Journal of Network Intelligence ©2020 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **5**, Number **2**, May **2020**

Dictionary Compression for Network Transmission Using Flexible Quantization Steps

Lei Liu, Yao Zhao, Chunyu Lin*, and Huihui Bai
Institute of Information Science
Beijing Jiaotong University, Beijing, China
*Correspending: cylin@bjtu.edu.cn

Received March 2019; revised December 2019

Abstract. *Dictionary learning is applied to various image and video processing tasks. However, the volume of the learned dictionary from large amounts of data leads it expensive to storage and transmit in networks. In this paper, we proposed a flexible dictionary compression scheme for face recognition. Firstly, a quantized K-SVD (Q-KSVD) algorithm for each dictionary basis (or atom) is proposed. To find the optimal dictionary subjected by the total bits, an algorithm is developed to select the bases according to their importance, denoted by the rate-recognition slopes of all bases. Moreover, a flexible quantization stepsize assignment method is proposed to further improve the compression efficiency. Each dictionary basis is taken as a small image so that we can employ the mature image coding method, and the basis quantization is integrated into various K-SVD-based dictionary learning algorithms. The dictionary can be compressed into various scales of bits, which is quite adaptable to the varied network environments. Face recognition results using four K-SVD-based algorithms show that the proposed scheme can achieve competitive performance with low bit rate.*
Keywords: Dictionary learning, Sparse representation, Quantized K-SVD, Compressed dictionary.

1. **Introduction.** Learning an overcomplete dictionary for sparse representation of signals has attracted growing interests in recent years. As learned from the signal directly, the redundant dictionary can offer a sparser representation for the signal than the fixed transforms such as the discrete cosine transform (DCT) or wavelet [1, 2, 3]. This makes dictionary learning based sparse representation popular in various applications such as image restoration [4, 5], image denoising [6], image/video compression [7, 8, 9, 10], image super-resolution [11, 12], and face/object classification [13, 14]. With the popular of the dictionary learning based spare representation, the dictionary need to be compressed in order to transmit in the networks.

With the rapid growth of signal acquisition, the dictionary learned from large amounts of data becomes quite large. This increases the complexity of the dictionary, as the sparsity-based algorithms such as matching pursuit (MP) [15] and orthogonal matching pursuit (OMP) [16], all comprise expensive dictionary-signal computations repeatedly. On the other hand, the large volume of the dictionary is hard to storage or transmit, especially in the mobile devices which with limited storage space and computational power. In [17], the K-SVD algorithm is proposed to effectively learn the dictionary form the training images, and the effectiveness is shown for the image denoising and compression. K-SVD is a popular dictionary learning algorithm and many variations or extensions have been proposed to improve the performance of the learned dictionary. In [18], an efficient and flexible dictionary structure is proposed to balance the complexity and adaptability of the dictionary, via a sparsity model of the dictionary atoms over a base dictionary. The dictionary is assumed to be represented by a fixed dictionary with sparse coefficient, *i.e.*, $D = \Phi A$, where Φ is the so-called base dictionary and A is the sparse code of the dictionary. Though the dictionary learning algorithms progressively improve the representation performance of the learned dictionary [19, 20, 21], it is still inadequate for some special tasks such as image classification, as only the representational capability is considered in the dictionary learning process, while the discriminative capability is neglected which is crucial for the recognition tasks. To deal with this problem, a variety of algorithms have been developed to combine the discriminative power and the representative power into the learned dictionary simultaneously. In [22], a discriminative K-SVD (D-KSVD) algorithm is proposed to learn a discriminative dictionary for face classification. The reconstruction error and the classification error are incorporated into the objective function together, thus the dictionary and the classifier are learned simultaneously. Training with the classifier, the learned dictionary obtains the discriminative power implicitly. In [23, 24], the label consistent K-SVD (LC-KSVD) algorithm is proposed, which incorporates the reconstruction error, the discriminative sparse coding error, and the classification error into the objective function together. It can learn the dictionary, the discriminative sparse code, and the classifier simultaneously. The discriminative sparse coding enforce the input image and the dictionary atom from the same class share a same label, thus the learned dictionary is endowed with the discriminative power explicitly. Owing to exploiting the label information of images, the learned discriminative dictionary can achieve more accurate recognition result.

Usually, the dictionary columns (known as atoms or bases) need to be normalized to have a unit l_2-norm. The floating-point dictionary is then directly stored and used for the following sparse decomposition. For one hand, the floating-point data type would consume a mass of memory space; for another hand, this would bring about massive overhead and extensive delay if the dictionary needs to communicate in the network environment. For example, a dictionary with a size of 570×576 learned by the K-SVD algorithm requires about 2.5 megabytes (MB) of storage space. For an overcomplete dictionary-based light field photography [25], the memory footprint of the learned dictionary consisting of 5,000 light field atoms is more than 110 MB. Some works attempt to eliminate the atoms with less importance in order to reduce the atom number [26, 27, 28], however, it is still far from adequate by saving the learned dictionary directly as floating-point numbers. Even though the dictionary can be sparse represented [18], there is still no explicit compression schemes for the dictionary. In [29], a quantized K-SVD (Q-KSVD) algorithm is proposed to reduce the dictionary storage space. This method treats the dictionary atoms as small images and uses the conventional image coding techniques to compress the dictionary atoms during the training process. However, this method applies a fixed quantization stepsize to all atoms, which is not optimal for different atoms, as each atom has variable importance at the sparse coding stage. The effective way to compress the dictionary still needs to be carefully investigated.

In this paper, we improve the compression scheme to further reduce the memory footprint needed to store the learned dictionary and reinforce its adaptability to the varied network environments. The dictionary atoms are treated as small images, and we use the conventional image method to compress them. Before normalization in the dictionary update stage, the dictionary atom is first converted to an image with values in the interval between $[0 \sim 255]$. A quantization stepsize assignment method is proposed to assign appropriate quantization stepsize for each basis, according to the contribution of the basis in the task. The dictionary image is then compressed by the conventional image coding techniques with the chosen quantization stepsizes [30, 31]. In order to guarantee the validity of the dictionary after compression, we combine the dictionary encoding steps with the update stage of the dictionary learning algorithms. Inspired by the scalable coding method used in JPEG 2000 [32], a bit rate threshold can be imposed on the compressed dictionary to restrict the amount of bits needed to encode the dictionary. By selecting the bases with higher importance (e.g., bases with larger rate-recognition slopes), the compressed dictionary ensures its optimality as far as the given bit rate target is attained. As shown in the experimental results, this scheme can get the optimum number of dictionary bases in terms of bit rate constraint, so that we can get various bit scales for different network requirements. This is quite preferred by the mobile network applications as the mobile network is heterogeneous and the bandwidth is vulnerable to changes. This dictionary compression approach can applied to various K-SVD-based dictionary learning algorithms. Finally, it is the reconstructed and normalized dictionary after scaling back that is used for the subsequent operations. The dictionary learned by the Q-KSVD method is evaluated on the face database of Extended YaleB. The test results show the proposed approach can come up with some sort of compromise between the recognition accuracy and the dictionary bit rate. It can obtain acceptable recogonition accuracy at very low bit rate of the compressed dictionary. Under some circumstances, it can even provide more precise recognition than the original dictionary.

2. **Background.** For a set of m-dimensional signals $Y = [y_1, ..., y_N] \in R^{m \times N}$, to find the sparse coefficients X of Y under a representative dictionary D with K bases, we can solve the following problem:

$$< D, X > = \arg\min_{D,X} \|Y - DX\|_2^2, \quad s.t. \ \forall i, \|x_i\|_0 \leq T, \tag{1}$$

where $D = [d_1, ..., d_K] \in R^{m \times K}$ is the learned dictionary ($K > m$), $X = [x_1, ..., x_N] \in R^{K \times N}$ is the coefficients, and T is the sparsity constraint for the coefficients.

To minimize the representative error under the sparsity constraint, the K-SVD algorithm can find the solution of the problem efficiently by an alternative manner [17]. Firstly, given D, the sparse coding stage tries to search for the sparse representation X through the solution of the following problem:

$$i = 1, 2, ..N, \ x_i = \arg\min_{x_i} \|Y - Dx_i\|_2^2, \quad s.t. \ \|x_i\|_0 \leq T. \tag{2}$$

Orthogonal matching pursuit (OMP) [16] can solve this problem efficiently. Secondly, after finding X, the coefficients are fixed and the dictionary will be updated. The update stage is accomplished by applying the singular value decomposition (SVD) on the bases of D one after another.

To balance the complexity and adaptability of the learned dictionary, a sparse dictionary structure [18] is proposed under the sparsity model of D over a base dictionary, *i.e.*, $D = \Phi A$, where Φ is fixed and called base dictionary, and A is sparse code of D over Φ.

Though many algorithms are developed to improve the representative power of the learned dictionary, it is not enough for the image recognition tasks as the dictionary is short of the discriminative power for different images from multiple classes. To solve this problem, many discriminative dictionary learning algorithms are proposed to explore the label information of the images from different classes. For example, by combining the classification error and the reconstruction error into the objective function together, the D-KSVD algorithm [22] is proposed as follows:

$$< D, W, X > = \arg\min_{D,W,X} \|Y - DX\|_2^2 + \beta\|H - WX\|_2^2, \quad s.t. \ \forall i, \|x_i\|_0 \leq T, \tag{3}$$

where H is the label matrix for images from various classes and W is the learned classifier. The first penalty term $\|Y - DX\|_2^2$ denotes the reconstruction error, and the second penalty term $\|H - WX\|_2^2$ denotes the classification error. β is a parameter to balance the two penalty terms. By combining $(Y, \sqrt{\beta}H)$ and $(D, \sqrt{\beta}W)$, Eq. (3) can be restructured into the standard K-SVD format:

$$< D, W, X > = \arg \min_{D,W,X} \left\| \begin{pmatrix} Y \\ \sqrt{\beta}H \end{pmatrix} - \begin{pmatrix} D \\ \sqrt{\beta}W \end{pmatrix} X \right\|_2^2, \quad s.t.\ \forall i, \|x_i\|_0 \leq T. \tag{4}$$

Besides the representative power, the learned dictionary also obtains the discriminative power, as the the objective function considers the representative and discriminative penalties at the same time.

Furthermore, by combing the discriminative sparse code error, together with the classification error and the reconstruction error, into the objective function, the LC-KSVD algorithm [23, 24] is proposed as follows:

$$\begin{aligned} < D, W, A, X > = \arg \min_{D,W,A,X} \|Y - DX\|_2^2 + \alpha\|C - AX\|_2^2 + \beta\|H - WX\|_2^2, \\ s.t.\ \forall i, \|x_i\|_0 \leq T, \end{aligned} \tag{5}$$

where C is a matrix with element $c_{i,j} = 1$, if the image y_i and the dictionary basis d_j share the same label (called label consistency); otherwise $c_{i,j} = 0$. $\|C - AX\|_2^2$ denotes the discriminative sparse code error, and A is a linear transform matrix, which enforces the images from the same class to have similar sparse coefficients. Also, Eq. (5) can be restructured into a standard K-SVD format:

$$< D, W, A, X > = \arg \min_{D,W,A,X} \left\| \begin{pmatrix} Y \\ \sqrt{\alpha}C \\ \sqrt{\beta}H \end{pmatrix} - \begin{pmatrix} D \\ \sqrt{\alpha}A \\ \sqrt{\beta}W \end{pmatrix} X \right\|_2^2, \quad s.t.\ \forall i, \|x_i\|_0 \leq T. \tag{6}$$

Both Eq. (4) and (6) become a K-SVD problem and can be solved directly.

3. **Quantized K-SVD with Flexible Quantization Stepsize.** In the literature of dictionary learning based sparse representation, each basis of the dictionary is normalized to a unit l_2-norm, and the dictionary is then simply saved in the form of floating-point numbers. When the amount of training data is large, the dictionary trained will be with a large number of bases, which could take up a lot of storage space. A naive way to directly compress the learned dictionary will be not optimal and the dictionary coding error will be susceptible to the following implementations. In this section, we consider how to make the learned dictionary compressible by modifying the K-SVD algorithm with quantization manipulations in the training process.

In order to reduce the memory space of the learned dictionary, the dictionary needs to be quantized before encoding, and the reconstructed version of the compressed dictionary is used for the following operations. As the dictionary does not need to be stored during the training, we disregard the entropy coding at this moment. Therefore we only consider the quantization of the dictionary during the dictionary training process.

Like the D-KSVD and LC-KSVD introducing extra discriminative and label consistent penalty terms into the objective function, we introduce the quantization penalty term into the K-SVD problem, and the objective function is formulated as follows:

$$< D_Q, X > = \arg \min_{D_Q,X} \|Y - D_Q X\|_2^2, \quad s.t.\ \forall i, \|x_i\|_0 \leq T, \tag{7}$$

where D_Q indicates the quantized dictionary after being reconstructed. Considering the quantization during the training process, the quantization distortion for the dictionary is constricted implicitly with the representational error, making the quantized dictionary more resilient in the sparse coding and the following operations.

For the K-SVD algorithm and its successors, the bases of the dictionary are updated one by one, and each basis is normalized after the update. To encode the basis of the dictionary by the available image compression techniques, we rescale each basis before its normalization into the interval between $[0 \sim 255]$ and consider it as a 2-D image. Then we employ the transform and quantization to the rescaled basis. It is the reconstructed versions, *i.e.*, the de-quantized and inverse transformed dictionary bases, are put into the following dictionary update. At the end of the dictionary training, we can entropy encode the quantized dictionary.

Furthermore, a bit rate constraint can be imposed on the dictionary, similar to the rate-distortion optimization problem in data compression. The objective function of the quantized dictionary with bit rate constraint is formulated as follows:

$$< D_Q, X > = \arg \min_{D_Q, X} \|Y - D_Q X\|_2^2, \quad s.t. \ \forall i, \|x_i\|_0 \leq T, \ \ R(D_Q) \leq T_R, \tag{8}$$

where D_Q indicates the compressed dictionary after being reconstructed, $R(D_Q)$ indicates the bits needed to encode the dictionary, and T_R is the network constraint using bits representation.

For the subsequent applications, we can decode each basis of the dictionary by the inverse operations. Each basis of the reconstructed dictionary is then normalized to a unit l_2-norm.

It is convenient to apply a fixed quantization stepsize to the coefficients of all bases. However, this method is not optimal as it neglects the variable significance of each basis. Thus, a flexible quantization stepsize assignment is desired to get optimal dictionary compression result. For this purpose, we define $f_i = ||x_i||_0$, where x_i is the sparse code corresponding to the i-th basis d_i in the dictionary, to denote the significance of d_i. A larger f_i means that the corresponding d_i is used more frequently in the sparse coding and of more significance, we then impose a small quantization stepsize to d_i to keep its stability, and vice versa. In order to make the bit rate of the final compressed dictionary within the given target, some bases from the original dictionary need to be discarded. Inspired by the rate-distortion-slope algorithm utilized in JPEG 2000 [32], we define a similar rate-recognition-slope metric to indicate the contribution of each basis for the recognition task. The importance of the i-th basis is evaluated by the slope s_i between the classification error and the bits needed to encode it:

$$s_i = \frac{||H - W'X'||_2^2}{R_i}, \tag{9}$$

where H is the class label matrix, W' is the classifier W excluding the i-th basis (column) w_i, X' is the sparse code X excluding the i-th row x_T^i, and R_i is the bits needed to encode d_i. Excluding the product of $w_i x_T^i$ can reflect the contribution of the corresponding basis d_i to the image classification. Note that, in this case, we need the entropy coding step during the training, as we need the bits information to calculate the contribution of each basis.

When all the rate-recognition slopes of the bases are calculated, we sort the slopes and select the bases with the largest slope values, as far as the bit rate target is achieved. The rest of the bases will be discarded. Then the next iteration of the dictionary training will be continued with the selected bases. To speed up the convergence, we preset two termination conditions as follows, and the algorithm will be terminated if either of them is reached:

$$\begin{aligned} RMSE(i+1) - RMSE(i) &\leq T_1, \\ Recog(i+1) - Recog(i) &\leq T_2, \end{aligned} \tag{10}$$

where $RMSE = \sqrt{\|Y - DX\|_2^2/(m \times N)}$ indicates the average reconstruction error, *Recog* is the classification accuracy for the current training images, and T_1 and T_2 are two threshold values. Finally, the Q-KSVD with flexible quantization stepsizes is summarized in Algorithm 1. Note that, if a fixed quantization stepsize is assigned to all the bases, the algorithm becomes the case of [29].

Algorithm 1 Quantized K-SVD with Flexible Quantization Stepsize

Input: Y, H, T_R
Output: D_Q, W

```
compute D^(0), X^(0), W^(0):
    compute D^(0), X^(0) using original K-SVD;
    compute W^(0) using (11);
repeat
    reshape each d_i into a 2-D image in the range of [0 ~ 255], transform it by DCT;
    compute f_i = ||x_i||_0, assign a quantization stepsize to the coefficients of d_i accordingly;
    entropy encode the quantized coefficients;
    compute s_i using (9);
    rank s_i, select from atom with max s_i, until the total bits reach T_R;
    reconstruct D_Q;
    update D_Q using original K-SVD;
    update W using (11);
until convergence condition in (10)
```

Once the dictionary is trained by the Q-KSVD, the bases can be rescaled to 2-D images and compressed by transform, quantization and entropy coding, just the same as the compression steps during the training process. For image classification, it is the reconstructed dictionary that is employed for the sparse coding of the test images.

The proposed dictionary compression scheme aforementioned is also suitable for other K-SVD-based algorithms, such as the D-KSVD [22] and the LC-KSVD [23, 24]. At this point, we only focus on the compression of the dictionary D but neglect A and W, as the discriminative transform matrix does not need to store and the classifier does not need to compress.

4. **Experimental Results.**

4.1. **Image Classification.**

For comparison, we first use the dictionary learned by the K-SVD algorithm for face recognition, and the classifier is learned as follows: :

$$W = (XX^T + \lambda I)^{-1} X H^T. \tag{11}$$

where X is the sparse coefficients trained by the original K-SVD, H is the class label matrix, and I is the identify matrix.

For D-KSVD and LC-KSVD, the trained dictionary D and classifier W cannot be directly used for a new image, as they are normalized jointly in the learning process. The

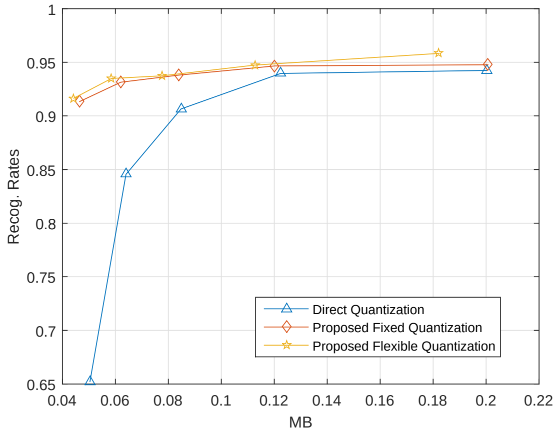

FIGURE 1. Recognition results by the direct and proposed dictionary compression schems with D-KSVD [22].

expected dictionary $\hat{D}$ and classifier $\hat{W}$ can be computed as follows [22]:

$$\begin{aligned} \hat{D} &= \{\frac{d_1}{\|d_1\|_2}\cdots\frac{d_K}{\|d_K\|_2}\}, \\ \hat{W} &= \{\frac{w_1}{\|d_1\|_2}\cdots\frac{w_K}{\|d_K\|_2}\}. \end{aligned} \tag{12}$$

For a test image y_i, the classification is based on the sparse coefficient x_i, which can be obtained by Eq. (2) with $\hat{D}$. Then, we can apply the classifier $\hat{W}$ to x_i and get the label vector of y_i:

$$l = \hat{W}x_i. \tag{13}$$

Finaly, the label of the testing image y_i is determined by the index of the largest element in l.

4.2. **Results with the Extended YaleB Database.** In this subsection, we incorporate the proposed dictionary compression scheme into four dictionary learning algorithms: K-SVD [17], D-KSVD [22], LC-KSVD1, and LC-KSVD2 [23, 24] (LC-KSVD contains two scenerios: LC-KSVD1 does not include the classification error term in the objective function and LC-KSVD2 includes this term). We test the proposed scheme through face classification problem, and the Extended YaleB database is adopted for the test. The database contains 2414 face images of 38 persons, about 64 images for each on average. The face images are pre-cropped and normalized with a size of 192×168. The feature used here is randomface [13], which projects the face image into a column with a randomly generated matrix, and the length of the feature is set to 576. The database is randomly splitted into two halves. One half contains 32 images per person is used for training the dictionary and the other is used for testing. The number of the dictionary bases is 570, and each basis has a length of 576. The weighting parameters in Eq. (3) and (5) are set to $\alpha = 16$ and $\beta = 4$. The sparsity assumed for the face image is set to $T = 30$. The termination thresholds are set to $T_1 = -0.05$ and $T_2 = 0.005$. A lapped transform-based codec [30, 31] with various quantization stepsizes (Qsteps) is employed for the dictionary compression.

Firstly, we test the validity of the compressed dictionary with D-KSVD [22], compared to the direct compression method, without any bit rate constraints. For the direct compression method, we compress the dictionary only once after the training of D-KSVD with the same compression steps used in the training process, e.g., rescale the basis into

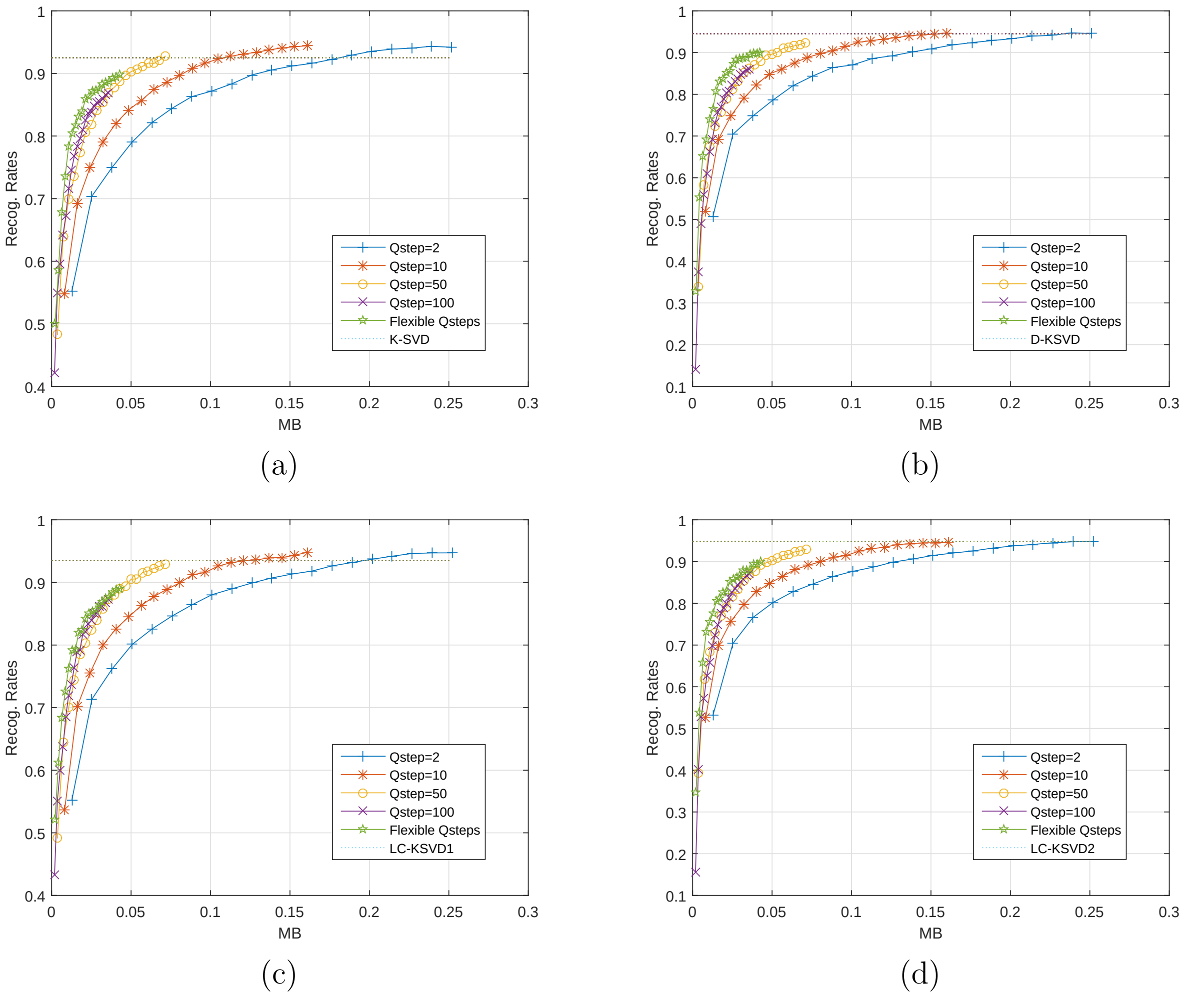

FIGURE 2. Recognition results by the compressed dictionary. (a) K-SVD; (b) D-KSVD; (c) LC-KSVD1; (d) LC-KSVD2.

an image, transform, quantization, and entropy coding. For the proposed method, we conduct a fixed quantization method, *i.e.*, assign a fixed stepsize to all bases, as a special case as [29] do. For the direct and fixed quantization methods, Qsteps = (5, 20, 40, 60, 80) are slected for test, respectively. For the flexible method, the quantization stepsize arrays which contain stepsize candidates near the given Qsteps are utilized. The results are shown in Fig. 1. From the results we can see that, the direct compression method is vulnerable to the coding distortion, especially at very low bit rates (or high Qsteps), as the recognition accuracy drops down quickly. On the other hand, the proposed dictionary compression scheme is quite robust to the coding distortion, as the recognition accuracy keeps at a high level with the decrease of the bit rate. This is maninly owing to the integration of the dictionary compression into the dictionary learning.

Secondly, we compress the bases with different Qsteps during the dictionary training, under the constraint of given bit rates. We calculate the rate-recognition slopes of all bases with the given Qstep, and select from the bases with largest slopes to meet the bit rate target. Both fixed and flexible quantization methods are conducted here. For fixed quantization, Qsteps = (2, 10, 50, 100) are adopted. The recognition results are shown in Fig. 2 and Fig. 3. As can be seen from Fig. 2, different tradeoffs between the recognition rate and the bit rate of the compressed dictionary can be achieved by the

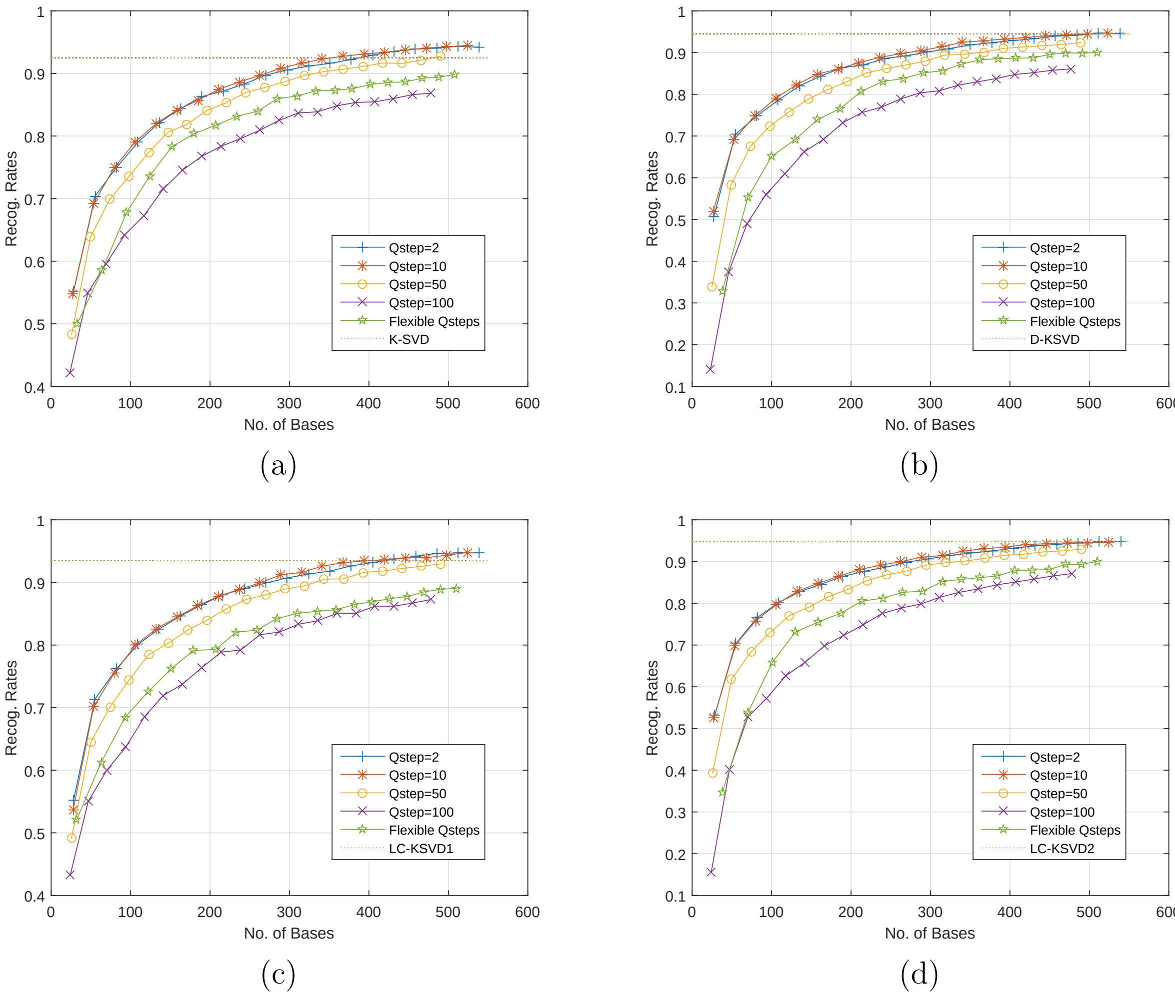

Figure 3. Recognition results by the compressed dictionary with different number of bases. (a) K-SVD; (b) D-KSVD; (c) LC-KSVD1; (d) LC-KSVD2.

proposed scheme, which is convenient for the dictionary with bit rate constraint. When Qsteps = (2, 10), and even 50, the proposed scheme can still achieve high recognition results, compared to the original dictionary. Sometimes the recognition results can even outperform the original ones, while the bit rates are much lower. This is may because that the noise in the learned dictionary can be removed by quantization to some content. However, with the increase of the quantization stepsize, the rate-recognition performance cannot continue to improve. This may because that the dictionary compression distortion caused by a large quantization stepsise cannot be eliminated by the training. However, the flexible quantization method can still achieve superior performance, especially at the low bit rate area. As can be seen from Fig. 3, the proposed scheme can adjust the dictionary size to adapt the bit rate constraint. For Qsteps = (2, 10, 50), even the number of dictionary bases reduced from 570 to about a half, it still achieves very high recognition results.

5. **Conclusions.** We propose a dictionary compression scheme for sparse representation, which can find the optimal dictionary when there is a bit rate constraint imposed on the dictionary. The basis compression steps are integrated into the dictionary training process, and the scheme is suitable for various K-SVD-based algorithms. Both fixed and flexible

quantization methods can achieve well-balanced performance. The fixed quantization is more simplicity while the flexible quantization is more competitive at low bit rate. Experiments on face recognition show the validity of the proposed scheme under various bit rate constraint, which is adaptable to the network applications. Future works include finding fine grained quantization stpsizes for different dictionary bases to further improve the performance and conducting the compressed dictionary to various applications.

Acknowledgment. This work is partially supported by the Fundamental Research Funds for the Central Universities (no. 2018JBZ001).

REFERENCES

[1] R. Rubinstein, A. M. Bruckstein, and M. Elad, Dictionaries for sparse representation modeling, *Proc. of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.
[2] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, Sparse representation for computer vision and pattern recognition, *Proc. of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
[3] I. Tosic and P. Frossard, Dictionary learning, *IEEE Signal Process. Mag.*, vol. 19, no. 7, pp. 27–38, 2011.
[4] W. Dong, L. Zhang, G. Shi, and X. Li, Nonlocally centralized sparse representation for image restoration, *IEEE Trans. Image Proc.*, vol. 22, no. 4, pp. 1620–1630, 2013.
[5] J. Mairal, M. Elad, and G. Sapiro, Sparse representation for color image restoration, *IEEE Trans. Image Process.*, vol. 17, no. 1, pp. 53–69, 2008.
[6] M. Elad and M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, *IEEE Trans. Image Proc.*, vol. 15, no. 12, pp. 3736–3745, 2006.
[7] H. Yang, W. Lin, and C. Deng, Learning based screen image compression, *Proc. of IEEE Int. Workshop on Multimedia Signal Proc.*, pp. 78–82, Banff, AB, Canada, 2012.
[8] J.-W. Kang, C.-C. Kuo, R. Cohen, and A. Vetro, Efficient dictionary based video coding with reduced side information, *Proc. of IEEE Int. Symp. Circ. and Syst.*, pp. 109–112, Rio de Janeiro, Brazil, 2011.
[9] M. Xu, S. Li, J. Lu, and W. Zhu, Compressibility constrained sparse representation with learnt dictionary for low bit-rate image compression, *IEEE Trans. Circ. Syst. Video Tech.*, vol. 24, no. 10, pp. 1743–1757, 2014.
[10] H. Xiong, Z. Pan, X. Ye, and C. W. Chen, Sparse spatio-temporal representation with adaptive regularized dictionary learning for low bit-rate video coding, *IEEE Trans. Circ. Syst. Video Tech.*, vol. 23, no. 4, pp. 710–728, 2013.
[11] J. Yang, J. Wright, T. S. Huang, and Y. Ma, Image super-resolution via sparse representation, *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, 2010.
[12] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. S. Huang, Coupled dictionary training for image super-resolution, *IEEE Trans. Image Process.*, vol. 21, no. 8. pp. 3467–3478, 2012.
[13] J. Wright, A. Y. Yang, A. Ganesh, S. S. Satry, and Y. Ma, Robust face recognition via sparse representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
[14] Y-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, Learning mid-level features for recognition, *Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 2559–2566, 2010.
[15] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.
[16] J. Tropp and A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
[17] M. Aharon, M. Elad, and A. Bruckstein, K-SVD: an algorithm for design overcomplete dictionaries for sparse representation, *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
[18] R. Rubinstein, M. Zibulevsky, and M. Elad, Double sparsity: learning sparse dictionaries for sparse signal approximation, *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, 2010.
[19] J. Sulam, V. Papyan, Y. Romano, and M. Elad, Multi-layer convolutional sparse modeling: pursuit and dictionary learning, *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4090–4104, 2018.
[20] V. Papyan, Y. Romano, J. Sulam, and M. Elad, Theoretical foundations of deep learning via sparse representations, *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 72–89, 2018.

[21] Y. Yankelevsky and M. Elad, Finding GEMS: multi-scale dictionaries for high-dimensional graph signals, *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1889–1901, 2019.

[22] Q. Zhang and B. Li, Discriminative K-SVD for dictionary learning in face recognition, *Proc. of IEEE Int. Conf. on Computer vision and Pattern Recognition*, San Francisco, CA, pp. 2691–2698, 2010.

[23] Z. Jiang, Z. Lin, and L. S. Davis, Learning a discriminative dictionary for sparse coding via label consistent K-SVD, *Proc. of IEEE Conf. on Computer vision and Pattern Recognition*, Colorado Springs, CO, pp. 1697–1704, 2011.

[24] Z. Jiang, Z. Lin, and L. S. Davis, Label consistent K-SVD: Learning a discriminative dictionary for recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[25] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, Compressive light field photography using overcomplete dictionaries and optimized projections, *ACM Trans. Graphics (TOG) - SIGGRAPH 2013 Conference Proceedings*, vol. 32, no. 4, pp. 1–12, 2013.

[26] R. Mazhar and P. D . Gader, EK-SVD: Optimized dictionary design for sparse representations, *Proc. of Int. Conf. on Pattern Recognition*, Tampa, FL, pp. 1–4, 2008.

[27] C. Rusu and B. Dumitrescu, Stagewise K-SVD to design efficient dictionaries for sparse representation, *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 631–634, 2012.

[28] M. Marsousi, K. Abhari, P. Babyn, and J. Alirezaie, An adaptive approach to learn overcomplete dictionaries with effieient numbers of elements, *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3272–3283, 2014.

[29] L. Liu, J. Liang, Y. Zhao, C. Lin, and H. Bai, Quantized dictionary for sparse representation, *Proc. of IEEE Int. Workshop on Multimedia Signal Proc.*, pp. 1–6, Xiamen, China, 2015.

[30] T. D. Tran, J. Liang, and C. Tu, Lapped transform via time-domain pre- and post-filtering, *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1557–1571, 2003.

[31] C. Tu and T. D. Tran, Context based entropy coding of block transform coefficients for image compression, *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1271–1283, 2002.

[32] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image compression fundamentals, standards, and practices*, Kluwer, Norwell, MA, 2002.