Neurocomputing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Letters

Bases sorting: Generalizing the concept of frequency for over-complete dictionaries

Chun-Guang Li [a], Zhouchen Lin [b,*,1], Jun Guo [a]

[a] Pattern Recognition and Intelligent System Lab., School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, PR China
[b] Key Laboratory of Machine Perception (MOE), School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, PR China

ARTICLE INFO

Article history:
Received 5 September 2012
Received in revised form
25 December 2012
Accepted 2 February 2013
Communicated by Qingshan Liu

Keywords:
Sparse representation
Over-complete dictionary
Frequency

ABSTRACT

We propose an algorithm, called Bases Sorting, to sort the bases of over-complete dictionaries used in sparse representation according to the magnitudes of coefficients when representing the training samples. Then the bases are considered to be ordered from low to high frequencies, thus generalizing the traditional concept of frequency for over-complete dictionaries. Applications are also shown.

1. Introduction

Nowadays sparse representation has attracted a lot of attention from signal processing, image processing, pattern recognition, and machine learning communities, and its wide applications have been found, e.g., blind source separation [1], image restoration and repairing [2–4], super-resolution [5], face recognition [6], and subspace segmentation [7]. The reader is referred to [8] for a detailed survey on applications of sparse representation.

Different from Fourier or wavelet transform, which depends upon pre-constructed or pre-defined bases, sparse representation uses data-adaptive over-complete bases, also called over-complete *dictionaries*. Besides being over-complete, current dictionaries are usually unstructured. In contrast, in Fourier or wavelet transform, the bases are well ordered according to their frequencies, which have an intuitive physical interpretation: low and high frequency bases correspond to slow-varying and fast-varying waveforms, respectively (Fig. 1). Note that the concept of frequency plays a critical role in the traditional signal processing theory. Unfortunately, for over-complete dictionaries such a concept is lacking. This incurs a lot of disadvantages. For example, one has to display an over-complete dictionary in a random order, and when performing data compression it is unclear which bases should be kept and which should be discarded. In this paper, we aim at generalizing the concept of "frequency" for over-complete dictionaries. Once the generalized frequency is defined, many concepts and techniques in the traditional signal processing may be transplanted to sparse representation.

For general over-complete dictionaries, we can no longer define their frequencies by looking at whether a basis is fast or slowly varying, because the bases may not be in waveforms. For example, for a given basis of 128-dimensional SIFT features [9] (see Fig. 2(b)) it is hard to say whether its "frequency" is low or high. This is nontrivial even for a dictionary learnt from raw image patches (see Fig. 2(a)). As a result, in the literature [2,5,10–14], people simply show and store their over-complete dictionaries in a random order. In this paper, we propose to define "frequency" for an over-complete dictionary by sorting the bases appropriately, so that the generalized frequency can indeed be a generalization of frequency in the traditional sense, when the dictionary reduces to the traditional bases. The algorithm is called Bases Sorting (BS). After sorting, the leading bases can be regarded as of "low frequency" while the ending bases regarded as of "high frequency." The sorting criterion is inspired by the $1/f$-power law [15] of the Fourier spectra of natural images. As the over-complete dictionary is data-adaptive, so should the orders of the bases be. Given training data, for each sample BS first computes the sparest representation coefficients with respect to the dictionary, then sorts the bases according to the magnitudes of the coefficients. The final order of a basis is its average position over the training data.

We validate our criterion and BS algorithm on a two-dimensional Discrete Cosine Transform (2D-DCT) dictionary. The experimental results confirm that the natural order of the 2D-DCT dictionary can be well recovered by BS. Moreover, the $1/f$-power

* Corresponding author. Tel.: +86 10 62753313.
E-mail address: zlin@pku.edu.cn (Z. Lin).
[1] This work was started when the first author was visiting Microsoft Research Asia.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.02.005

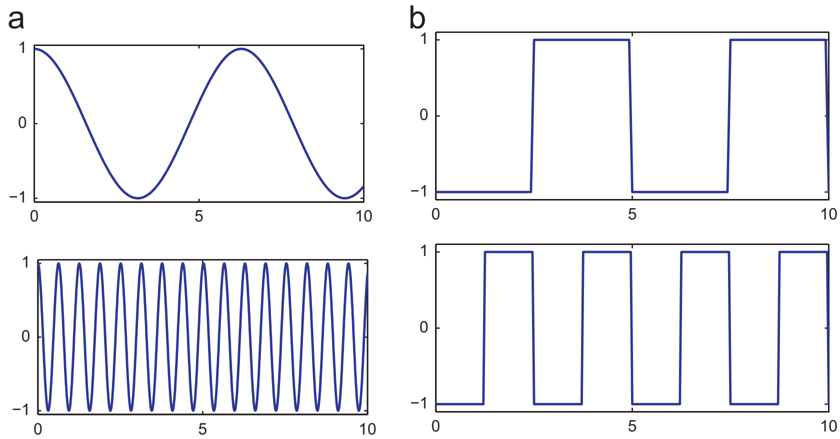

Fig. 1. The low frequency bases and high frequency bases. (a) Cosine functions. (b) Haar wavelets.

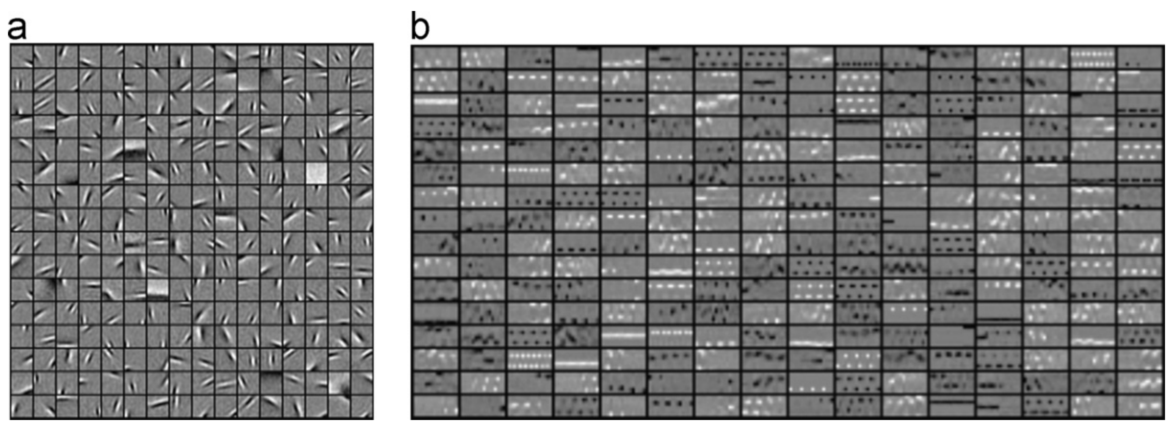

Fig. 2. Unsorted dictionaries. (a) Dictionary learnt from 14×14 raw image patches. (b) Dictionary learnt from 128-dimensional SIFT features. (**The images in this paper are best viewed on screen!**).

law can also be preserved for general over-complete dictionaries, e.g., those learnt from raw image patches and SIFT features. These suggest that the traditional concept of frequency is successfully generalized by using our BS algorithm to sort the bases. We further apply the generalized frequency to dictionary visualization and data compression to show the usefulness of generalized frequency.

2. Sparse representation

Denote the data set as $X=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n]$, where $\mathbf{x}_i\in\mathbb{R}^p$, $i=1,\ldots,n$, and the dictionary $U=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_m]$, where $\mathbf{u}_j\in\mathbb{R}^p$, $j=1,\ldots,m$. Each column $\mathbf{u}_j$ of U is called a basis or an atom. The dictionary is often over-complete, i.e., $p\ll m$, and is learnt from training data, e.g., by K-SVD [10] or the Lagrange dual method [11].

The core of the sparse representation theory is to represent a data vector as the sparest linear combination of the bases in an over-complete dictionary. A basic formulation is as follows:

$$\min_{\mathbf{v}} \ \|\mathbf{v}\|_0 \quad \text{s.t.} \ \ \mathbf{x}=U\mathbf{v}, \tag{1}$$

where $\|\mathbf{v}\|_0$ denotes the l_0-norm[1] of $\mathbf{v}$, which is the number of non-zero entries in $\mathbf{v}$.

Problem (1) is NP-hard. There has been a lot of effort on solving (1). Donoho [16] proved that the following convex program:

$$\min_{\mathbf{v}} \ \|\mathbf{v}\|_1 \quad \text{s.t.} \ \ \mathbf{x}=U\mathbf{v} \tag{2}$$

is equivalent to (1) under rather general conditions, where $\|\mathbf{v}\|_1$ denotes the l_1-norm of $\mathbf{v}$, which is the sum of absolute values of the entries in $\mathbf{v}$, and there have been many algorithms to solve (2) efficiently. See [17] for a review. To tolerate a certain level of noise, problem (2) can be relaxed as

$$\min_{\mathbf{v}} \ \|\mathbf{v}\|_1 \quad \text{s.t.} \ \ \|\mathbf{x}-U\mathbf{v}\|_2^2<\epsilon, \tag{3}$$

where $\epsilon>0$ is an estimated noise level.

Unlike Fourier and wavelet transforms in which the order of bases are analytically predefined as frequency, the bases $\mathbf{u}_j$'s in the dictionary U are usually orderless because there is no counterpart definition of "frequency" for over-complete dictionaries. In the next section, we will introduce our criterion and algorithm to sort the bases in an over-complete dictionary. Then the bases are considered to be arranged from low to high frequencies. In this way, we generalize the traditional concept of frequency.

3. Criterion and algorithm to sort an over-complete dictionary

It has been observed that statistics of many man-made or natural objects, such as city size, incomes, word frequencies, earthquake

[1] Strictly speaking, l_0-norm is not a norm but a pseudo-norm.

Please cite this article as: C.-G. Li, et al., Bases sorting: Generalizing the concept of frequency for over-complete dictionaries, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2013.02.005

magnitudes, and the vertex degrees in the World Wide Web, obey a power law distribution [18–20], i.e., small magnitudes are very common, whereas large magnitudes are very rare. In the study of natural image statistics, Field [15] discovered a 1/*f*-power law, i.e., by performing Fourier transform on natural images the magnitudes of the Fourier coefficients at frequency *f* (averaged over orientations) are proportional to 1/*f*. For wavelet transforms, the magnitudes of wavelet coefficients also obey this law.

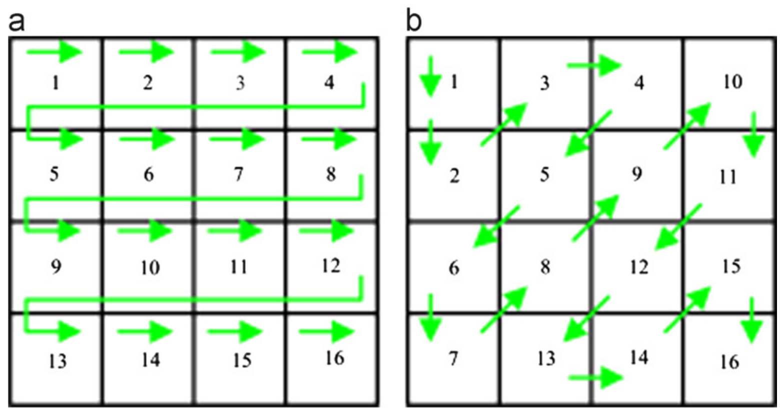

Fig. 3. Two schemes to arrange the sorted bases. (a) The row-by-row arrangement. (b) The zigzag arrangement.

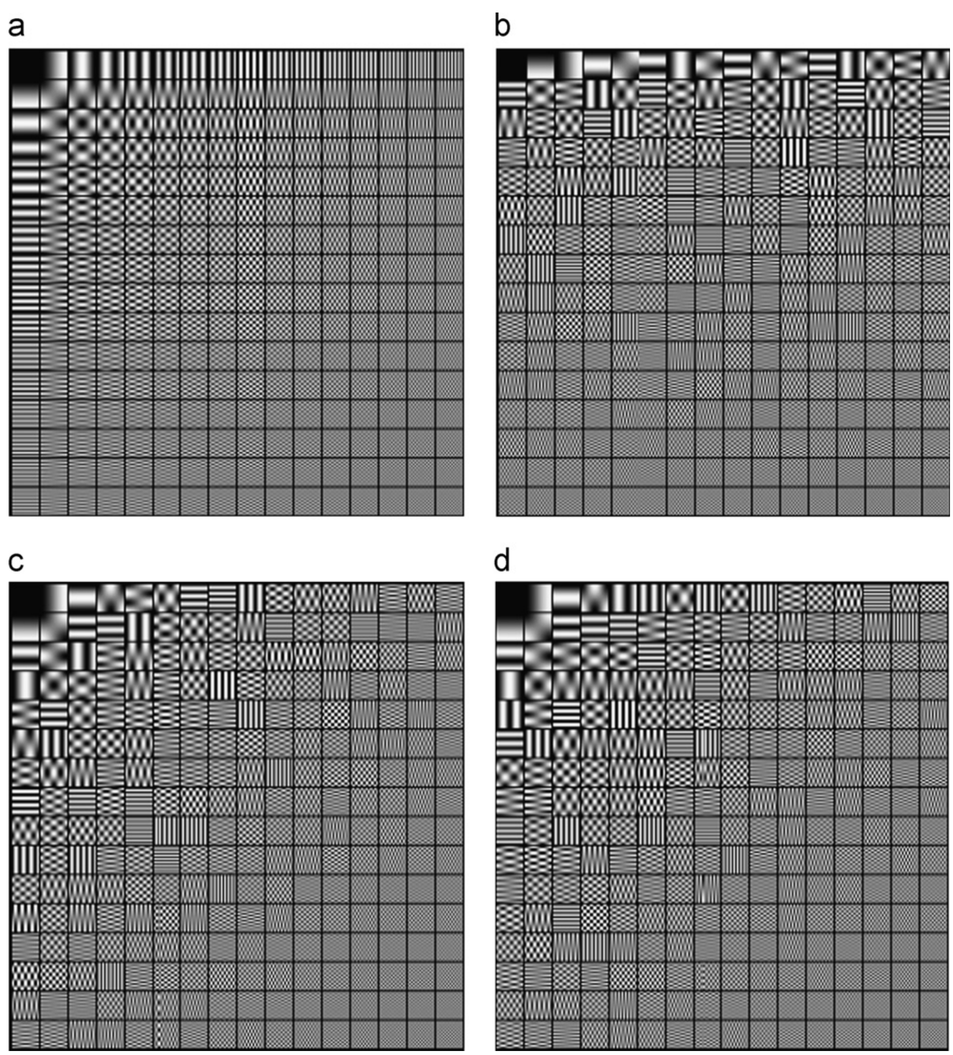

Fig. 4. Validation experiments on 256 standard 2D-DCT bases. (a) The standard 2D-DCT bases in the natural order. (b) and (c) are the results of BS shown in two arrangements. (d) The result of baseline shown in zigzag arrangement.

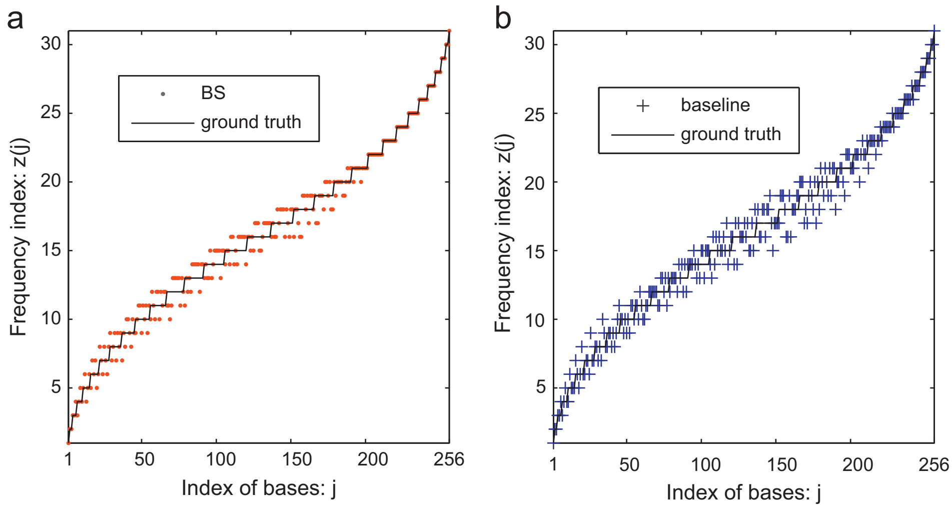

Fig. 5. Frequency indices of the 256 2D-DCT bases sorted by (a) BS and (b) baseline, respectively.

Inspired by the above observations, we propose to sort the bases in an over-complete dictionary so that the magnitudes of the sparsest representation coefficients of data are also in a descending order. In this way, the indices of bases can be naturally *interpreted* as frequencies.

According to this principle, a straightforward way, called the baseline method, is to first compute the average magnitudes of the representation coefficients of the training samples and then sort the bases in the descending order of the averaged magnitudes. However, we will show that such a method is *not* the best approach. A much better way is to first sort the bases according to the magnitudes of the representation coefficients for each training sample, then record the position of each basis, and finally resort the bases according to their averaged positions. We call our method the Bases Sorting (BS) algorithm, as detailed below.

Given an over-complete dictionary U which is learnt from data set X, for $\mathbf{v} \in \mathbb{R}^m$, $\text{supp}(\mathbf{v}) = \{j \in I | v_j \neq 0\}$ denotes the support of $\mathbf{v}$, where $I = \{1,2,\ldots,m\}$. For the sparsest representation $\mathbf{v}$ of $\mathbf{x} \in X$ w.r.t. U, i.e., $\mathbf{v}$ is the solution to (1), we define the non-zero components $\{v_j | j \in \text{supp}(\mathbf{v})\}$ as the *active coefficients* of $\mathbf{x}$ and the subset of bases $U^{(\mathbf{x})} = \{\mathbf{u}_j | j \in \text{supp}(\mathbf{v})\}$ as ***x**-activated bases*. Then we can sort $U^{(\mathbf{x})}$ in an descending order of the absolute values of their coefficients $\{|v_j|\}$. The indices of the $\mathbf{x}$-activated bases are recorded as $\pi(U^{(\mathbf{x})})$. Then the final order of bases are the expectation of the orders, i.e.,

$$\pi(U) = \mathbb{E}[\pi(U^{(\mathbf{x})})], \tag{4}$$

where $\mathbb{E}[\cdot]$ is the expectation operator.

Formally our BS algorithm consists of four steps:

1. For each data vector $\mathbf{x}_i \in X$, calculate its sparsest representation vector $\mathbf{v}_i$ with respect to the dictionary U.
2. Determine the order $\pi(U^{(\mathbf{x}_i)})$ of bases in $U^{(\mathbf{x}_i)}$ by sorting the magnitudes of the corresponding active coefficients $\{|v_j| | j \in \text{supp}(\mathbf{v}_i)\}$ of $\mathbf{x}_i$ in descending order.
3. Record all the order $\pi(U^{(\mathbf{x}_i)})$ to obtain an averaged frequency order $\pi(U)$.
4. Sort the bases in U in an ascending order of $\pi(U)$.

The pseudo-code of BS algorithm is presented in Algorithm 1. Note that as calculating the sparsest representation of training samples is a part of dictionary learning, sorting the bases is simply a by-product of dictionary learning and actually does not add to the computation load.

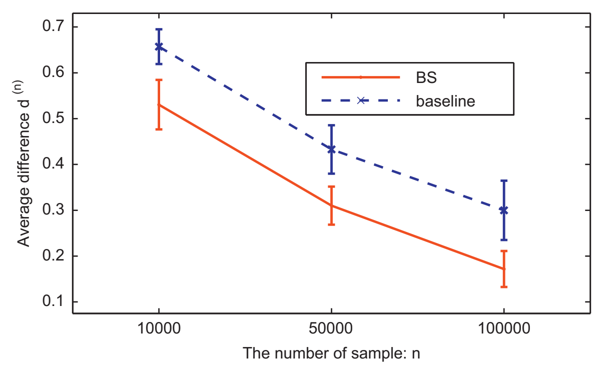

Fig. 6. Stability comparison of the estimated generalized frequency order.

Algorithm 1. Bases Sorting (BS) algorithm.

1: **Input**: training data matrix $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, dictionary $U = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$, and parameter $\epsilon > 0$.
2: Initialize the accumulation buffer: $\mathbf{s} = (0, \ldots, 0)$.
3: **for** $i=1$ to n **do**
4: Calculate the sparsest representation of $\mathbf{x}_i$ by solving

$$\mathbf{v}_i = \arg\min_{\mathbf{v}} \|\mathbf{v}\|_1, \quad s.t.\ \|\mathbf{x}_i - U\mathbf{v}\|_2^2 < \epsilon. \tag{5}$$

5: Identify the support of sparse representation vector $\mathbf{v}_i$ and the set of active bases:

$$I^{(i)} = \{j | v_{ij} \neq 0, j = 1, \ldots, m\},$$
$$U^{(\mathbf{x}_i)} = \{\mathbf{u}_j | j \in I^{(i)}\}. \tag{6}$$

6: Compute the magnitude vector $\mathbf{a}_i$ of active coefficients, i.e., $a_{ij} = |v_{ij}|$.
7: Sort the activated bases in $U^{(\mathbf{x}_i)}$ according to the descending order of the magnitude vector $\mathbf{a}_i$:

$$\pi(U^{(\mathbf{x}_i)}) = \text{sort}(I^{(i)}, \mathbf{a}_i). \tag{7}$$

8: Update the accumulation buffer: $s_j = s_j + 1$ for all $j \in I^{(i)}$.
9: **end for**
10: Compute the averaged order by

$$\pi(U) = \sum_{i=1}^{n} \pi(U^{(\mathbf{x}_i)})./\mathbf{s}, \tag{8}$$

where ./ is element-wise division.
11: Sort U in an ascending order of $\pi(U)$.

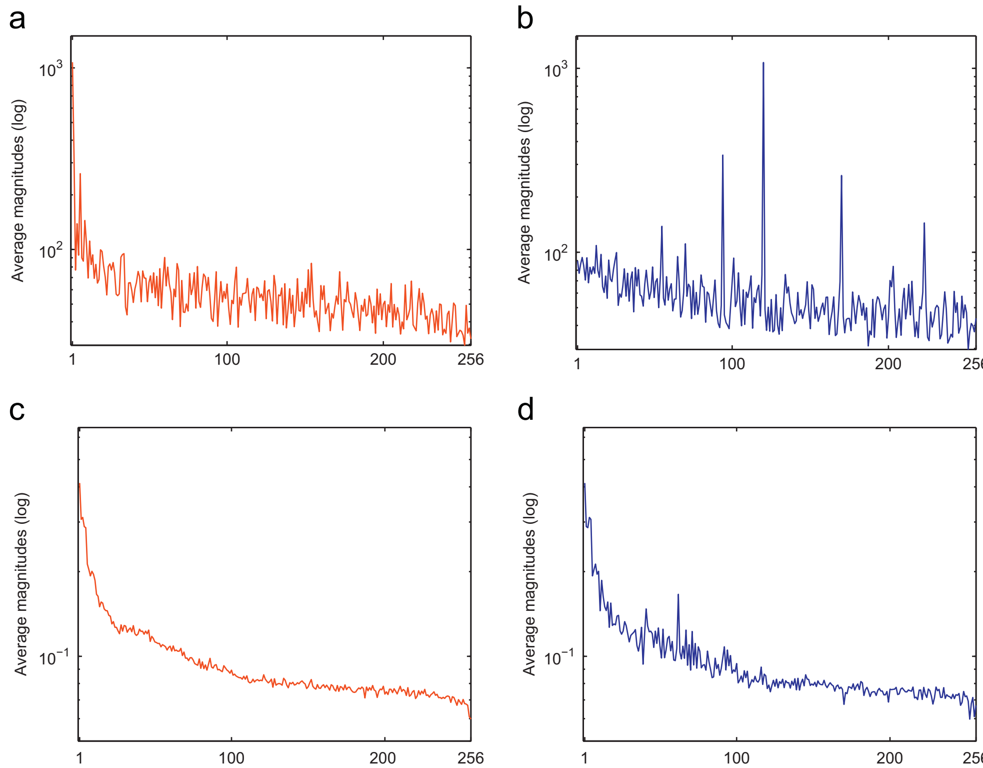

Fig. 7. The average spectra of the test data in log scale. The top and bottom rows are the results of using the sorted dictionary consisting of 256 bases learnt from raw image patches and SIFT features, respectively. The left and right columns are the average spectra defined by BS and baseline, respectively.

With the sorted dictionary U, we can say that $\mathbf{u}_j$ is *of frequency* j and the bases in U are sorted from low to high frequencies. The sparsest representation $\mathbf{v}$ of a data vector $\mathbf{x}$, arranged according to the corresponding bases, is called the *spectrum* of $\mathbf{x}$. If the active bases of a data vector $\mathbf{x}$ are all low, middle, or high frequency bases, we can say that $\mathbf{x}$ is *a low, middle, or high frequency signal*, respectively. Thus the traditional concepts related to frequency are naturally generalized.

4. Experiments

We first validate our criterion and BS algorithm by testing with 2D-DCT bases to show that our generalized concept of frequency is consistent with the traditional sense when the bases are waveform. Moreover, the $1/f$-power law can be preserved by our generalized frequency. We further apply the generalized frequency to dictionary visualization and data compression to show its usefulness in practice.

In our experiments, we use the Lagrange dual method [11] to learn the dictionary and adopt the Feature-Sign algorithm [11] to find the sparsest representation in Algorithm 1. Throughout the experiments, we will compare our BS algorithm with the baseline method to show the advantages of BS algorithm.

To visualize sorted dictionaries when the bases are 2D signals, we arrange their bases in two schemes, row-by-row and zigzag, as illustrated in Fig. 3.

4.1. Validations

A natural question about our generalized frequency is whether it is consistent with the classical concept of frequency. To answer this question, we first apply our BS algorithm to sort the standard 2D-DCT bases. The training data X consist of 200,000 image patches of 16×16 pixels, which are randomly sampled from the images of Scene-15.[2] The dictionary U simply consists of the 256 standard 2D-DCT bases (see Fig. 4(a)), rearranged into 256-dimensional vectors.

The 256 2D-DCT bases sorted by the BS algorithm are displayed in Fig. 4(b) and (c), in row-by-row and zigzag schemes, respectively. As can be seen, the 2D-DCT bases are organized in a good order: the visually low frequency bases are in the front and the visually high frequency bases are at the end.

As the BS results on the 2D-DCT bases are not identical to the natural order, it is not easy to tell visually how much they differ. To compute the difference between them quantitatively, we recall that the 2D-DCT bases on the same anti-diagonals in Fig. 4(a) are considered of the same frequency. So we define the following zigzag step function $z(j)$ that maps the j-th sorted basis to the $z(j)$-th anti-diagonal (or called the $z(j)$-th frequency index):

$$z(j)=\begin{cases} 1, & j=1,\\ 2, & j=2,3,\\ 3, & j=4,5,6,\\ \cdots & \\ 29, & j=251,252,253,\\ 30, & j=254,255,\\ 31, & j=256. \end{cases} \tag{9}$$

The zigzag step functions of both the natural order and generalized frequency order are shown in Fig. 5(a). We can see that they agree with each other fairly well. The average difference between their zigzag step functions is only 0.4430. For comparison, we present the results of baseline in Figs. 4(d) and 5(b). It can be seen

[2] Scene-15 is a benchmark data set for scene classification, available at http://www-cvr.ai.uiuc.edu/ponce_grp/data/.

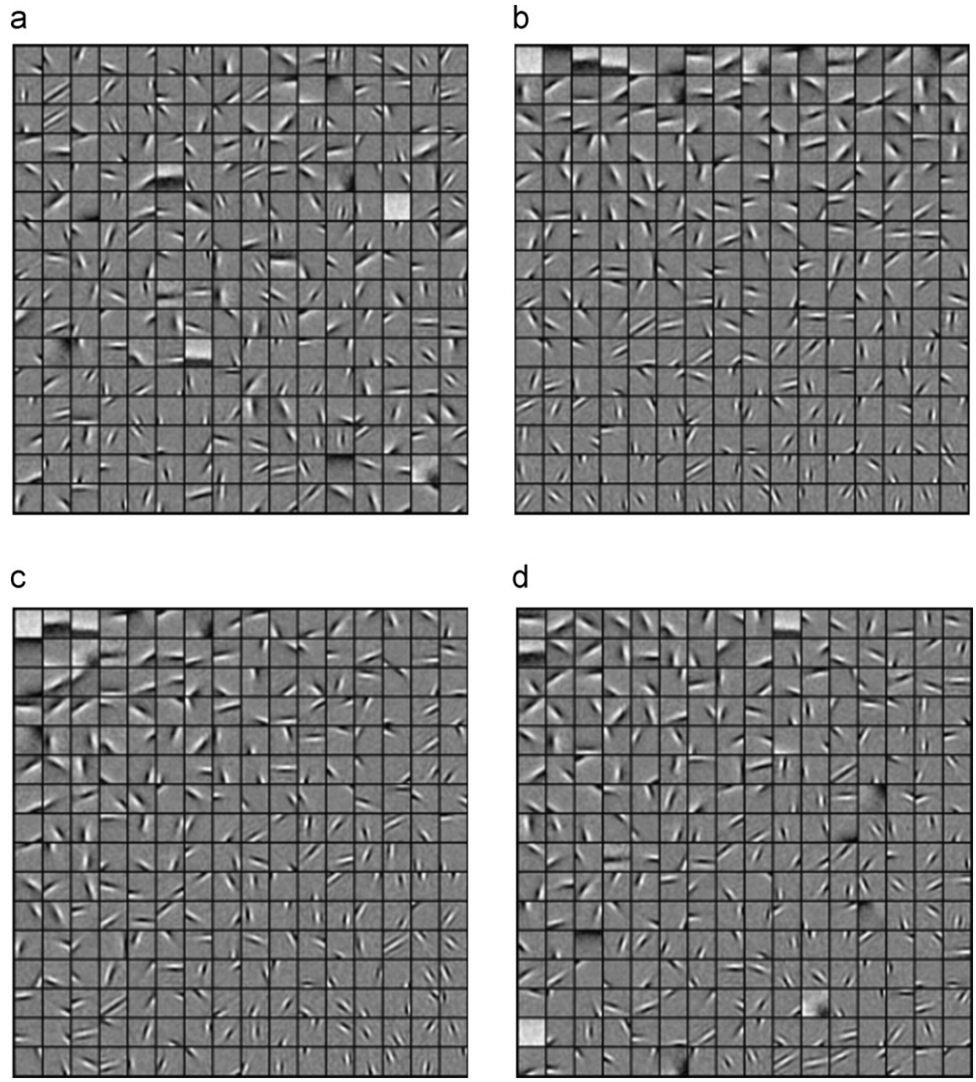

Fig. 8. Visualization of an over-complete dictionary of size 256, learnt from 50,000 image patches. (a) The orderless display. (b) and (c) are the sorted results by BS in row-by-row and zigzag arrangements, respectively. (d) The result of the baseline in zigzag arrangement.

that baseline is worse than BS, especially when the index is between 100 and 250. Actually, the average difference between the zigzag step functions of baseline and natural order is 0.6172.

The generalized frequency defined by BS is actually also more robust than that by baseline. To show this, we compute the zigzag step functions of bases sorted using different numbers of training samples and then calculate their difference as follows:

$$d^{(n)} = \frac{1}{m} \| \mathbf{z}^* - \mathbf{z}^{(n)} \|_1, \tag{10}$$

where $\mathbf{z}^*$ is the zigzag step function of the generalized frequency order estimated by using 200,000 training samples, $\mathbf{z}^{(n)}$ is the zigzag step function estimated by using n training samples, and m is the number of bases in U. We randomly select $n = 10^4$, 5×10^4, 10^5 training samples to estimate the generalized frequency order and calculate the distances $d^{(n)}$. The means and standard variances of distance $d^{(n)}$ over 20 trials are displayed in Fig. 6. We can see that the generalized frequency order by BS is much more stable than that by baseline when the number of training samples changes. This is because BS uses the expectation of order to define its generalized frequency order.

To test whether the 1/f-power law can be preserved by our generalized frequency, we randomly sample one million raw image patches from the images of Scene-15 as test data and shown in Fig. 7(a) the average magnitudes of the coefficients of the bases when representing the *test* data. We can see that the average magnitudes fall off quickly when the generalized frequency increases. The fall-off curve is very close to Laplacian. This suggests that the 1/f-power law is indeed preserved by our generalized frequency. For the baseline method, there are several high peaks in the "mid-and-high frequencies," as shown in Fig. 7(b). Similar phenomenon can be observed on the dictionary learnt from 50,000 128-D SIFT features (Fig. 7(c) and (d)).

These experiments suggest that our generalized frequency is indeed a good generalization of the traditional frequency.

4.2. Dictionary visualization

A straightforward application of generalized frequency is to display the bases in a good order.

First, we randomly sample 50,000 image patches of 14×14 pixels from 10 natural scene images[3] and learn an over-complete dictionary with 256 bases (Fig. 8(a)). The dictionary sorted by our BS algorithm is displayed in Fig. 8(b) and (c) in two arrangement

[3] The 10 images are available at http://ai.stanford.edu/~hllee/softwares/nips06-sparsecoding.htm.

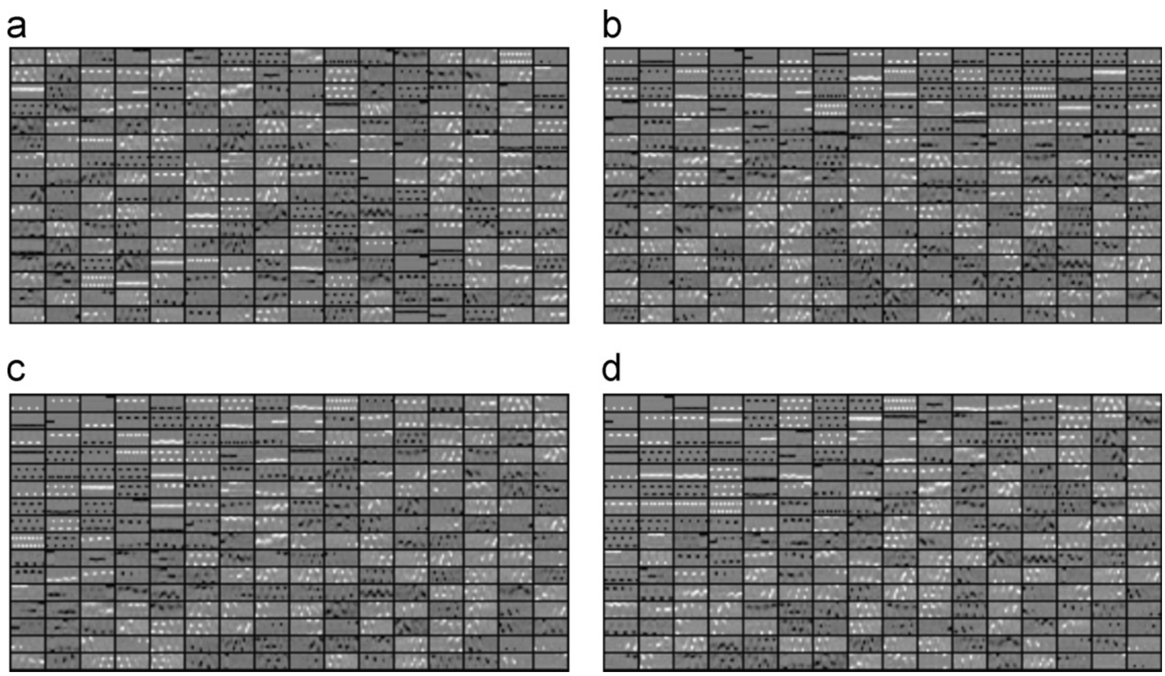

Fig. 9. Visualization of a dictionary of size 256, learnt from 50,000 128-dimensional SIFT features. (a) The orderless display. (b) and (c) are the sorted results by BS in row-by-row and zigzag arrangement schemes, respectively. (d) The result of the baseline shown in zigzag arrangement.

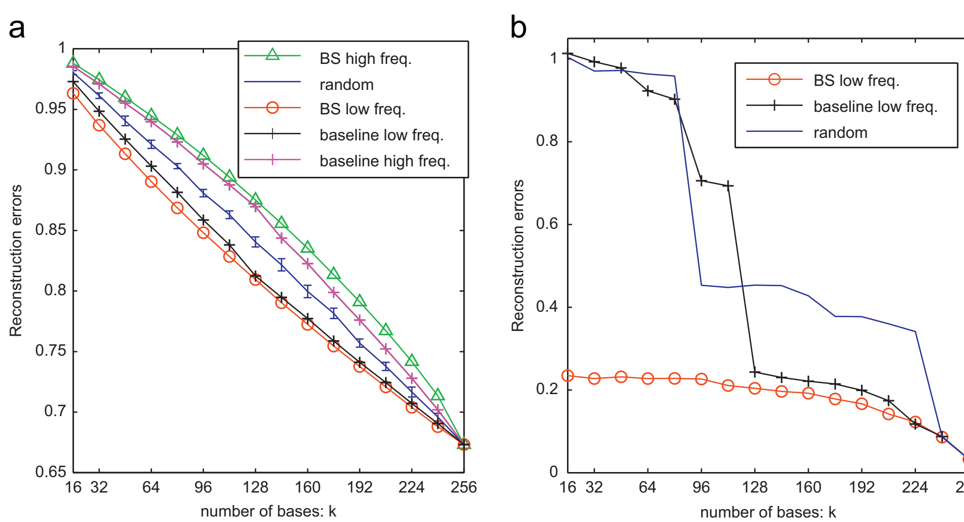

Fig. 10. Average reconstruction errors of different choices of k bases. (a) Average reconstruction errors on training samples. That the reconstruction error is not zero when all the bases are used is because we have used a noise tolerance parameter in dictionary learning (see (5)). (b) Average reconstruction errors of reconstructing test images by truncated k bases.

schemes. We can see that visually low frequency bases are sorted before visually high frequency ones. The bases sorted by the baseline method are presented in Fig. 8(d). As can be seen, the bases are not sorted properly. For example, the lowest frequency basis (i.e., the brightest block) is not at the leading position. So our BS algorithm can yield better sorting result than the baseline.

Second, we learn an over-complete dictionary on 50,000 128-dimensional SIFT features [9] (Fig. 9(a)) randomly sampled from images in the Scene-15 data set and sort it by BS. The learnt bases look like Morse code. The sorted dictionary is displayed in Fig. 9(b) and (c) in two arrangement schemes. Compared with the orderless display in Fig. 9(a), the sorted bases are in an interesting order, which seems to change from simple pattern (e.g., regularly spaced dots or dashes) to complex patterns (e.g., irregularly spaced dots). The bases sorted by the baseline method are presented in Fig. 9(d). It is difficult to tell the difference between the results of BS and baseline intuitively because SIFT feature is not visually perceivable.

These experiments show that BS can find the intrinsic order among the bases of an over-complete dictionary.

4.3. Data compression

DCT and Discrete Wavelet Transform (DWT) have been widely used in data compression thanks to an important property of DCT and DWT. Namely, the high frequency coefficients can be truncated and the data can still be well approximated by the low frequency ones. In this section we show that over-complete dictionaries also have such a property in the sense of generalized frequency.

Please cite this article as: C.-G. Li, et al., Bases sorting: Generalizing the concept of frequency for over-complete dictionaries, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2013.02.005

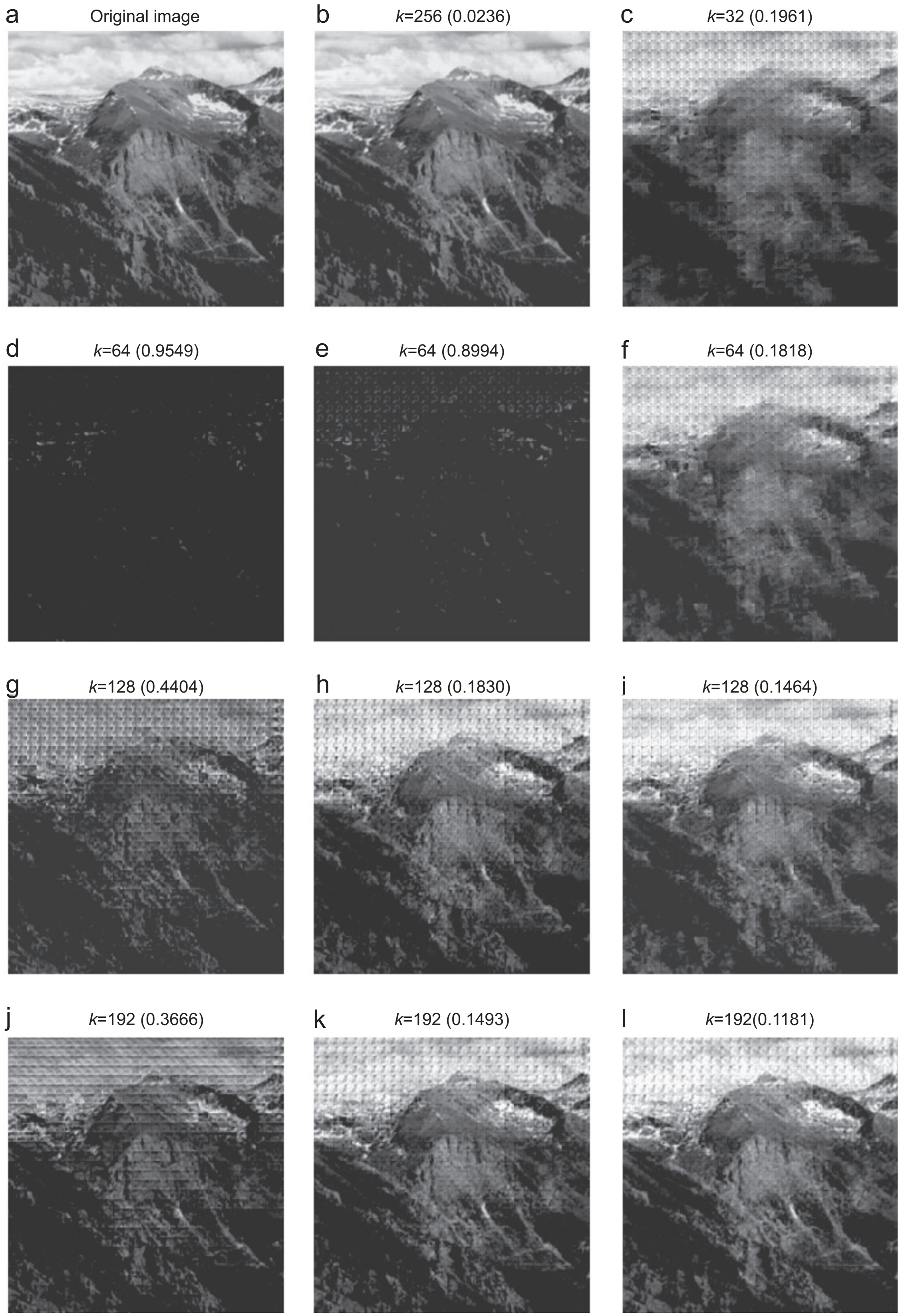

Fig. 11. Reconstructed images by using truncated k bases. The values in brackets are the reconstruction errors. (a) The original image. (b) Reconstructed by using all 256 bases. (d), (g), and (j) are reconstructed by using k randomly chosen bases. (e), (h), and (k) are reconstructed by using the k lowest frequency bases defined by the baseline method. (c), (f), (i), and (l) are reconstructed by the k lowest frequency bases defined by BS.

We prepare an over-complete dictionary learnt from raw image patches as in Section 4.2. Then we compute the average reconstruction errors using a part of the bases. The bases are chosen in three ways:

1. Randomly choose k bases from the dictionary.
2. Choose the k lowest frequency bases of the dictionary.
3. Choose the k highest frequency bases of the dictionary.

The average reconstruction errors on the training samples are presented in Fig. 10(a). One can observe that using the k lowest frequency bases always leads to the best reconstruction accuracy, while using the k highest frequency bases results in the worst accuracy, and using random k bases is between these two choices. Moreover, the reconstruction errors using k lowest frequency bases defined by BS are consistently lower than those by baseline. Although the difference between BS and baseline on the training samples is not salient, their difference is drastic on the testing samples (Fig. 10(b)), especially when k is small.

For visual comparison, we reconstruct an image with the dictionary in Fig. 8 and display in Fig. 11 the images reconstructed by using k lowest frequency bases. Clearly, using the k lowest frequency bases defined by BS yields much better visual quality than the baseline, which is only slightly better than using k randomly chosen bases. This is because BS can result in better average spectra that facilitate data compression. Namely, high frequencies have lower representation magnitudes, as having been shown in Fig. 7.

5. Conclusions

We have proposed a novel criterion and an algorithm to sort the bases in an over-complete dictionary, and accordingly defined the generalized frequency as the indices of the sorted bases. We have validated that our generalized frequency can be a good generalization of the traditional concept of frequency used in Fourier and wavelet transforms. We have also applied generalized frequency to dictionary visualization and data compression. We believe that based on our generalized frequency more concepts in the traditional signal processing theory can be transplanted to sparse representation. This will be our future work.

Acknowledgment

C.-G. Li is supported by the Natural Science Foundation of China (NSFC) under Grant nos. 61005004 and 61175011, the 111 Project under Grant no. B08004, and the Fundamental Research Funds for the Central Universities under Grant no. 2012RC0108. Z. Lin is supported by the NSFC under Grant nos. 61272341, 61231002, and 61121002. J. Guo is supported by the NSFC under Grant no. 61273217.

References

[1] Y. Li, A. Cichocki, S. Amari, Analysis of sparse representation and blind source separation, Neural Comput. 16 (6) (2004) 1193–1234.
[2] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Trans. Image Process. 15 (12) (2006) 3736–3745.
[3] J. Starck, M. Elad, D. Donoho, Image decomposition via the combination of sparse representation and a variational approach, IEEE Trans. Image Process. 14 (10) (2005) 1570–1582.
[4] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration, IEEE Trans. Image Process. 17 (1) (2008) 53–69.
[5] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[6] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[7] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: CVPR, 2009, pp. 2790–2797.
[8] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, Proc. IEEE 98 (6) (2010) 1031–1044.
[9] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.
[10] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[11] H. Lee, A. Battle, R. Raina, A.Y. Ng, Efficient sparse coding algorithms, in: NIPS, 2007, pp. 801–808.
[12] R. Rubinstein, A. Bruckstein, M. Elad, Dictionaries for sparse representation modeling, Proc. IEEE 98 (6) (2010) 1045–1057.
[13] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, J. Mach. Learn. Res. 11 (2010) 19–60.
[14] J. Mairal, F. Bach, J. Ponce, Task-driven dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 791–804.
[15] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am. A 4 (1987) 2379–2394.
[16] D. Donoho, For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution, Comm. Pure Appl. Math. 59 (6) (2006) 797–829.
[17] A. Yang, A. Ganesh, S. Sastry, Y. Ma, Fast l_1-minimization algorithms and an application in robust face recognition: a review, Technical Report UCB/EECS-2010-13, EECS Department, University of California, Berkeley, February 2010.
[18] M. Levy, S. Solomon, New evidence for the power-law distribution of wealth, Physica A: Stat. Mech. Appl. 242 (1) (1997) 90–94.
[19] A. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[20] L. Adamic, B. Huberman, Power-law distribution of the world wide web, Science 287 (5461) (2000) 2115.

Chun-Guang Li received his B.E. degree in telecommunication engineering from Jilin University in 2002 and Ph.D. degree in signal and information processing from Beijing University of Posts and Telecommunications (BUPT) in 2007. Currently he is a lecturer with the School of Information and Communication Engineering, BUPT, and as a member of Pattern Recognition and Intelligent System (PRIS) lab. He visited the Visual Computing Group, Microsoft Research Asia, from July 2011 to April 2012. His research interests focus on statistical machine learning and high dimensional data analysis, including manifold learning, sparse and low-rank model, and visual object categorization. He is a member of the IEEE, ACM, and CCF.

Zhouchen Lin received the Ph.D. degree in applied mathematics from Peking University in 2000. Currently, he is a professor at the Key Laboratory of Machine Perception (MOE), School of Electronics Engineering and Computer Science, Peking University. He is also a chair professor at Northeast Normal University. Before March 2012, he was a lead researcher in the Visual Computing Group, Microsoft Research Asia. He was a guest professor at Shanghai Jiaotong University, Beijing Jiao Tong University, and Southeast University. He was also a guest researcher at the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include computer vision, image processing, computer graphics, machine learning, pattern recognition, and numerical computation and optimization. He is an associate editor of Int. J. Computer Vision and Neurocomputing and a senior member of the IEEE.

Jun Guo received B.E. and M.E. degrees from Beijing University of Posts and Telecommunications (BUPT), China, in 1982 and 1985, respectively, Ph.D. degree from the Tohuku-Gakuin University, Japan, in 1993. At present he is a professor and the dean of School of Information and Communication Engineering, BUPT. His research interests include pattern recognition theory and application, information retrieval, content based information security, and network management. He has published over 200 papers, some of them are on world-wide famous journals or conferences including SCIENCE, Nature's online Journal of Scientific Reports, IEEE Trans. on PAMI, IEICE, ICPR, ICCV, SIGIR, etc. His book "Network management" was awarded by the government of Beijing city as a recommended textbook for higher education in 2004.

Please cite this article as: C.-G. Li, et al., Bases sorting: Generalizing the concept of frequency for over-complete dictionaries, Neurocomputing (2013), http://dx.doi.org/10.1016/j.neucom.2013.02.005