Rate-Distortion Based Sparse Coding for Image Set Compression

Xinfeng Zhang†, Weisi Lin†, Siwei Ma$^{\ddagger *}$, Shiqi Wang‡ , Wen Gao$^{\ddagger *}$

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

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

* *Peking University Shenzhen Graduate School, Shenzhen, China*

†{xfzhang, wslin}@ntu.edu.sg, $^{\ddagger *}${swma, sqwang, wgao}@pku.edu.cn

***Abstract*—In this paper, we propose a novel image set compression approach based on sparse coding with an ordered dictionary learned from perceptually informative signals. For a group of similar images, one representative image is first selected and transformed into wavelet domain, and then its AC components are utilized as samples to train an over-complete dictionary. In order to improve compression efficiency, the dictionary atoms are reordered according to their frequency used in sparse approximation of the representative image. In addition, a rate-distortion based sparse coding method is proposed to distribute atoms among different image patches adaptively. Experimental results show that the proposed method outperforms JPEG and JPEG2000 up to 6+ dB and 2+ dB, respectively.**

***Index Terms*—Image compression, dictionary, sparse coding, rate-distortion**

I. Introduction

With the explosive increase of images, image compression plays a more and more important role in image storage and transmission. Traditional image compression achieves high compact image representation by reducing various redundancies within images, e.g., spatial redundancy and psychovisual redundancy. However, with the development of cloud storage, a large amount of images with similar contents are uploaded to servers, and this wastes a lot of space due to set redundancy existing among similar images.

Existing image set compression methods usually reduce set redunce by inter-prediction between different images [1]. Due to the irregular variations between similar images, e.g., large scale motions, luminance changes, the tradition methods are difficult to get high efficient prediction via local motion estimation. Recently, image sparse coding with learned over-complete dictionaries shows promising results on image compression [2] [3] by representing images with dictionary atoms compactly. In this compression framework, a dictionary is firstly learned from a lot of images, and then an image is compressed by representing its non-overlapping image patches as linear combination of very few dictionary atoms, which is called sparse coding. In [4], Skretting and Engan compared compression efficiency with pixel domain dictionary and wavelet domain dictionary and proposed a more general sparse coding based compression method with wavelet domain dictionary, which is obtained by the Recursive Least Squares Dictionary Learning Algorithm (RLS-DLA). Since the sparse approximation with learned dictionary is more efficient for images, which are similar with those used in dictionary learning, image compression with learned dictionaies can reduce set redundancy efficiently by sparsing coding image patches, while avoid the limitation of local motion estimation. Therefore, sparse coding with learned dictionary is more suitable for solving the image set compression problem, which compresses a group of similar images and improve compression performance by reducing set redundancy.

In this paper, we propose a rate-distortion based sparse coding method to deal with the image set compression problem and reduce the set redundancy by an ordered dictionary. For a group of similar images, one representative image is firstly selected and compressed with traditional method, e.g. JPEG. Then, the decoded representative image is further transformed into wavelet domain and divided into overlapping patches with removing DC components, which are utilized as samples for dictionary learning. In order to reduce statistical redundancy, we reorder the dictionary atoms according to their frequency in sparse coding for the representative image, which makes the frequently used atoms centralized to the front. In addition, considering different sparsity of image patches, we propose a rate-distortion based orthogonal matching pursuit (RD-OMP) method to approximate image by progressively increasing atoms for different image patches according to their rate-distortion costs.

The remainder of this paper is organized as follows. In Section 2, we briefly review the background of sparse coding based image compression. Section 3 first introduces the proposed sparse coding based image set compression framework, and then introduces dictionary reordering and the rate-distortion based sparse coding method. Experimental results are reported in Section 4 and Section 5 concludes the paper.

II. Review Of Sparse Coding Based Image Compression

Sparse representation with learned dictionary compresses images by describing image patches with linear combinations of rarely dictionary atoms. Considering the good adaptability of learned over-complete dictionaries, many sparse coding based image compression methods have been proposed in literatures, e.g., [5], [6]. In general, the over-complete dictionary is learned from a collection of image patches, $\{\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_L\}$, by solving the following optimization problem,

$$f(\mathbf{D}, \boldsymbol{\alpha}) = \min_{\mathbf{D},\boldsymbol{\alpha}} \sum_i^L \|\boldsymbol{y}_i - \mathbf{D}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_0, \quad s.t.\|\boldsymbol{d}_i\|_2 \le 1. \tag{1}$$

Here $\boldsymbol{y}_i \in R^M$ is the vectorization of the i^{th} image patch and $\mathbf{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, , \boldsymbol{d}_N] \in R^{M\times N}(M < N)$ is the learned dictionary with N atoms $\{\boldsymbol{d}_i\}$. $\alpha_i \in R^N$ is the coefficient vector corresponding to image patch $\boldsymbol{y}_i$, in which there is very few non-zero values, and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_L]$. λ is the regularization parameter to trade off the data fitting term (first term of Eqn.(1)) and the regularized sparsity term (second term of Eqn.(1)).

The joint optimization problem in Eqn.(1) can be solved by iteratively optimizing the sparse coding coefficient matrix $\boldsymbol{\alpha}$ and dictionary $\mathbf{D}$, respectively, i.e., solving the sub-optimization problems in Eqn.(2) and Eqn.(3):

$$\boldsymbol{\alpha}^k = \min_{\boldsymbol{\alpha}} \sum_i^L \|\boldsymbol{y}_i - \mathbf{D}^{k-1}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_0, \tag{2}$$

$$\mathbf{D}^k = \min_{\mathbf{D}} \sum_i^L \|\boldsymbol{y}_i - \mathbf{D}\alpha_i^k\|_2^2, \tag{3}$$

where α^k and $\mathbf{D}^k$ represent the k^{th} iteration results. Many optimization algorithms have been proposed to solve the above problems in Eqn.(2) and Eqn.(3), e.g., orthogonal matching pursuit (OMP) [7] for the sparse coding in Eqn.(2), and RLS-DLA [8] for the dictionary update in Eqn.(3).

For sparse coding based image compression, one image, $\mathcal{I}$, is firstly divided into non-overlapping patches, $\{\boldsymbol{x}_i\}$, and their DC components are compressed with DPCM method. The AC components are represented by sparse approximation with the learned dictionary, and the non-zero coefficients are further quantized and entropy coded to reduce the psychovisual redundancy and statistical redundancy, respectively. A given error threshold, ϵ, is usually utilized as stopping criterion for the following sparse coding problem,

$$\min \|\alpha_i\|_0, \quad s.t.\|\boldsymbol{x}_i - \mathbf{D}\alpha_i\|_2 \le \epsilon, i = 1, 2, ..., L. \tag{4}$$

Due to the sparsity of the coefficients, $\{\alpha_i\}$ are usually coded with the *run-level* method.

III. Rate-Distortion Based Sparse Coding For Image Set Compression

The proposed image set compression framework is illustrated in Fig.1. For a group of similar images, one representative image is first compressed with traditional image compression method, e.g. JPEG. Then, the decoded representative image is transformed into wavelet domain with 9/7 wavelet filter bank [9] and divided into overlapping image patches. These image patches after removing DC components are utilized to train dictionary with the RLS-DLA method. In order to improve the efficiency of the dictionary learning, the samples with very low energy are removed from training set. Finally, the dictionary atoms are reordered according to their frequency utilized in approximating the representative image.

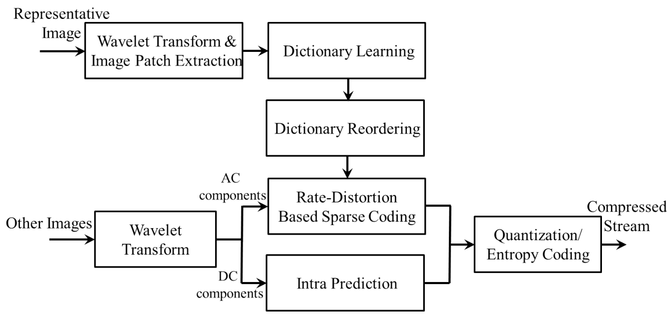


Fig. 1. The proposed image set compression framework based on sparse coding with ordered dictionary.

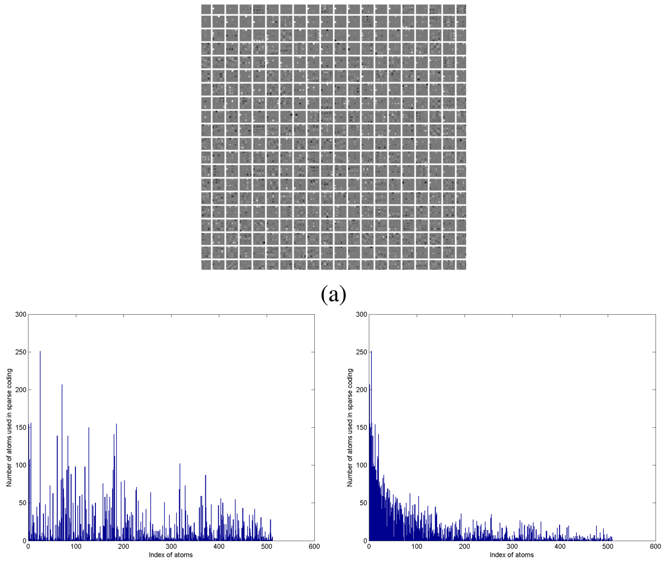
For other images in the image set, they are first transformed into wavelet domain and divided into non-overlapping patches. The DC components are compressed with intra prediction method, and the AC components are represented by sparse coding with the ordered dictionary based on rate-distortion costs. Finally, the prediction residuals in DC component and the non-zero coefficients of sparse coding for AC components are quantized and entropy coded with Huffman coding method. In the proposed method, the dictionary does not need to be transmitted, which can be learned at the decoder from the reconstructed representative image.

A. Dictionary Reordering

The traditional orthogonal transform, e.g., DCT, the atoms of which are usually arranged according the frequency variation. The non-zero coefficients of the transformed images with DCT usually distribute at the front of the frequency bands, which makes the positions of the coefficients be coded efficiently. However, this good property does not hold for the learned dictionary, the atoms of which are almost irregularly distributed, e.g., Fig.2 (a) illustrates the atoms in the learned dictionary and Fig.2 (b) illustrates the histogram of the atoms used in image representation. Although the atoms are irregular distributed, they have different frequencies utilized in sparse representation of images. Based on the assumption that images with similar contents usually share similar atoms in sparse coding, we reorder the atoms according to their frequencies used in approximating the representative image to improve the coding efficiency for the non-zero coefficients. When applying the ordered dictionary to other images, the non-zero coefficients are centralized to the front atoms of the dictionary and the variance of their *run* in run-level coding decreases significantly illustrated in Fig.2 (c), which implies coding gain.

B. Rate-Distortion Based Sparse Coding

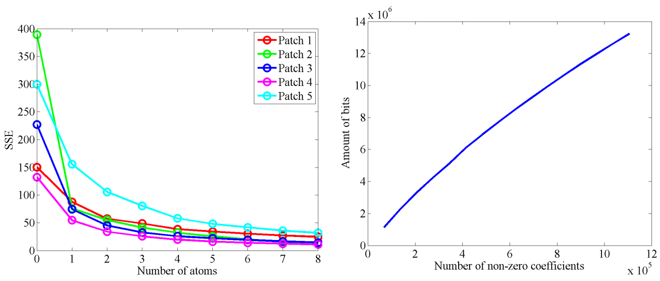
Classical sparse coding methods, e.g., OMP and LARS, they solve the optimization problem in Eqn. (4), by minimizing the number of non-zero coefficients at a given a distortion for every image patch independently. However, for different



(a) (b) (c)

Fig. 2. (a) The atoms in dictionary learned from wavelet domain. (b) The histogram of the atoms used in sparse coding for image, *Mail.Room*, with unordered dictionary, the variance of the *run* of non-zero coefficients, 2.1×10^4, (c) the histogram of the atoms used in sparse coding with ordered dictionary, the variance of the *run* of non-zero coefficients, 1.3×10^4.

image patches, they have different distortion reduction rate when increasing one atom for them, Fig. 3 (a) illustrating the relationship between distortion variation and number of atoms used in sparse coding for different image patches. Although the distortion of image patches decreases along with the number of atoms increasing, the distortion reduction rate is different for image patches. In addition, the more atoms used in sparse coding, the more bits are needed. Fig. 3 (b) shows the relationship between the L_0-norm of coefficient and the bits used to code these coefficients. The number of atoms used in sparse coding is proportional to the bits used in coding these image patches, which can be used as an coding rate estimation.



(a) (b)

Fig. 3. (a) The relationship between distortion reduction and the number of atoms used in sparse coding. (b) The relationship between the amounts of atoms used in sparse coding and the compression bits for the corresponding image patch.

Therefore, in order to improve the compression performance, we extend the OMP method by jointly decomposing a batch of image patches, or even all the image patches. Then, the independent sparse coding problem in Eqn. (4) is transformed into the following rate-distortion optimization problem,

$$\begin{aligned} &\min R, \quad s.t. D \leq \epsilon, \\ &where \quad R = \sum_{i=1}^{L} \|\alpha_i\|_0, \quad D = \sum_{i=1}^{L} \|\boldsymbol{x}_i - \mathbf{D}\alpha_i\|_2^2 \end{aligned} \tag{5}$$

The optimization problem in Eqn. (5) can be solved by the greedy algorithm. We apply OMP method to increase one atom for every image patch, and then find the patch with largest distortion reduction to increase one atom for it. The procedure can be performed iteratively until the distortion is not larger than the given threshold, ϵ. In order to reduce the complexity, we can process a batch of image patches at a time instead of all image patches. If each batch only includes one image patch, our proposed sparse coding method degenerates to the traditional OMP method. The proposed rate-distortion based sparse coding algorithm is formulated in **Algorithm 1**,

Algorithm 1 Rate-Distortion Based Orthogonal Matching Pursuit

Input:

Dictionary $\mathbf{D}$=$\{\boldsymbol{d}_i\}$;

Image patches: $\{\boldsymbol{x}_i\}$ (N is the number of batches, $\boldsymbol{x}_{ik}$ is the k^{th} batch);

Distortion threshold:ϵ;

Variables: red_dist_i, cur_dist_i and pre_dist_i, representing the distortion reduction, current distortion and previous distortion for image patches, $\{\boldsymbol{x}_i\}$.

for $i = 1 : N$ **do**

1. Initialize the distortion reduction variable red_dist_i and pre_dist_i for each patch as follows,
   (i) Compute $c_j =< \boldsymbol{x}_i, \boldsymbol{d}_j >$, and find the maximum denoted as c_{j_max} and the corresponding $\boldsymbol{d}_{j\_max}$;
   (ii) $red\_dist_i = \|\boldsymbol{x}_i - c_{j\_max} \times \boldsymbol{d}_{j\_max}\|_2$, $pre\_dist_i = \|\boldsymbol{x}_i\|_2$ and $cur\_dist_i = pred\_dist_i - red\_dist_i$;
2. Apply OMP algorithm to increase one atom for every patches in $\{\boldsymbol{x}_i\}$;
3. Find the current distortion for every patches, cur_dist_i;
4. Update the variables, red_dist_i = $pre\_dist_i -$ cur_dist_i;
5. Find the corresponding patch with maximum red_dist_i, the index of which is denoted as i_max;
6. Update the coefficient, and set $pre\_dist_i = cur\_dist_i$;
7. **if** $\sum pre\_dist_i > \epsilon$ **then** goto step 2;

end for

Output: The sparse coding coefficients: $\{\boldsymbol{\alpha}\}$

IV. Experimental Results

In order to verify the compression performance of the proposed method, we compare our method with the most popular image compression standards, i.e., JPEG[1] and JPEG 2000[2]

[1] JPEG codec, http://www.ijg.org

[2] JPEG2000 codec, http://www.openjpeg.org

using some different types of image sets, including buildings, man, mountains, and indoor scene. We also compared with learned dictionary based image coding method with traditional OMP [7] (denoted as *Dictionary*), the dictionaries of which are also learned from representative images but without dictionary reordering. The test image resolution varies from 800x600 to 3072x2048. The dictionary is 64×512 for each image set, and is learned from 8×8 image patches trasformed with wavelets. Fig. 4 shows the curve of the average PSNR and bitrate for all the images in every image set. Our proposed method achieves up to 6+ dB and 2+ dB compared with JPEG and JPEG 2000. The image number in every image set is 20, 10, 7, 5, 10 and 6 corresponding to Fig.(a)-Fig.(f). Fig. 5 illustrates the subjective comparison for an image compressed with different methods at 0.15 bpp. Our proposed method obtains more visual pleased image with very low bitrate.

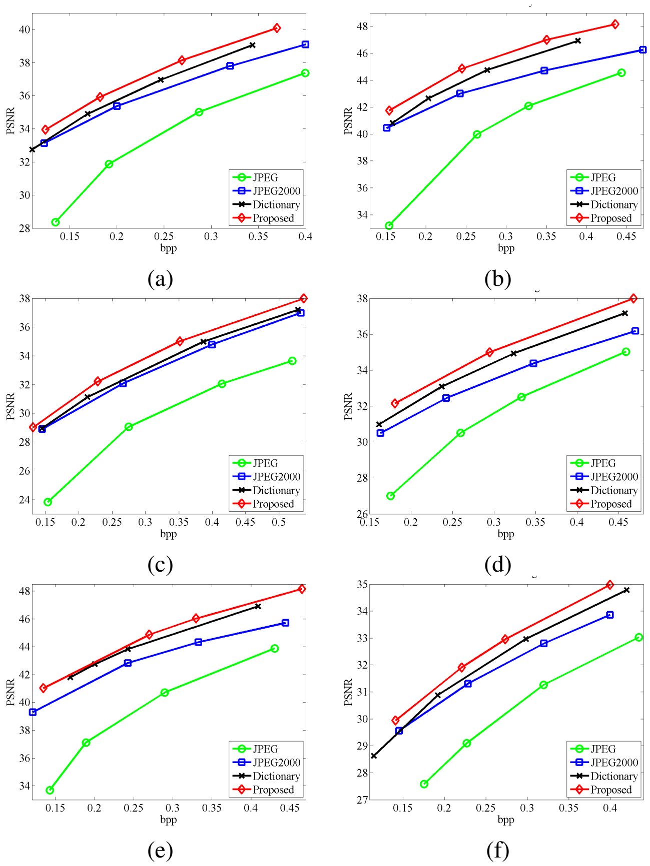


Fig. 4. Rate-distortion performance compared with JPEG and JPEG2000 in terms of PSNR for several sets of images. (a) *RockBoat.Summer.Palace*, (b) *Castle.Entry*, (c) *Mall.Room*,(d) *Wadham.College*,(e) *Man*, (f) *Mount.Huang*.

V. Conclusion

In this paper, we have proposed a new sparse coding based compression method for image set. The learned dictionaries are reordered based on representative images of image sets. And the images are approximated by jointly decomposing a batch of image patches on the dictionaries. Based on experimental results, the proposed method significantly improves the performance of image set compression.

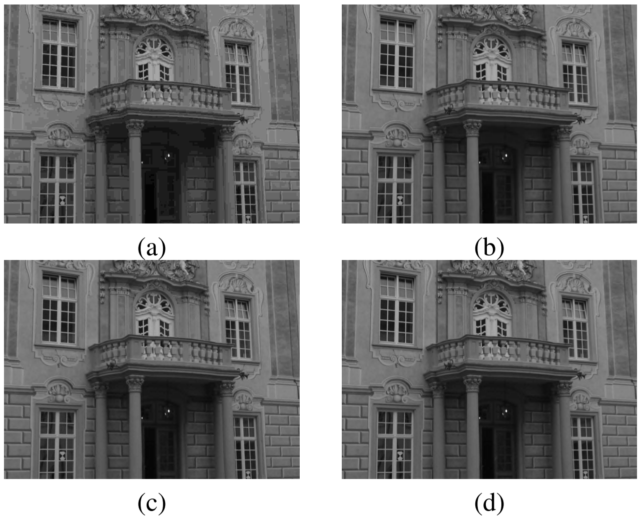


Fig. 5. Visual results for subjective quality comparison, a part of *Castle.Entry*, (a) JPEG at 0.16 bpp, 33.09 dB, (b) JPEG2000 at 0.15bpp, 39.93 dB, (c) *Dictionary* method, 0.15bpp 39.90dB, (d)Proposed method, 0.15 bpp, 40.87dB.

Acknowledgment

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Singapore, under its interactive Digital Media (IDM) Strategic Research Programme. This work was also supported in part by the National Basic Research Program of China (973 Program, 2015CB351800).

References

[1] X. Zhang, Y. Zhang, W. Lin, S. Ma, and W. Gao, "An inter-image redundancy measure for image set compression," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015, pp. 1274–1277.

[2] R. Neff and A. Zakhor, "Dictionary approximation for matching pursuit video coding," in *2000 International Conference on Image Processing, 2000. Proceedings*, vol. 2, Sep. 2000, pp. 828–831 vol.2.

[3] G. Monaci, P. Jost, P. Vandergheynst, and P. V, *Image Compression with Learnt Tree-Structured Dictionaries*, 2004.

[4] K. Skretting and K. Engan, "Image compression using learned dictionaries by RLS-DLA and compared with k-SVD," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 1517–1520.

[5] O. Bryt and M. Elad, "Compression of facial images using the k-SVD algorithm," *Journal of Visual Communication and Image Representation*, vol. 19, no. 4, pp. 270–282, May 2008.

[6] Y. Sun, X. Tao, Y. Li, and J. Lu, "Dictionary learning for image coding based on multisample sparse representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 11, pp. 2004–2010, Nov. 2014.

[7] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *1993 Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, 1993*, Nov. 1993, pp. 40–44 vol.1.

[8] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2121–2130, Apr. 2010.

[9] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, pp. 485–560, 1992.