# A Novel Image Denoising Method Based on Sparse Representation and Incremental Dictionary Learning for Large-scale Dataset

ABSTRACT. *Denoising of images is one of the most basic tasks of image processing. The approach which we concentrate on is based on sparse and redundant representations over a learned overcomplete dictionary. It has been proved that the K-singular value decomposition (K-SVD) algorithm is a highly effective method of training the overcomplete dictionaries for sparse signal representation. However, when the training datasets become extremely large in scale, this algorithm will be no longer effective as it is batch algorithm which deals with all the training samples at each iteration. In this paper we present an efficient incremental learning alternative implementation of this algorithm, which both accelerates it and adapts gracefully to large datasets with millions of training samples. The denoising experiments conducted with both large-scale image and training dataset demonstrate that our proposed method leads to faster performance and better dictionaries.* **Keywords:** Image denoising, Sparse representation, Incremental learning, Overcomplete dictionary

1. **Introduction.** Images are one of the significantly important ways to get information for us. However, in the practical application, images are often suffering from a variety of noise, so that denoising of images becomes one of the most basic tasks of image processing which has been extensively studied in the past several decades. Denoising is the simplest problem among the family known as Inverse Problems, aiming to recover a high quality signal from a degraded version of it. In this paper, we address the classic image denoising problem: An ideal image is measured in the presence of an additive zero-mean white and homogeneous Gaussian noise, with standard deviation. The image denoising problem is important, not only because of the evident applications it serves. Being the simplest possible inverse problem, it provides a convenient platform over which image processing ideas and techniques can be assessed. Indeed, numerous contributions in the past several years addressed this problem from diverse points of view. Statistical estimators, spatial adaptive filters, stochastic analysis, partial differential equations, transform-domain methods and other approximation theory methods are some of the many directions explored in studying this problem. In this paper, we have no intention to provide a survey of this activity. Instead, we intend to explore the method that utilizes sparse and redundant representations over trained overcomplete dictionaries for image or image sequence denoising, extending the work reported by Elad et al. [1] . Recently several algorithms have been proposed for learning such

sparse representation dictionaries from the training data such as method of optimal directions (MOD) [2] , K-SVD [3] and etc. Researchers have shown that these algorithms can achieve outstanding performances in image compression [4] and denoising.

However, most of these methods for dictionary learning are iterative batch algorithms, which deal with all the training samples at each iteration to minimize the objective function under sparsity constraints. Therefore, another problem we may encounter is that when the training set becomes very large, these methods are no longer efficient. To overcome this bottleneck, an online algorithm for dictionary learning which applies stochastic approximation method has been proposed in the literature [5] . To address these issues, we propose an incremental learning approach that processes one sample (or a small subset) of the training set at a time. This is particularly important in the context of image and video processing, where it is common to learn dictionaries adapted to training data that may include several millions of small patches. Our proposed approach is expected to lead to benefits both in the denoising performance and the computational complexity, when compared to previous batch dictionary learning algorithm.

The outline of this paper is organized as follows. In section II, we briefly introduce the principles of sparse and redundant representations and their deployment to image denoising. Then we discuss the batch dictionary learning algorithm and propose the incremental dictionary learning method for large-scale training dataset in section III. In section IV, the experimental results are given compared to classic competitive denoising algorithms with different kinds of dataset. Conclusions and future works are presented in the final section.

2. **Denoising via Sparse Representation.** Consider the problem of estimation of x from the observed signal y with additional noise n, thus

$$\mathrm{y}=\mathrm{x}+\mathrm{n}, \quad \text{where} \quad \mathrm{n} \sim N\{0, \ \sigma^2 \mathrm{I}\} \tag{1}$$

Where, n denotes the observation zero-mean white Gaussian noise and I denotes the identity matrix. We desire to design an algorithm that can remove the noise from y, getting as close as possible to the original image x. image denoising method based on sparse representation assume that x has a sparse representation over an overcomplete dictionary $\Phi$, i.e. $\mathrm{x}=\Phi\alpha$ with a small $\|\alpha\|_0^0$ (the number of nonzero elements of a vector) and also assume that a good estimation on the energy of the present noise $\|\mathrm{n}\|_2^2 \le \varepsilon^2$ is provided.

The sparsest representation we are looking for, is simply

$$\hat{\alpha} = \arg\min_{\alpha} \|\alpha\|_0^0, \ s.t. \|\mathrm{y}\text{-}\Phi\alpha\|_2^2 \le \varepsilon^2 \tag{2}$$

Some pursuit algorithm such as orthogonal matching pursuit (OMP) [6] and basis pursuit (BP) [7] algorithms can be used to resolve this optimization problem. Once the sparsest solution of Eq. 2 has been found with the stated algorithm, we can recover the approximate image by $\hat{\mathrm{x}}=\Phi\hat{\alpha}$. Sparse representation of noised image is conducted on the trained dictionary or predefined dictionary. In summary, image denoising includes three steps:

1) Training samples: The training data were constructed as a set of block patches of size 8x8 or 16x16 pixels, taken from the noised image or a database of natural images. Each block is rearranged into a column y and get the matrix $Y=[y_1, y_2 ... y_n]$.

2) Dictionary training: We applied the MOD, K-SVD or our proposed incremental learning algorithm to train an adaptive overcomplete dictionary of size L. The coefficients were

computed using OMP with a fixed number, where the maximal number of coefficients is K. Note that better performance can be obtained by switching to Lasso [8] . We concentrated on OMP because of its simplicity and fast execution.

3) Image denoising via sparse representation: We conduct adaptive sparse representation of noised image and get denoised x', rearrange each column of x' into an image block, combine all the blocks and average the overlapped pixel, get the denoised image $\hat{x}$ .

3. **Dictionary learning.** The MOD or K-SVD algorithms is both batch dictionary learning algorithms often used to find the optimal dictionary D that leads to the lowest reconstruction error given a fixed sparsity factor K. In practice, it has been observed that K-SVD converges with less number of iterations than MOD. It motivates us to select the K-SVD algorithm to be compared with our proposed incremental method. We now briefly describe the K-SVD algorithm which was proposed by Aharon et al. [3] .

The K-SVD algorithm aims to iteratively improve the dictionary to achieve sparser representations of the signals in Y, by solving the optimization problem.

$$\min_{D,X}\{\|Y - D\mathrm{A}\|_F^2\} \quad s.t. \forall i, \|\alpha_i\|_0 \leq K \tag{3}$$

The K-SVD algorithm summarized in Algorithm 1 uses two basic steps, which together constitute the algorithm iteration and each iteration involves the all training samples: (i) the signals in Y are sparse coded given the current dictionary, producing the sparse representations matrix A, and (ii) the dictionary atoms are updated given the current sparse representations. The sparse coding part is commonly implemented using OMP or Lasso. The dictionary update is performed one atom at a time, optimizing the target function for each atom individually while keeping the rest fixed. The problem can be solved directly via SVD decomposition.

Algorithm 1 K-SVD

1: Input: Signal set Y, initial dictionary $D_0$, target sparsity K, number of iterations k.
2: Output: Dictionary D and sparse matrix A such that $Y \approx DA$
3: Init: Set D:= $D_0$
4: for n = 1 … k do
5: $\quad \forall \mathrm{i}: \ \alpha_{\mathrm{i}} := \mathrm{Arg}\min_{\alpha}\{\|\mathrm{y}_{\mathrm{i}} - D\alpha\|_2^2\} \quad s.t. \|\alpha\|_0 \leq \mathrm{K}$
6:     for j = 1…L do
7:         Apply SVD decomposition to update $D_j$
8:     end for
9: end for

Large-scale training set is a reasonable extension from human beings in learning from experiences such as large image denoising or video inpainting task. However the aforementioned batch dictionary learning schemes fail to handle large-scale dataset problem. For this reason, we explored the incremental dictionary learning method for large-scale training set. Inspired by Bottou [9] and Mairal [10] use the expected objective function to replace the original empirical objective function, we obtain a novel dictionary learning problem:

$$\min_{D}\frac{1}{2}\mathrm{E}_{\mathrm{y}}[\|y - D\hat{\alpha}\|_2^2] \tag{4}$$

Where, $\hat{\alpha}$ denotes the sparse coefficients computed in the sparse coding stage. To solve the above problem, we propose an incremental learning algorithm for dictionary updating which processes one training sample or a small subset at a time.

The approach we propose in this paper is a block-coordinate descent algorithm. The overall algorithm is summarized in Algorithm 2. In this algorithm, the i.i.d. samples $y_t$ are drawn from an unknown probability distribution p(y) sequentially. However, since the distribution p(y) is unknown, obtaining such i.i.d. samples may be very difficult. The common trick in our algorithm to obtain such i.i.d. samples is to cycle over a randomly permuted training set. The sparse coding steps are same as K-SVD algorithm. The dictionary update steps are different from the K-SVD algorithm. The updated dictionary $D_t$ is computed by minimizing the following cost function.

$$f_t(D) \triangleq \frac{1}{t}\sum_{i=1}^{t}\frac{1}{2}\left\|y_i - D\alpha_i\right\|_2^2 + \lambda\left\|\alpha_i\right\|_1^1 \quad (5)$$

Where, the vectors $\alpha_i$ are computed during the previous spare coding steps of the algorithm.

Algorithm 2 Incremental dictionary learning

1: Input: $y \in R^m \sim p(y)$ (a strategy to draw i.i.d samples of p), $\lambda \in R$ (regularization parameter), T (number of iterations).
2: Output: Dictionary D
3: Init: Set D:= $D_0 \in R^{m\times k}$ (initial dictionary), $A_0$:= 0, $B_0$:= 0 (reset the “past” information).
4: for t = 1 to T do
5: Draw $y_t$ from p(y).
6: Sparse coding: compute $a_t$ using OMP
7: $A_t \leftarrow A_{t-1} + a_t a_t^T$
8: $B_t \leftarrow B_{t-1} + y_t a_t^T$
9: Compute $D_t$ according to Eq. 5, with $D_{t-1}$ as warm restart
10: end for

4. **Experimental Validation.** In this section, we present experiments on natural images to demonstrate the efficiency of incremental dictionary learning and verify our method to the application of image denoising. Firstly, incremental dictionary training experiments are conducted on a standard image dataset. Secondly, we apply the proposed method to denoise a large-scale image with zero-mean white and homogeneous Gaussian noise.

4.1. **Incremental training on the image patches.** For our experiments, we randomly select $1\times10^6$ patches with size $8\times8$ from the standard image dataset showed as Figure 1 to form the training dataset. These patches are equally divided into two subsets A and B. We use the two subsets to incrementally train a $64\times256$ dictionary and assess the performance.

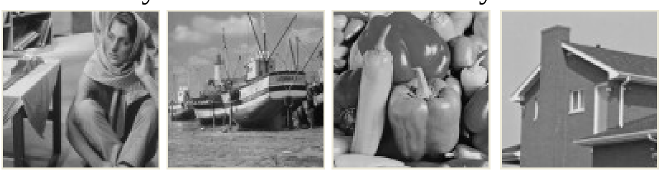

FIGURE 1. The images used for training the dictionary.

In the following dictionary training process, we normalize the patches to have unit $\ell_2-norm$ and $\lambda=1.2$ in all of our experiments. We use the first part of the training set to train a dictionary. Then we use subsets B to continually train the dictionary. The learned dictionaries are showed as Figure 2. In the first training stage, the time of computation for dictionary learning is 18.467 seconds and the value of loss function of Eq. 5 is 0.2745. In the second training stage, the time of computation for dictionary learning is 19.11 seconds and the value of loss function is 0.2735.

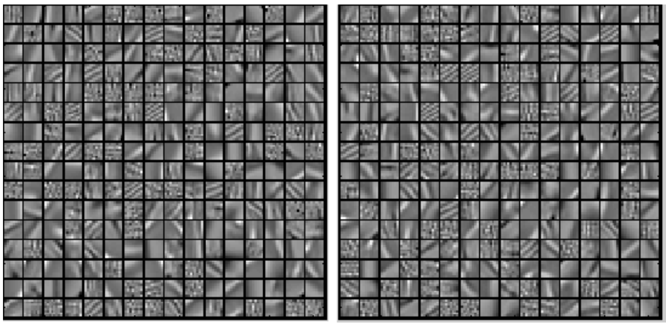

FIGURE 2. (left) Learned dictionary for 8×8 image patches, trained using the first half of the training set. (right) Learned dictionary for 8×8 image patches, continually trained using the second half of the training set.

4.2. **Denoising on large-scale image.** Our last experiment demonstrates that our algorithm can be used for a large-scale image denoising task from the corrupted 4-Megapixel image (2560×1600) of Figure 3. We learn an overcomplete dictionary with 1024 elements using the roughly $1\times10^6$ blocks with size 16×16 from the noisy image. Once the dictionary has been learned, we denoise the corrupted image, averaging the denoised blocks when they overlap in the result, using the sparse representation denoising technique described in section II. Our intent here is of course not to evaluate our learning procedure in denoising tasks, which would require a thorough comparison with state-the-art techniques on standard datasets. Instead, we just wish to demonstrate that the proposed method can indeed be applied to a realistic, non-trivial denoising task on a large-scale image.

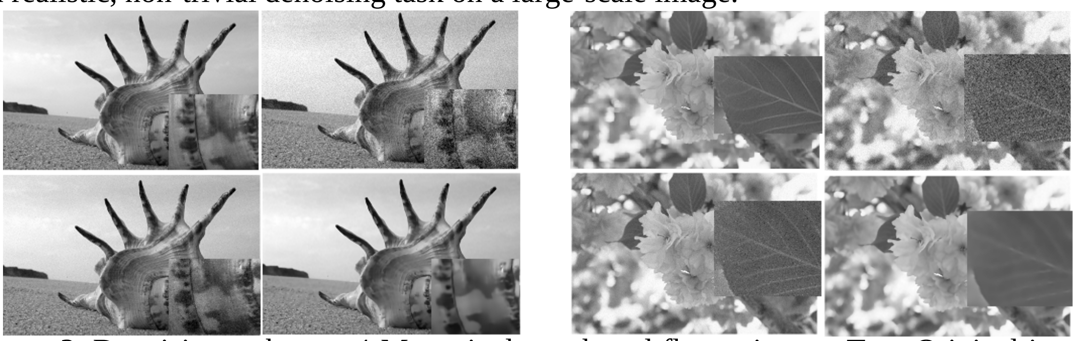

FIGURE 3. Denoising task on a 4-Megapixel conch and flower image. Top: Original image (left) and noisy image (right). Bottom: Wiener filter denoised image (left) and our proposed method denoised image (right).

The denoising performances are reported in Tables 1-2 compared to classic wiener filter. The experimental results in Table 1 shows the peak signal-to-noise ratios (PSNRs) of the noisy image, wiener filter denoised image and our proposed method denoised image conducted on different images. The experimental results in Table 2 shows the PSNRs of the noisy image, wiener filter denoised image and our proposed method denoised image conducted on the same image with different noise size. As can be seen, our proposed method outperforms the classic image denoising filter in different settings.

TABLE 1. Denoising performance on different image

| | noisy image | wiener filter | our proposed method |
|---|---|---|---|
| conch image | 22.11db | 29.47db | 33.67db |
| flower image | 22.11db | 29.52db | 36.41db |

TABLE 2. Denoising performance on different noise size

| | noisy image | wiener filter | our proposed method |
|---|---|---|---|
| conch image | 28.12db | 35.31db | 36.97db |
| conch image | 24.61db | 31.95db | 35.11db |
| conch image | 22.11db | 29.47db | 33.67db |
| conch image | 20.17db | 27.53db | 32.56db |

5. **Conclusions.** In this paper we introduce a new image denoising framework based on sparse representation using incremental learning algorithm for obtaining dictionaries adapted to large-scale training dataset. Preliminary experiments demonstrate that it is effective on image denoising tasks that may involve millions of training samples. In our future work, we plan to use the proposed learning framework for image sequence and video restoration tasks with variable datasets size.