and all of them can be categorized into the separately updated matching pursuit algorithms. The first one is the standard OMP algorithm [25], where the iteration process stops until the ℓ_0 norm of coefficients reaches the limited value L. The second one is similar but the stop criterion is determined by the error energy,

$$\boldsymbol{A}_i = \underset{\boldsymbol{A}_i}{\arg\min} \|\boldsymbol{A}_i\|_0, \text{ s.t. } \|\boldsymbol{S_i} - \boldsymbol{D}\boldsymbol{A}_i\|_2^2 < \epsilon. \tag{21}$$

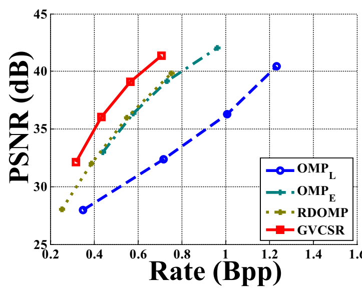
Those two methods are denoted as OMP_L and OMP_E, respectively. In addition, the



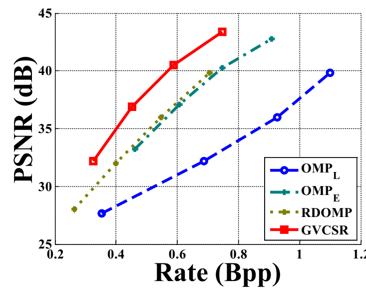
(a) Completeness $\gamma = 24$



(b) Completeness $\gamma = 32$



(c) Completeness $\gamma = 40$



(d) Completeness $\gamma = 48$

Figure 3: Rate-distortion performance comparisons with three matching pursuit based algorithms, in terms of dictionaries with different completeness values.

RDOMP method [16, 17] based on the probability distribution of coefficients is also compared.

In Fig. 3, the rate-distortion comparisons are illustrated in terms of different values of completeness, where the completeness is defined as $\gamma \triangleq \frac{M}{N}$ assuming the dictionary $\boldsymbol{D} \in \mathbb{R}^{N \times M}$. It can be observed that the rate-distortion performance can be significantly improved by the proposed algorithm. This may benefit from the global optimization of the proposed method that jointly considers the distortion and coding rate. Furthermore, from the figure one can also discern that the improvements are more obvious for larger γ. Obviously, the sparse coefficients tend to become more sparse for larger γ. Thereby the matching pursuit based methods may suffer from their potential instability due to the increasing independency among coefficients, while the proposed method can effectively solve this by global optimization to significantly reduce the coding bits.

Finally, we evaluate the performance of the GVCSR based dictionary learning scheme described in Section 2.3 by comparing it with the state-of-the-art algorithms, including MOD [9], KSVD [10], ODL [11] and RLS [12]. For fair comparison, the initial dictionary in each algorithm is the same, which is randomly selected from the training set. The maximum iteration time is also set to be identical. The GVCSR based sparse coding method is employed for generating the sparse coefficients. The averaged results of actual coding rate and Peak Signal to Noise Ratio (PSNR) on the CSIQ dataset are shown in Fig. 4, from which we can see that the proposed scheme can achieve impressive improvements in terms of rate-distortion performance.

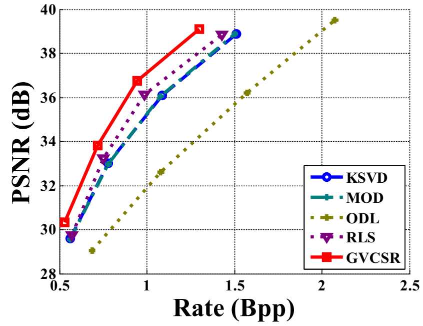


Figure 4: Performance comparisons with the state-of-the-art dictionary learning algorithms, including MOD [9], KSVD [10], ODL [11] and RLS [12].

4 Conclusion

In this work, we present a novel Globally Variance-Constrained Sparse Representation (GVCSR) for rate-distortion joint optimization. To achieve this, a variance-constraint term that can accurately predict the coding rate of the sparse coefficients is introduced into the optimization process. Subsequently, we propose to use the Alternating Direction Method of Multipliers (ADMM) to effectively solve this model in a scientifically sound way. In this manner, the rate-distortion jointly optimized sparse representation can be achieved, leading to higher compression efficiency. Furthermore, experimental results have shown that the GVCSR model can achieve better RD performance comparing with the state-of-the-art sparse coding and dictionary learning algorithms.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grants 61625301, 61571017, 61231002, and 61632001, and in part by the National Basic Research Program of China (973 Program) under Grants 2015CB351800 and 2015CB352502.

References

[1] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[2] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.

[3] X. Zhang, S. Wang, S. Ma, S. Liu, and W. Gao, "Entropy of primitive: A top-down methodology for evaluating the perceptual visual information," in *Visual Communications and Image Processing (VCIP)*, 2013, pp. 1–6.

[4] X. Zhang, S. Wang, S. Ma, and W. Gao, "Towards accurate visual information estimation with entropy of primitive," in *IEEE Int. Symposium on Circuits and Systems (ISCAS)*, 2015.

[5] S. Ma, X. Zhang, S. Wang, J. Zhang, H. Sun, and W. Gao, "Entropy of primitive: From sparse representation to visual information evaluation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. PP, no. 99, pp. 1–1, 2015.

[6] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[7] J. Zhang, D. Zhao, and W. Gao, “Group-based sparse representation for image restoration,” *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, 2014.

[8] X. Zhang, S. Wang, K. Gu, T. Jiang, S. Ma, and W. Gao, “Sparse structural similarity for objective image quality assessment,” in *IEEE Int. Conf. Systems, Man, and Cybernetics*, 2015.

[9] K. Engan, S. O. Aase, and J. Hakon Husoy, “Method of optimal directions for frame design,” in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 5, 1999, pp. 2443–2446.

[10] 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, 2006.

[11] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online Learning for Matrix Factorization and Sparse Coding,” *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.

[12] K. Skretting and K. Engan, “Recursive least squares dictionary learning algorithm,” *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2121–2130, 2010.

[13] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[14] J. Z. Salvatierra, *New sparse representation methods; application to image compression and indexing.* Human-Computer Interaction, 2010.

[15] J. Zhang, S. Ma, R. Xiong, D. Zhao, and W. Gao, “Image primitive coding and visual quality assessment,” in *Advances in Multimedia Information Processing–PCM*, 2012, pp. 674–685.

[16] P. Vandergheynst and P. Frossard, “Adaptive entropy-constrained matching pursuit quantization,” in *IEEE Int. Conf. Image Processing (ICIP)*, vol. 2, 2001, pp. 423–426.

[17] T. Ryen, G. M. Schuster, and A. K. Katsaggelos, “A rate-distortion optimal coding alternative to matching pursuit,” in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 2002, pp. III–2177.

[18] 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, 2014.

[19] X. Liu, D. Zhai, D. Zhao, and W. Gao, “Image super-resolution via hierarchical and collaborative sparse representation,” in *Data Compression Conference (DCC)*. IEEE, 2013, pp. 93–102.

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations & Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[21] G. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov 1998.

[22] C. E. Shannon, *A mathematical theory of communication.* McGraw-Hill,, 1974.

[23] Z. Lin, R. Liu, and Z. Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” in *Advances in neural information processing systems*, 2011, pp. 612–620.

[24] E. C. Larson and D. M. Chandler, “Categorical image quality (CSIQ) database.” [Online]. Available: http://vision.okstate.edu/?loc=csiq

[25] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.