A Consensus Framework for Convolutional Dictionary Learning based on L1 Norm Error

Mizuki Takanashi* and Yoshimitsu Kuroki†
* National Institute of Technology (KOSEN), Kurume College, Fukuoka, Japan
E-mail: a3711mt@kurume.kosen-ac.jp
† E-mail: kuroki@kurume-nct.ac.jp

***Abstract*—Convolutional Sparse Representations (CSRs) approximate an original signal with a sum of convolutions of dictionary filters and their coefficient maps. Convolutional Dictionary Learning (CDL) is a problem to get a set of convolutional dictionaries for CSRs. An effective way to get high fidelity dictionaries for any images is using enormous images for learning; however, there is a limitation of memory capacity for normal CDL. This paper tackles robust dictionary design with the l_1 norm error on the error term instead of the l_2 norm, which is generally used for CDL, for an enormous number of learning images. Furthermore, our method employs a consensus framework to decrease the memory consumption. The number of learning images without the consensus frame work for dictionary learning is up to about 100 at most, but our method obtains the dictionaries using more learning images: 1,000 and 10,000 in the experiments. As for the dictionary fidelity, the dictionary designed with the l_1 error term for 100 test images generates about 3 dB higher PSNR images than that with the l_2 error term at equivalent sparseness of coefficients.**

I. Introduction

Sparse Representation for images [1] is used in various fields: pattern recognition, computer vision, and image processing. A sparse representation approximates a vectorized signal $\boldsymbol{s} \in \mathbb{R}^N$ as $\boldsymbol{s} \approx D\boldsymbol{x}$, where $D \in \mathbb{R}^{N\times M}$ is a dictionary matrix, and $\boldsymbol{x} \in \mathbb{R}^M$ is the corresponding coefficient having only a few non-zero elements. To compute sparse representation on high dimension images, the approximation is applied for non-over lapped blocks to reduce computational load, but the generated dictionary filters are not robust against image shift.

An alternative representation is Convolutional Sparse Representation (CSR) [2], which approximates $s_k \in \mathbb{R}^N$ as

$$\boldsymbol{s}_k \approx \sum_{m=1}^{M} \boldsymbol{d}_m * \boldsymbol{x}_{k,m}, \tag{1}$$

where $\boldsymbol{d}_m \in \mathbb{R}^L\,(L < N)$ is a dictionary filter and $\boldsymbol{x}_{k,m} \in \mathbb{R}^N$ is the corresponding coefficient map. Hereafter, we call a set of filters a dictionary. Please note that CSR models an entire signal with the sum of the pairs of the M common filters and their coefficient maps. To obtain both the filters and the maps in CSR, most algorithms minimize a cost function consisting of a weighted sum of a error and a regularization terms. The former and the latter terms are usually estimated by the l_2 norms and the l_1 norms respectively. It is infeasible to obtain simultaneously both dictionary filters and corresponding coefficients, and then general solutions divide the optimization process into two sub-optimization procedures: the filter optimization for fixed coefficients and the coefficient optimization for fixed dictionary. Both procedures are shown as the following;

$$\underset{\{\boldsymbol{d}_m\}}{\arg\min}\ \frac{1}{2}\sum_{k=1}^{K}\left\|\sum_{m=1}^{M}\boldsymbol{d}_m * \boldsymbol{x}_{k,m} - \boldsymbol{s}_k\right\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{d}_m\|_2 \le 1, \tag{2}$$

$$\underset{\{\boldsymbol{x}_{k,m}\}}{\arg\min}\ \frac{1}{2}\sum_{k=1}^{K}\left\|\sum_{m=1}^{M}\boldsymbol{d}_m * \boldsymbol{x}_{k,m} - \boldsymbol{s}_k\right\|_2^2 + \lambda\sum_{m=1}^{M}\sum_{k=1}^{K}\|\boldsymbol{x}_{k,m}\|_1 \quad \text{s.t.} \quad \|\boldsymbol{d}_m\|_2 \le 1, \tag{3}$$

where K is the number of learning images.

To design a dictionary which well expresses features for any images, it is effective to increase the number of learning images. However, the amount of memory consumption also increases, and then [3] proposes a consensus framework to suppress the memory consumption on the cost function using the l_2 error term for a large number of learning images. Another method to get a better dictionary is employing the l_1 norm error instead of the l_2 norm. Some works reported that CDL based on the l_1 norm error improves the accuracy of the designed dictionary [4] [5] [6]. The CDL problem with the l_1 norm error is defined as

$$\underset{\{\boldsymbol{d}_m\}}{\arg\min}\ \sum_{k=1}^{K}\left\|\sum_{m=1}^{M}\boldsymbol{d}_m * \boldsymbol{x}_{k,m} - \boldsymbol{s}_k\right\|_1 \quad \text{s.t.} \quad \|\boldsymbol{d}_m\|_2 \le 1. \tag{4}$$

In this study, we propose a consensus framework of CDL with the l_1 norm error.

II. Convolutional Dictionary Learning

A. CDL based on l_1 norm error

A dictionary designed with the l_1 norm errror is obtained by solving (4). The convolution of the filters with the coefficients can be written in the simple form as

$$\sum_{m=1}^{M}\boldsymbol{d}_m * \boldsymbol{x}_{k,m} = X_k\boldsymbol{d}, \tag{5}$$

by defining $X_k \in \mathbb{R}^{N\times NM}$ and $\boldsymbol{d} \in \mathbb{R}^{NM}$ as

$$X_k = \begin{pmatrix} x_{k,m,1} & x_{k,m,N} & \cdots & x_{k,m,2} \\ \vdots & x_{k,m,1} & \ddots & \vdots \\ \vdots & \vdots & \ddots & x_{k,m,N} \\ x_{k,m,N} & x_{k,m,N-1} & \cdots & x_{k,m,1} \end{pmatrix}, \quad (6)$$

$$\boldsymbol{d} = \begin{pmatrix} P\boldsymbol{d}_1 \\ \vdots \\ P\boldsymbol{d}_M \end{pmatrix}, \quad (7)$$

where $x_{k,m,i}$ is the i-th element of the vector $\boldsymbol{x}_{k,m}$. The matrix $P \in \mathbb{R}^{NM\times NM}$ is a zero padding matrix, which extends dictionary filters to the same size as that of the coefficients:

$$P\boldsymbol{d}_m = \begin{pmatrix} \boldsymbol{d}_m \\ 0 \end{pmatrix}, \quad (8)$$

and enables to calculate the convolution by the Hadamard product in the Fourier domain. Then, by defining $X \in \mathbb{R}^{NK\times NM}$ and $\boldsymbol{s} \in \mathbb{R}^{NK}$ as

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix}, \quad \boldsymbol{s} = \begin{pmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_K \end{pmatrix}, \quad (9)$$

we can rewrite (4) in the simple form:

$$\arg\min_{\{\boldsymbol{d}\}} \|X\boldsymbol{d} - \boldsymbol{s}\|_1 \quad \text{s.t.} \quad \boldsymbol{d} \in C_{PN}, \quad (10)$$

where the constraint set C_{PN} is given by

$$C_{PN} = \{\boldsymbol{d} \in \mathbb{R}^{NM} : \left(I - PP^T\right)\boldsymbol{d} = \boldsymbol{0}, \|\boldsymbol{d}\|_2 \leq 1\}. \quad (11)$$

In this equation, $I \in \mathbb{R}^{NM\times NM}$ is the identity matrix.

One of the conventional CDL algorithms uses a convex solver ADMM (Alternating Direction Method of Multipliers) [7]. We can rewrite (4) to apply ADMM as

$$\begin{aligned} &\arg\min_{\{\boldsymbol{d}\},\{\boldsymbol{g}_0\},\{\boldsymbol{g}_1\}} \|\boldsymbol{g}_0\|_1 + \iota_{C_{PN}}(\boldsymbol{g}_1) \\ &\quad \text{s.t.} \quad \begin{pmatrix} X \\ I \end{pmatrix}\boldsymbol{d} - \begin{pmatrix} \boldsymbol{g}_0 \\ \boldsymbol{g}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{s} \\ 0 \end{pmatrix}, \end{aligned} \quad (12)$$

where ι is the indicator function defined as

$$\iota_A(\boldsymbol{x}) = \begin{cases} 0 & \text{if} \quad \boldsymbol{x} \in A \\ \infty & \text{if} \quad \boldsymbol{x} \notin A \end{cases}. \quad (13)$$

The ADMM procedures at iteration i for (12) is shown as the following fives steps:

$$\boldsymbol{d}^{i+1} = \arg\min_{\{\boldsymbol{d}\}} \|X\boldsymbol{d} - \boldsymbol{g}_0^i - \boldsymbol{s} + \boldsymbol{h}_0^i\|_2^2 + \|\boldsymbol{d} - \boldsymbol{g}_1^i + \boldsymbol{h}_1^i\|_2^2, \quad (14)$$

$$\boldsymbol{g}_0^{i+1} = \arg\min_{\{\boldsymbol{g}_0\}} \|\boldsymbol{g}_0\|_1 + \frac{\rho}{2}\|X\boldsymbol{d}^{i+1} - \boldsymbol{g}_0 - \boldsymbol{s} + \boldsymbol{h}_0^i\|_2^2, \quad (15)$$

$$\boldsymbol{g}_1^{i+1} = \arg\min_{\{\boldsymbol{g}_1\}} \iota_{C_{PN}}(\boldsymbol{g}_1) + \frac{\rho}{2}\|\boldsymbol{d}^{i+1} - \boldsymbol{g}_1 + \boldsymbol{h}_1^i\|_2^2, \quad (16)$$

$$\boldsymbol{h}_0^{i+1} = \boldsymbol{h}_0^i + X\boldsymbol{d}^{i+1} - \boldsymbol{g}_0^{i+1} - \boldsymbol{s}, \quad (17)$$

$$\boldsymbol{h}_1^{i+1} = \boldsymbol{h}_1^i + \boldsymbol{d}^{i+1} - \boldsymbol{g}_1^{i+1}, \quad (18)$$

where $\rho \in \mathbb{R}$ is a parameter. The solution of (14) shown as

$$(X^TX - I)\boldsymbol{d}^{i+1} = X^T(\boldsymbol{g}_0^i + \boldsymbol{s} - \boldsymbol{h}_0^i) + (\boldsymbol{g}_1^i - \boldsymbol{h}_1^i) \quad (19)$$

is numerically relieved by the Sherman-Morrison formula as

$$\begin{aligned} \boldsymbol{d}^{i+1} = (\boldsymbol{g}_1^i - \boldsymbol{h}_1^i) - X^T(I + XX^T)^{-1} \\ (\boldsymbol{g}_0^i + \boldsymbol{s} - \boldsymbol{h}_0^i - X(\boldsymbol{g}_1^i - \boldsymbol{h}_1^i)). \end{aligned} \quad (20)$$

The efficient approach to solve the above equation is to compute it in the Fourier domain. The $\boldsymbol{g}_0$ update is obtained via the soft-thresholding [8]:

$$\boldsymbol{g}_0 = S_{1/\rho}\left(X\boldsymbol{d}^{i+1} - \boldsymbol{s} + \boldsymbol{h}_0^i\right), \quad (21)$$

where the soft-thresholding is given by

$$S_\gamma(X) = \text{sign}(X) \odot \max(0, \|X\| - \gamma). \quad (22)$$

In (16), $\boldsymbol{g}_1$ is optimized by using the proximal operator defined as

$$\boldsymbol{g}_1 = \text{prox}_{C_{PN}}(\boldsymbol{d}^{i+1} + \boldsymbol{h}_1^i), \quad (23)$$

where the proximal operator is

$$\text{prox}_{C_{PN}}(\boldsymbol{d}) = \begin{cases} PP^T\boldsymbol{d} & \text{if } \left\|PP^T\boldsymbol{d}\right\|_2 \leq 1 \\ \text{otherwise} & \end{cases}. \quad (24)$$

The five steps continue until the variables converge.

Resulting dictionaries with the l_1 norm error is robust against outliers compared to the l_2 norm error and less sensitive to images with extreme features.

B. Consensus Framework

A CDL problem is typically formulated as (2). To obtain $\boldsymbol{d}$ with ADMM, a temporal variable $\boldsymbol{g}$ is involved as

$$\frac{1}{2}\arg\min_{\{\boldsymbol{d}\},\{\boldsymbol{g}\}} \|X\boldsymbol{d} - \boldsymbol{s}\|_2^2 + \iota_{C_{PN}}(\boldsymbol{g}) \quad \text{s.t.} \quad \boldsymbol{d} = \boldsymbol{g}. \quad (25)$$

Dictionary design with a large amount of images requires enormous memory since the size of X is $NK \times NM$, although it is effective to design the dictionary accurately. For a solution to this problem, [3] proposes a consensus framework to CDL on the l_2 error. The consensus framework designs the dictionary $\boldsymbol{d}_k$ for each signal $\boldsymbol{s}_k$; thereby, (25) is divided into K subproblems on index k as

$$\arg\min_{\{\boldsymbol{d}_k\},\{\boldsymbol{g}\}} \frac{1}{2}\|X_k\boldsymbol{d}_k - \boldsymbol{s}_k\|_2^2 + \iota_{C_{PN}}(\boldsymbol{g}) \quad \text{s.t.} \quad \boldsymbol{d}_k = \boldsymbol{g}. \quad (26)$$

To combine each subproblem for the whole learning images, we modify X_{con} and $\boldsymbol{d}_{con}$ as

$$X_{con} = \begin{pmatrix} X_1 & 0 & 0 & \cdots \\ 0 & X_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (27)$$

$$\boldsymbol{d}_k = \begin{pmatrix} P\boldsymbol{d}_{1,k} \\ \vdots \\ P\boldsymbol{d}_{M,K} \end{pmatrix}, \quad \boldsymbol{d}_{con} = \begin{pmatrix} \boldsymbol{d}_1 \\ \vdots \\ \boldsymbol{d}_K \end{pmatrix}, \quad (28)$$

where $\boldsymbol{d}_k \in \mathbb{R}^{NM}$ is the dictionary corresponding to the input signal $\boldsymbol{s}_k$. Then, (25) in the standard ADMM form is given by

$$\begin{aligned} &\underset{\{\boldsymbol{d}_{con}\},\{\boldsymbol{g}\}}{\arg\min}\ \frac{1}{2}\|X_{con}\boldsymbol{d}_{con} - \boldsymbol{s}\|_2^2 + \iota_{C_{PN}}(\boldsymbol{g}) \\ &\qquad\qquad \text{s.t.} \quad \boldsymbol{d}_{con} - E\boldsymbol{g} = 0, \end{aligned} \tag{29}$$

where matrix $E \in \mathbb{R}^{NMK \times NM}$ is defined as

$$E = \begin{pmatrix} I & I & \cdots \end{pmatrix}^T. \tag{30}$$

The updates of the ADMM procedure are as the followings:

$$\begin{aligned} \boldsymbol{d}_{con}^{i+1} = \underset{\{\boldsymbol{d}_{con}\}}{\arg\min}\ &\frac{1}{2}\|X_{con}\boldsymbol{d}_{con} - \boldsymbol{s}\|_2^2 + \\ &\frac{\rho}{2}\|\boldsymbol{d}_{con} - E\boldsymbol{g}^i + \boldsymbol{h}_{con}^i\|_2^2, \end{aligned} \tag{31}$$

$$\boldsymbol{g}^{i+1} = \underset{\{\boldsymbol{g}\}}{\arg\min}\ \iota_{C_{PN}}(\boldsymbol{g}) + \frac{\rho}{2}\|\boldsymbol{d}_{con}^{i+1} - E\boldsymbol{g} + \boldsymbol{h}_{con}^i\|_2^2, \tag{32}$$

$$\boldsymbol{h}_{con}^{i+1} = \boldsymbol{d}_{con}^{i+1} - E\boldsymbol{g}^{i+1} + \boldsymbol{h}_{con}^i. \tag{33}$$

Again, the $\boldsymbol{d}_{con}$ update step can be computed for each input signal $\boldsymbol{s}_k$ as

$$\boldsymbol{d}_k^{i+1} = \underset{\{\boldsymbol{d}_k\}}{\arg\min}\ \frac{1}{2}\|X_k\boldsymbol{d}_k - \boldsymbol{s}_k\|_2^2 + \frac{\rho}{2}\|\boldsymbol{d}_k - \boldsymbol{g}^i + \boldsymbol{h}_k^i\|_2^2, \tag{34}$$

and the solution is

$$\boldsymbol{d}_k^{i+1} = \left(X_k^\mathsf{T} X_k + \rho \boldsymbol{I}\right)^{-1}\left(X_k^\mathsf{T}\boldsymbol{s}_k + \rho\left(\boldsymbol{g}^i - \boldsymbol{h}_k^i\right)\right). \tag{35}$$

Same as (20), we can apply the Sherman-Morrison formula to calculate the inverse matrix as

$$\begin{aligned} \boldsymbol{d}_k^{i+1} = \left(\boldsymbol{g}^i - \boldsymbol{h}_k^i\right) - X_k^\mathsf{T}&\left(X_k X_k^\mathsf{T} + \rho\boldsymbol{I}\right)^{-1} \\ &\left\{X_k\left(\boldsymbol{g}^i - \boldsymbol{h}_k^i\right) - \boldsymbol{s}_k\right\}. \end{aligned} \tag{36}$$

Note that memory consumption for computing (36) does not vary with the number of learning images. Equation (32) has the form of the ADMM consensus problem, thus the $\boldsymbol{g}$ update step can be rewritten as

$$\begin{aligned} \boldsymbol{g}^{i+1} = &\underset{\{\boldsymbol{g}\}}{\arg\min}\ \iota_{C_{PN}}(\boldsymbol{g}) + \\ &\frac{K\rho}{2}\left\|\boldsymbol{g} - K^{-1}\left(\sum_{k=1}^{K}\boldsymbol{d}_k^{i+1} + \sum_{k=1}^{K}\boldsymbol{h}_k^i\right)\right\|_2^2, \end{aligned} \tag{37}$$

and the solution is obtained via the proximal operator;

$$\boldsymbol{g}^{i+1} = \mathrm{prox}_{C_{PN}}\left(K^{-1}\left(\sum_{k=1}^{K}\boldsymbol{d}_k^{i+1} + \sum_{k=1}^{K}\boldsymbol{h}_k^i\right)\right). \tag{38}$$

Since E is the set of a identity matrix, we can update $\boldsymbol{h}$ for every $\boldsymbol{h}_k$ of the learning image $\boldsymbol{s}_k$ independently.

III. Proposed Method

In this study, we apply the consensus framework to CDL based on the l_1 norm error. The normal form of CDL with the l_1 norm error is shown as (4). Same as the application of ADMM to CDL with the l_2 norm error, we design the dictionary $\boldsymbol{d}_k$ for each input signal $\boldsymbol{s}_k$. An optimization problem is formulated as

$$\begin{aligned} &\underset{\{\boldsymbol{d}_k\},\{\boldsymbol{g}_{0k}\},\{\boldsymbol{g}_1\}}{\arg\min}\ \|\boldsymbol{g}_{0k}\|_1 + \iota_{C_{PN}}(\boldsymbol{g}_1) \\ &\qquad \text{s.t.} \begin{pmatrix} X_k \\ I \end{pmatrix}\boldsymbol{d}_k - \begin{pmatrix} \boldsymbol{g}_{0k} \\ \boldsymbol{g}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{s}_k \\ 0 \end{pmatrix}. \end{aligned} \tag{39}$$

By using the block diagonal matrix X_{con} and the vector $\boldsymbol{d}_{con}$ defined in (27) and (28), the above problem is rewritten in the standard ADMM form as

$$\begin{aligned} &\underset{\{\boldsymbol{d}_{con}\},\{\boldsymbol{g}_{0con}\},\{\boldsymbol{g}_1\}}{\arg\min}\ \|\boldsymbol{g}_{0con}\|_1 + \iota_{C_{PN}}(\boldsymbol{g}_1) \\ &\quad \text{s.t.} \begin{pmatrix} X_{con} \\ I \end{pmatrix}\boldsymbol{d} - \begin{pmatrix} \boldsymbol{g}_{0con} \\ E\boldsymbol{g}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{s} \\ 0 \end{pmatrix}. \end{aligned} \tag{40}$$

The update steps are followings:

$$\begin{aligned} \boldsymbol{d}_{con}^{i+1} = \underset{\{\boldsymbol{d}_{con}\}}{\arg\min}\ &\|X_{con}\boldsymbol{d}_{con} - \boldsymbol{g}_{0con}^i - \boldsymbol{s} + \boldsymbol{h}_{0con}^i\|_2^2 + \\ &\|\boldsymbol{d}_{con} - E\boldsymbol{g}_1^i + \boldsymbol{h}_{1con}^i\|_2^2, \end{aligned} \tag{41}$$

$$\begin{aligned} \boldsymbol{g}_{0con}^{i+1} = &\underset{\{\boldsymbol{g}_{0con}\}}{\arg\min}\ \|\boldsymbol{g}_{0con}\|_1 + \\ &\frac{\rho}{2}\|X_{con}\boldsymbol{d}_{con}^{i+1} - \boldsymbol{g}_{0con} - \boldsymbol{s} + \boldsymbol{h}_{0con}^i\|_2^2, \end{aligned} \tag{42}$$

$$\boldsymbol{g}_1^{i+1} = \underset{\{\boldsymbol{g}_1\}}{\arg\min}\ \iota_{C_{PN}}(\boldsymbol{g}_1) + \frac{\rho}{2}\|\boldsymbol{d}_{con}^{i+1} - E\boldsymbol{g}_1 + \boldsymbol{h}_{1con}^i\|_2^2, \tag{43}$$

$$\boldsymbol{h}_0^{i+1} = \boldsymbol{h}_0^i + X_{con}\boldsymbol{d}_{con}^{i+1} - \boldsymbol{g}_{0con}^{i+1} - \boldsymbol{s}, \tag{44}$$

$$\boldsymbol{h}_{1con}^{i+1} = \boldsymbol{h}_{1con}^i + \boldsymbol{d}_{con}^{i+1} - \boldsymbol{E}g_1^{i+1}. \tag{45}$$

Since X_{con} is the block diagonal matrix, $\boldsymbol{d}_{con}$ can be computed for every input signal as

$$\begin{aligned} \boldsymbol{d}_k^{i+1} = \underset{\{\boldsymbol{d}_k\}}{\arg\min}\ &\|X_k\boldsymbol{d}_k - \boldsymbol{g}_{0k}^i - \boldsymbol{s}_k + \\ &\boldsymbol{h}_{0k}^i\|_2^2 + \|\boldsymbol{d}_k - \boldsymbol{g}_1^i + \boldsymbol{h}_{1k}^i\|_2^2. \end{aligned} \tag{46}$$

This Equation has the same form as (14), and thus the solution is obtained in the same way as

$$\begin{aligned} \boldsymbol{d}_k^{i+1} = (\boldsymbol{g}_1^i - \boldsymbol{h}_{1k}^i) - X_k^T&(I + X_k X_k^T)^{-1} \\ (\boldsymbol{g}_{0k}^i + \boldsymbol{s}_k - &\boldsymbol{h}_{0k}^i - X_k(\boldsymbol{g}_1^i - \boldsymbol{h}_{1k}^i)). \end{aligned} \tag{47}$$

Similarly, the $\boldsymbol{g}_0$ update step is formulated for every input signal as

$$\begin{aligned} \boldsymbol{g}_{0k}^{i+1} = &\underset{\{\boldsymbol{g}_{0k}\}}{\arg\min}\ \|\boldsymbol{g}_{0k}\|_1 + \\ &\frac{\rho}{2}\|X_k\boldsymbol{d}_k^{i+1} - \boldsymbol{g}_{0k} - \boldsymbol{s}_k + \boldsymbol{h}_{0k}^i\|_2^2, \end{aligned} \tag{48}$$

and the solution is

$$\boldsymbol{g}_{0k} = S_{1/\rho}\left(X_k\boldsymbol{d}_k^{i+1} - \boldsymbol{s}_k + \boldsymbol{h}_{0k}^i\right). \tag{49}$$

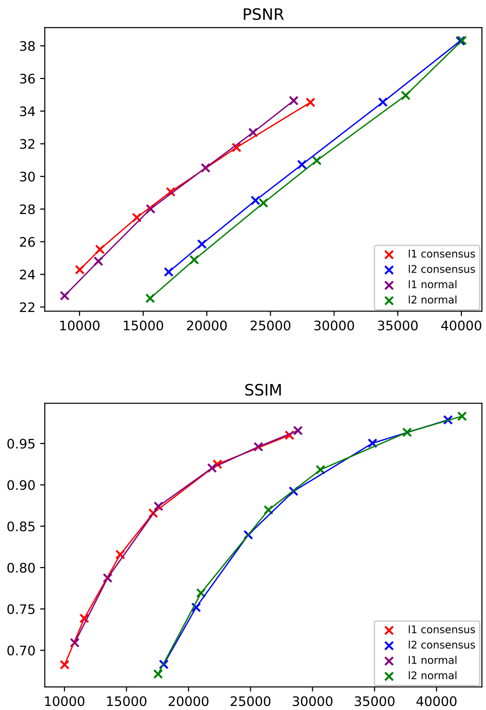

Fig. 1: l_0 norm versus image quality with 100 test images.

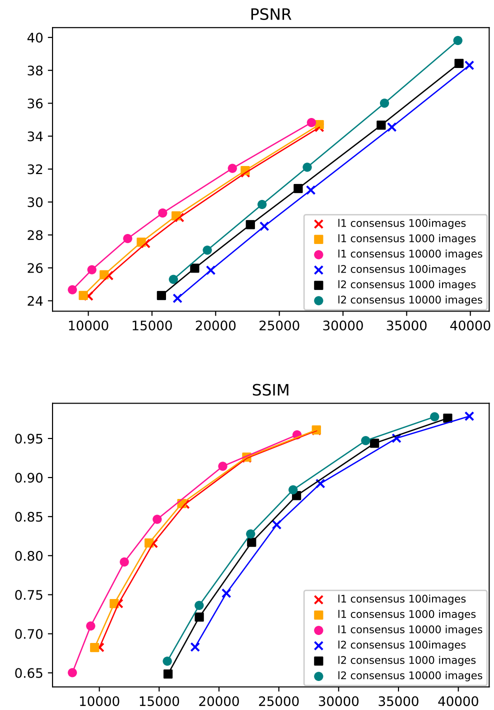

Fig. 2: Reconstructed image quality using the consensus form of CDL with varying the number of learning images: 100, 1,000, and 10,000.

To solve the $\boldsymbol{g}_1$ update step, we apply the consensus ADMM problem to (43) as

$$\boldsymbol{g}_1^{i+1} = \arg\min_{\{\boldsymbol{g}_1\}} \iota_{C_{PN}}(\boldsymbol{g}_1) + \frac{K\rho}{2}\left\| \boldsymbol{g}_1 - K^{-1}\left(\sum_{k=1}^{K} \boldsymbol{d}_k^{i+1} + \sum_{k=1}^{K} \boldsymbol{h}_{1k}^{i}\right)\right\|_2^2 . \tag{50}$$

This problem is solved via the proximal operator as

$$\mathrm{prox}_{C_{PN}}\left(K^{-1}\left(\sum_{k=1}^{K} \boldsymbol{d}_k^{i+1} + \sum_{k=1}^{K} \boldsymbol{h}_{1k}^{i}\right)\right). \tag{51}$$

The $\boldsymbol{h}_0$ and the $\boldsymbol{h}_1$ update steps are solved for every signal independently, since X is the block diagonal matrix and E is the set of the identity matrix.

We compute these update steps in the Fourier domain to compute the convolution by the Hadamard product. The space complexity for the proposed method is $O(MN)$, while that for normal CDL is $O(KMN)$.

IV. Experiments

We compared the four types of CDL: consensus form of CDL based on the l_1 norm error (l_1 consensus), the l_2 norm error (l_2 consensus), the normal form of CDL based on the l_1 norm error (l_1 normal) and the l_2 norm error (l_2 normal). To compare the performance of dictionaries designed by each method, we reconstructed images with these dictionaries and compared the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) between original testing images and the reconstructed images. The testing images did not used for the dictionary design. We employed " SPORCO " [9] for the implement.

Fig. 3: Examples of reconstructed images.

The experiments used ImageNet [10], which includes more than 14 million labeled color images. We used 100,000 images of the dataset for learning images, and 100 images for testing images. For preprocessing, we converted the images to gray scale images, cropped into square, and resized them to 256×256 arrays. The number of dictionary filters is 6, and its size is 32×32 pixels. We determined the parameters as $\rho = 50$ and $\lambda = 1$ via experiments using the small number of images.

Fig. 1 and Fig. 2 compare the 4 methods in the l_0 norm of the coefficient maps in the horizontal axes, where vertical axes shows reconstructed test image quality (PSNR or SSIM). Fig. 1 graphs the performance of the dictionaries generated each method with 100 learning images. This result tells us that the consensus-framework CDLs work equivalently to the normal CDLs. Fig. 2 shows the performance of the dictionaries with the consensus form of CDL with varying the number of learning images. The normal methods both in the l_2 and the l_1 norms failed to obtain dictionaries due to the memory limitation. The result also shows that the more images we use for learning, the higher quality images all the methods obtain; furthermore, the dictionaries generated with the l_1 error outperforms those of the l_2 error. Fig. 3 illustrates examples of reconstructed images with the l_1 and the l_2 errors.

V. Conclusion

In this study, we propose the method to compute CDL based on the l_1 norm error for an enormous number of learning images by employing the consensus framework. The proposed method generates equivalent dictionaries with the normal CDL, and the reconstructed image quality outperforms those using the l_2 error. The number of learning images of the normal CDL is about 100 at most due to the memory limitation, but that of the proposed method exceeds the limitation.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 20K11878.

References

[1] J. Mairal, F. R. Bach and J. Ponce, ”Sparse Modeling for Image and Vision Processing,” Now Foundations and Trends, 2014, DOI: 10.1561/0600000058.

[2] B. Wohlberg, ”Efficient Algorithms for Convolutional Sparse Representations,” in IEEE Transactions on Image Processing, vol.25, no.1, pp.301-315, Jan. 2016, doi: 10.1109/TIP.2015.2495260.

[3] C. Garcia-Cardona and B. Wohlberg, ”Convolutional Dictionary Learning: A Comparative Review and New Algorithms,” IEEE Transactions on Computational Imaging, vol.4, no.3, pp.366-381, Sep. 2018, DOI:10.1109/TCI.2018.2840334.

[4] S. Wang, Q. Liu, Y. Xia, P. Dong, J. Luo, Q. Huang, and D.D. Feng, ”Dictionary learning based impulse noise removal via L1–L1 minimization,” Signal Processing, vol.93, no.9, pp.2696-2708, 2013, ISSN 0165-1684, DOI:10.1016/j.sigpro.2013.03.005.

[5] S. Mukherjee, R. Basu, and C. S. Seelamantula, ”l1-K-SVD: A robust dictionary learning algorithm with simultaneous update,” Signal Processing, vol.123, pp.42-52, 2016, ISSN 0165-168, DOI:10.1016/j.sigpro.2015.12.008.

[6] B. Wohlberg, ”Convolutional sparse representations as an image model for impulse noise restoration,” 2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), Bordeaux, France, 2016, pp.1-5, DOI:10.1109/IVMSPW.2016.7528229.

[7] S. Boyd N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and sta-tistical learning via the alternating direction method of multipliers,” Foundations and Trends in Machine learning vol.3, no.1, pp.1-122, 2010, DOI: 10.1561/2200000016.

[8] N. Parikh and S. Boyd, “Proximal algorithms,” Foundations and Trends in Optimization, vol.1, no.3, pp.127–239, 2014. DOI:10.1561/2400000003.

[9] B. Wohlberg, ”SPORCO”, http://brendt.wohlberg.net/software/SPORCO/, July 2021.

[10] O. Russakovsky*, J. Deng*, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg and L. Fei-Fei ”ImageNet Large Scale Visual Recognition Challenge,” IJCV, 2015.