A STUDY ON CLUSTERING-BASED IMAGE DENOISING: FROM GLOBAL CLUSTERING TO LOCAL GROUPING

Mohsen Joneidi[a], *Mostafa Sadeghi*[a], *Mojtaba Sahraee-Ardakan*[a]
Massoud Babaie-Zadeh[a], *Christian Jutten* [b*]

[a]Electrical Engineering Department, Sharif University of Technology, Tehran, IRAN.
[b]GIPSA-Lab, Grenoble, and Institut Universitaire de France, France.

ABSTRACT

This paper studies denoising of images contaminated with additive white Gaussian noise (AWGN). In recent years, clustering-based methods have shown promising performances. In this paper we show that low-rank subspace clustering provides a suitable clustering problem that minimizes the lower bound on the MSE of the denoising, which is optimum for Gaussian noise. Solving the corresponding clustering problem is not easy. We study some global and local sub-optimal solutions already presented in the literature and show that those that solve a better approximation of our problem result in better performances. A simple image denoising method based on dictionary learning using the idea of gain-shaped K-means is also proposed as another global suboptimal solution for clustering.

***Index Terms*—** Image denoising, data clustering, dictionary learning, sparse representation

1. INTRODUCTION

Consider the problem of estimating a clean version of an image contaminated with additive white Gaussian noise (AWGN). A general approach for denoising is to divide the noisy image into some (overlapping) small blocks, then de-noise each block, and finally obtain the overall estimate of the clean image by averaging the de-noised blocks [6].

The signal model is as follows,

$$\mathbf{y}_i = \mathbf{z}_i + \mathbf{n}_i \tag{1}$$

where $\mathbf{y}_i \in \mathbb{R}^N$ is the vector form of the ith block of the noisy image, $\mathbf{z}_i$ is the vector form of the ith block of the original image, and $\mathbf{n}_i$ is a zero-mean AWGN, each of its enteries having variance of σ^2.

Numerous methods have already been proposed in the literature for image denoising. Some methods are based on defining a neighbourhood for each block and weighted averaging, as in [1-4] which work in the spatial domain. The method proposed in [5], named as BM3D, is similar to [1-4], but its processing is performed in the frequency domain. BM3D constructs a three-dimensional matrix for each image block by grouping those two-dimensional blocks that are similar to it. Then, a 3D-DCT filtering is performed which provides a good estimate of the clean version of each block.

Elad and Aharon [6] suggested a new approach based on dictionary learning. They used K-Singular Value Decomposition (K-SVD) algorithm to produce a global dictionary using the noisy image blocks. The clean estimate of each de-noised block is estimated by decomposing noisy blocks in the obtained dictionary using a sparse coding algorithm.

Local grouping or similar blocks clustering are important factors in the success of some methods including [2], [5], [8] and [12]. Dictionary learning based denoising methods can also be seen as performing some kind of image blocks clustering. For example, K-SVD is a generalization of K-means clustering algorithm. So, the clustering or grouping has a crucial role in image denoising which will be studied in more details in this paper.

In this paper, we introduce the problem of efficient clustering based on minimizing the mean square error (MSE) lower bound of denoising derived in [11]. Although the minimization of this lower bound does not guarantee that the estimation error is minimized, some state of the art methods are in fact implicitly solving our derived problem. Some of the previously suggested clustering-based approaches are studied and their performances are compared by performing simulations. In this way, we actually provide a rough justification for why some clustering-based methods perform better than others. Moreover, inspired by the obtained problem, we propose a simple but efficient image denoising algorithm based on global clustering.

The rest of the paper is organized as follows. In Section 2 clustering of image blocks is studied and the efficient clustering problem is introduced. We then discuss some global clustering solutions for the resulting problem in Section 3. Section 4 studies some algorithms based on local grouping instead of global clustering. Finally, in Section 5, we numerically compare the performances of the solutions.

*This work has been partially funded by the Iran National Science Foundation (INSF) under Contract 91004600 and by the European project ERC-2012-AdG320684-CHESS.

2. NOISY IMAGE BLOCKS CLUSTERING

As mentioned in the previous section, grouping or clustering of similar blocks is an important factor in success of some recent works. The lower bound of the MSE of image denoising has been studied in [10] and [11]. This lower bound for a block $\mathbf{z}_i$ belonging to the kth cluster, whose data indices are in Ω_k, is as follows

$$E\left[\|\mathbf{z}_i - \widehat{\mathbf{z}}_i\|_2^2\right] \geq \text{trace}\left[\left(\mathbf{J}_i + \hat{\mathbf{C}}_k^{-1}\right)^{-1}\right] \tag{2}$$

where, $\mathbf{J}_i$ is the Fisher information matrix (FIM) for $\mathbf{z}_i$, $\widehat{\mathbf{z}}_i$ is the de-noised estimate of $\mathbf{z}_i$, and $\hat{\mathbf{C}}_k$ is the estimated covariance matrix for the kth cluster. For *i.i.d.* and zero-mean Gaussian noise, authors of [11] calculated $\mathbf{J}_i$ as

$$\mathbf{J}_i = \frac{|\Omega_k|}{\sigma^2}\mathbf{I},\ \forall i \in \Omega_k \tag{3}$$

where, $|\Omega_k|$ is the number of members of the kth cluster.

The question we are going to answer is "which type of clustering is efficient?". To this end, by simple calculations, we firstly rewrite the right side of (2) in terms of the eigenvalues of $\hat{\mathbf{C}}_k$

$$E\left[\|\mathbf{z}_i - \widehat{\mathbf{z}}_i\|_2^2\right] \geq \frac{\sigma^2}{|\Omega_k|}\sum_j \frac{\lambda_j^k}{\lambda_j^k + \frac{\sigma^2}{|\Omega_k|}} \tag{4}$$

where λ_j^k is the jth eigenvalue of the estimated covariance matrix of the k^{th} cluster. We introduce the following cost function which is a summation of the right-side of (4) over all clusters (note that each cluster contributes to the cost function proportional to the number of its members, so it is multiplied by $|\Omega_k|$)

$$f(\Omega) = \sum_{k=1}^{K}\sum_j \frac{\lambda_j^k}{\lambda_j^k + \frac{\sigma^2}{|\Omega_k|}} \tag{5}$$

where $\Omega = \{\Omega_1, ..., \Omega_K\}$. A clustering problem aims to find Ω, i.e., the set of indices indicating the members of each cluster. We define our efficient clustering problem as follows,

$$\min_{\Omega} f\left(\Omega\right) \tag{6}$$

Assume that $\sigma^2/|\Omega_k|$ is small compared to non-zero λ_j^k's. By this assumption, $\sum_j \frac{\lambda_j^k}{\lambda_j^k + \sigma^2/|\Omega_k|}$ approximates $\|\boldsymbol{\lambda}^k\|_0$ where $\|\cdot\|_0$ denotes the number of non-zero entries of a vector and $\boldsymbol{\lambda}^k = [\lambda_1^k, \ldots, \lambda_N^k]$. In other words, problem (6) clusters the data into some *low-rank subspaces* and guarantees that most of the eigenvalues are zero for each cluster. So, we propose to replace (6) with our new problem

$$\min_{\Omega} f_0\left(\Omega\right) \triangleq \sum_k \|\boldsymbol{\lambda}^k\|_0 \tag{7}$$

This problem tries to collect data in a cluster such that most of their covariance matrix eigenvalues be zero. Solving this problem is not easy. We study some sub-optimal solutions previously used in image denoising, with a new simple solution which is proposed by this paper. We also roughly compare them with the introduced clustering problem (7).

3. GLOBAL CLUSTERING

A well-known clustering method is the family of K-means clustering algorithms [13], which have been used by K-LLD [12] for image denoising. K-means clustering algorithm solves the following problem

$$\min_{\mathbf{D}} \sum_{k=1}^{K}\sum_{j\in\Omega_k} \|\mathbf{y}_j - \mathbf{d}_k\|_2^2 \tag{8}$$

where, $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_K]$. This problem can be written in the following form which is a matrix factorization

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2\ ,\ \forall i,j:\ \|\mathbf{x}_i\|_0 = 1,\ x_i^j \in \{0,1\} \tag{9}$$

where, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L]$ (L is the number of blocks), $\mathbf{x}_i$ is the ith column of $\mathbf{X}$, and x_i^j is the jth entry of $\mathbf{x}_i$. This problem implies that all entries of each $\mathbf{x}_i$ must be equal to zero except one of them. The non-zero element is forced to be 1. This restriction does not exist in the so-called gain-shaped variant of K-means [13], which solves the following problem

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \text{ subject to } \forall i:\ \|\mathbf{x}_i\|_0 = 1 \tag{10}$$

This problem is a K-rank1 subspace (K-lines) clustering. As can be seen in Fig. 1 (b) and (d), the obtained clusters by gain-shaped K-means is in agreement with problem (7). This is because only one eigenvalue of each cluster's covariance matrix is non-zero. For Fig. 1 (b) and (d) $f_0(\Omega_{\text{Kmeans}}) = 5$, $f_0(\Omega_{\text{gain shaped Kmeans}}) = 3$, respectively.

Inspired by the simple approach (10), a suboptimal solution for (7) can be obtained. We propose to construct the dictionary using the obtained cluster centroids and dominant principal components (PCs) of each cluster (generally, natural images do not perfectly lie on rank-1 subspace as in Fig. 1. So, the proposed dictionary also contains dominant PCs spanning details of each cluster). Those PCs should be added to the dictionary that their corresponding eigenvalues are greater than the noise variance. The noisy image blocks are then denoised in a way similar to the framework used in [6]. It will be shown in Section 5 that this leads to a fast and efficient denoising algorithm.

Another approach for clustering is *dictionary learning* in sparse signal representation, which aims to solve the following problem

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \text{ subject to } \forall i:\ \|\mathbf{x}_i\|_0 \leq \tau \tag{11}$$

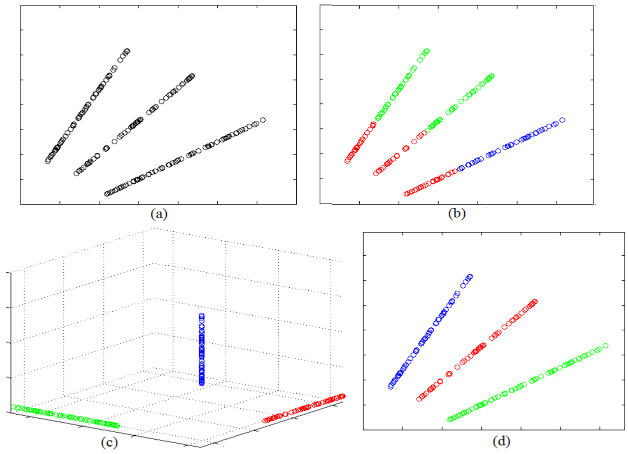

Fig. 1. Comparison of clustering in raw data domain and in the sparse-domain transformed data (as used in CSR and LSSC) for some 2D data. (a) Raw data. (b) K-means clustering on raw data (K=3). (c) K-means clustering on sparse-domain transformed data using an over-complete dictionary having 3 atoms. (d) Reconstruction of the data from their sparse representations in (c), in the case of these data Gain-shaped K-means directly results in (d).

Algorithm 1 Image denoising based on gain-shaped K-means

1: **Task** Denoise image patches $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{L}$ from AWGN ($\mathbf{y}_i = \mathbf{z}_i + \mathbf{n}_i$, $\forall i$).
2: Learning K cluster centroids using K-subspace [13].
3: Construct the dictionary, $\mathbf{D}$, by the obtained cluster centroids and significant PCs in each cluster.
4: Sparse code $\mathbf{y}_i$'s on $\mathbf{D}$: $\mathbf{y}_i \approx \mathbf{D}\mathbf{x}_i$.
5: Estimate $\mathbf{z}_i$'s by $\hat{\mathbf{z}}_i = \mathbf{D}\mathbf{x}_i$.
6: Construct the denoised image using $\{\hat{\mathbf{z}}_i\}_{i=1}^{L}$.

K-SVD is a well-known dictionary learning algorithm. Low-rank subspaces found by K-SVD have overlaps. It means that corresponding to each subset of the columns of $\mathbf{D}$, there is a low-rank subspace that K-SVD learns. Data that used the same subset lie on a low-rank subspace but K-SVD learns a very large number of low-rank subspaces for a set of training data such that many of them are empty or low populated (refer to Fig. 2, top). Actually, clusters found by K-SVD include the data that have used the same dictionary columns. Note that these clusters are not guaranteed to be low-rank. In the simulation results we will see that our proposed method based on gain-shaped K-means outperforms K-SVD.

4. LOCAL GROUPING

The derived problem (7) describes a suitable global clustering problem, while the state of the art algorithms do not perform global clustering, but instead use local patch-grouping.

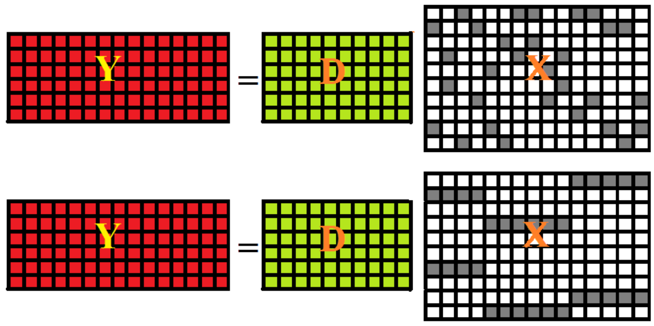

Fig. 2. Top: K-SVD approximates data by a union of rank-2 subspaces. No rank-2 cluster can be found. Bottom: Group sparsity constraint on $\mathbf{X}$. There are three rank-2 clusters.

Translating global clustering to local grouping converts the problem to,

$$\mathtt{G}_i = \min_{\mathtt{G}} \|\boldsymbol{\lambda}_{\mathtt{G}}\|_0 \text{ subject to } |\mathtt{G}| \geq \tau, \mathtt{G} \in \mathtt{W}_i\ , i \in \mathtt{G} \qquad (12)$$

where, $\mathtt{G}_i$ is group of blocks corresponding to the ith block, $\boldsymbol{\lambda}_{\mathtt{G}}$ is the vector of the eigenvalues of the covariance matrix of $\mathtt{G}_i$ and $\mathtt{W}_i$ is a window around the i^{th} block. The last constraint implies that the ith block must be a member of $\mathtt{G}_i$. An equivalent form of (12) can be written as,

$$\mathtt{G}_i = \max_{\mathtt{G}} |\mathtt{G}| \text{ subject to } \|\boldsymbol{\lambda}_{\mathtt{G}}\|_0 \leq \tau, \mathtt{G} \in \mathtt{W}_i\ , i \in \mathtt{G} \qquad (13)$$

BM3D, a high performance image denoising algorithm, implicitly uses (13) in order to perform local grouping. The similarity criterion used in BM3D for performing local grouping is novel, in which firstly blocks are transformed using an orthonormal transformation (e.g., DCT or DFT), then a projection onto a low-rank subspace is performed using hard-thresholding of the coefficients of each block. In the new transformed space, a simple Euclidean distance criterion determines similar blocks to the ith block. Similar blocks to the ith one lie nearly on a low-rank subspace, thus many of $\lambda_{\mathtt{G}_i}$'s are about zero and the constraint of (13) is satisfied.

The idea behind (13) can be used in another way different from what BM3D uses. These denoising algorithms first perform grouping using a rough criterion, e.g., Euclidean distance, then in the main denoising algorithm they obtain a low-rank representative for each group and use it. The algorithm suggested by Dong *et al.* (clustering based sparse representation or CSR) [7] which solves the following problem, is an example of these types of algorithms

$$\min_{\mathbf{X},\mathbf{B}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \gamma_1 \sum_i \|\mathbf{x}_i\|_0 + \gamma_2 \sum_{k=1}^{K} \sum_{j \in \mathtt{G}_k} \|\mathbf{x}_j - \mathbf{b}_k\|_2^2 \qquad (14)$$

where $\mathbf{B} = [\mathbf{b}_k]$, and $\mathbf{b}_k$ is the centroid of the kth group. Note that (14) does not optimize the dictionary. In fact, firstly a global dictionary using K-means and PCA is learned which is then used by this problem to simultaneously perform local grouping and sparse coding, in an iterative procedure. The first and second terms in (14) are similar to K-SVD problem, but the last term clusters the sparse-domain transformed data. Figure 3 illustrates the effect of clustering data in the sparse domain rather than the raw data. The traditional K-means is not able to cluster data in low-rank subspaces for raw data ($f_0(\Omega_{\text{Kmeans}}) = 5$), but in the case of the sparse-domain transformed data, it successfully performs clustering ($f_0(\Omega_{\text{CSR}}) = 3$). Contrary to K-SVD, in which the members of a cluster have used one column of $\mathbf{D}$, problem (14) encourages the clustering to put data that have the same sparse representations (structures) in one cluster.

Another local grouping based method is a novel approach, called learned simultaneous sparse coding (LSSC) [9], that simultaneously performs group sparse coding [14] and grouping the similar patches. Group sparse coding implies that the blocks within a group have similar sparse representations, like CSR. This is achieved by jointly decomposing groups of similar signals on subsets of the learned dictionary (as previously explained, K-SVD fails to achieve this goal. See Fig. 2 for comparison). They proposed the following cost function,

$$\min_{\mathbf{X}_k} \sum_{k=1}^{K} \|\mathbf{X}_k\|_{p,q} \text{ s.t. } \forall k: \sum_{i \in \mathsf{G}_k} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i^k\|_2 \leq \epsilon \qquad (15)$$

where, $\mathbf{X}_k$ is the coefficient matrix of the kth cluster, $\mathbf{x}_i^k$ is the ith column of $\mathbf{X}_k$, and $\|\mathbf{X}\|_{p,q} = \sum_i \|\mathbf{x}_{[i]}\|_q^p$, with $\mathbf{x}_{[i]}$ being the ith row of $\mathbf{X}$. Minimizing $\|\mathbf{X}\|_{p,q}$ with $p = 1$ and $q = 2$ (that is, the ℓ_1 norm of the vector containing the ℓ_2 norms of the rows) implies that the number of engaged rows of $\mathbf{X}$ will be limited. In other words, this cost function encourages the data to have the same coefficient supports in a cluster. As the data in the same cluster can be decomposed by few bases, the rank of the data matrix in the same cluster will be minimized. Thus a solution for (15) tries to minimize (12). In other words, $\sum \|\mathbf{X}_k\|_{p,q}$ approximates $\|\boldsymbol{\lambda}_{\mathsf{G}k}\|_0$.

5. SIMULATION RESULTS

In this section, denoising results of some recent methods are presented. Comparisons are performed separately for global and local methods. K-SVD and our simple gain-shaped K-means are presented as global methods. The presented local methods include those introduced in [5], [7], [9] and [12]. Our method is simulated similar to the framework of [6]. Running time of K-subspaces (for identification of K-rank1 subspaces) is about 40% of K-SVD for 20,000 blocks extracted from a 512×512 image. Both algorithms have the same amount of error for the training set (depending on the noise variance) but their size of dictionary are different. Performance comparison of these algorithms can be seen in Table 1. we have used the Peak Signal to Noise Ratio (PSNR[1]) as the performance criterion. The PSNR values were averaged over 5 experiments, corresponding to 5 different realizations of AWGN.

Table 1. Image denoising performance of global methods in PSNR (dB). In each cell, right: K-SVD [12], left: our simple method based on gain-shaped K-means

σ/SNR	Lena		Barbara		House		Boat	
5/34.16	38.71	38.60	38.08	38.22	39.59	39.37	37.25	37.22
10/28.14	35.60	35.47	34.68	34.42	36.54	35.98	33.85	33.64
20/22.11	32.57	32.38	30.98	30.83	33.68	33.20	30.52	30.36

Table 2. Image denoising performance of local methods in PSNR (dB). Upper right: K-LLD [12], upper left: LSSC [9], bottom right: CSR [7], bottom left: BM3D [5]

σ/SNR	Lena		Barbara		House		Boat	
5/34.16	38.69	38.01	38.48	37.26	39.93	37.63	37.35	35.96
	38.72	38.74	38.31	38.43	39.83	39.98	37.28	37.31
10/28.14	35.83	35.20	34.97	33.30	36.96	35.09	34.02	33.16
	35.93	35.90	34.98	35.10	36.71	36.88	33.92	33.88
20/22.11	32.90	32.37	31.57	28.93	34.16	32.66	30.89	30.17
	33.05	32.96	31.75	31.78	33.77	33.86	30.88	30.78

In Table 2, the results of local methods are compared. As can be seen, the methods of [7] and [9] discussed in section 3, show good performances. Recently [15] investigated a comprehensive comparison of different image denoising methods. They have shown numerically that BM3D, SCR and LSSC have the best results.

In natural images, far away blocks have generally different patterns, so, using all blocks may result in inappropriate clustering. Moreover, non-overlapped clusters obtained by global methods are not as flexible as the overlapped groups. On the other hand, local grouping assigns appropriate groups to each block. Although local methods have better performance, global methods are able to extract salient features of images and use it easily for de-nosing. By comparing the results of tables 1 and 2, we see that the performance of the proposed global method is just about 0.5dB lower than promising local methods, which is not a high difference. However, a common good property of both global and local methods is that they exploit the low-dimensional characteristics of clusters/groups in order to design a suitable denoising algorithm.

[1] PSNR is defined as $10\log_{10}(255^2/\text{MSE})$ and measured in dB.

6. CONCLUSION

This paper studied the problem of image denoising based on clustering. Our goal was to obtain an appropriate clustering to be exploited in denoising algorithms. We derived low-rank subspace clustering as an efficient clustering problem and suggest a new simple solution. We also studied some existing solutions that approximately solve the resulting problem by global clustering or local block grouping. Clustering based on sparse representations, as used by some previous works, is a good idea to solve low-rank subspace clustering. We saw that the state of the art methods for denoising are based on local grouping and they approximate our derived problem in order to obtain suitable groups of blocks.

7. REFERENCES

[1] D. van de Ville, and M. Kocher, "SURE-based non-local means," *IEEE Signal Process. Letters*, vol. 16, no. 11, pp. 973–976, Nov. 2009.

[2] T. Tasdizen, "Principal neighbourhood dictionaries for non-local means image denoising," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2649–2660, Dec. 2009.

[3] R. Vignesh, B. T. Oh, and C.-C. J. Kuo, "Fast non-local means (NLM) computation with probabilistic early termination," *IEEE Signal Process. Letters*, vol. 17, no. 3, pp. 277–280, Mar. 2010.

[4] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel Regression for Image Processing and Reconstruction," *IEEE Trans. on Image Process.* vol. 16, no. 2, pp. 349–366, Feb. 2007.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[6] M. Elad, and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[7] W. Dong, X. Li, D. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.457–464, June 2011.

[8] M. Sadeghi, M. Joneidi, M. Babaie-Zadeh, and C. Jutten, "Sequential subspace finding: A new algorithm for learning low-dimensional linear subspaces," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, Marrakesh, Morocco, 2013.

[9] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proceedings of IEEE International Conference on Computer Vision*, pp. 2272–2279, 2009.

[10] P. Chatterjee, and P. Milanfar, "Patch-based near-optimal image de-noising," *IEEE Trans. on Image Process.*, vol. 21, no. 4, pp. 1635–1649, April 2012.

[11] P. Chatterjee, and P. Milanfar, "Practical bounds on image de-noising: From estimation to information," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1221–1233, May 2011.

[12] P. Chatterjee and P. Milanfar, "Clustering-based denoising with locally learned dictionaries," *IEEE Trans. on Image Process.*, vol. 18, no. 7, pp. 1438–1451, 2009.

[13] A. Gersho, and R. M. Gray, Vector Quantization and Signal Compression, Springer, 1992.

[14] J. A. Tropp, "Algorithms for simultaneous sparse approximation," *Proc. of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

[15] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Transactions on Cybernetics,* vol. 44, no. 7, pp. 1001–1013, 2014.