IMPROVING K-SVD DENOISING BY POST-PROCESSING ITS METHOD-NOISE

Yaniv Romano

Department of Electrical Engineering
Technion, Haifa 32000, Israel
yromano@tx.technion.ac.il

Michael Elad

Department of Computer Science
Technion, Haifa 32000, Israel
elad@cs.technion.ac.il

ABSTRACT

Various patch-based image denoising algorithms have been shown to be very effective. Nevertheless, in most cases the difference between the noisy image and its denoised version (called "method-noise") still contains traces of the original image content. In this paper we propose a novel technique for improving the K-SVD denoising results. Our scheme starts by applying the K-SVD on the given noisy image. Then, for each patch, we recover the "stolen" image content information from the method-noise by performing iterations of denoising using the same atoms that represent the first-stage denoised patch. Experimental results demonstrate the efficiency of this technique.

***Index Terms*—** Image denoising, method-noise, sparse representations, dictionary, K-SVD

1. INTRODUCTION

Cleaning additive noise from signals or images (known as denoising) is a classical and long-studied problem in signal processing. Consider a given measurement signal $\mathbf{y} \in \mathbb{R}^n$ obtained from the clean signal $\mathbf{x} \in \mathbb{R}^n$ by a contamination of the form $\mathbf{y} = \mathbf{x} + \mathbf{v}$. In this paper we shall restrict our discussion to zero mean i.i.d. Gaussian noise $\mathbf{v} \in \mathbb{R}^n$, where σ is known. The denoising goal is to recover $\mathbf{x}$ from $\mathbf{y}$, which can be viewed as the need to separate between the original signal $\mathbf{x}$ and the additive noise $\mathbf{v}$.

Signal/image processing algorithms rely heavily on data-models (also referred to as priors in a Bayesian context) [1, 2]. The evolution of these models and the progress made in recent years resulted in a massive improvement in many practical applications, and in particular denoising. There are many noise cleaning algorithms, and many of today's state-of-the-art methods share in common the fact that they are patch-based, e.g. the K-SVD [3], BM3D [4], NLM [5], LSSC [6], many variants of these, and there are others [7–13]. In this paper we will pay special attention to the K-SVD and propose a method to boost its performance.

This research was supported by the European Research Council under EU's 7th Framework Program, ERC Grant agreement no. 320649, and by the Intel Collaborative Research Institute for Computational Intelligence.

(a) (b)

Fig. 1. K-SVD denoised image (left) and the obtained method-noise (right) for Lena with $\sigma = 25$.

The difference between the noisy and the denoised image is often referred to as the "method-noise" or the "residual" image, given by $\Delta^0 = \mathbf{y} - \hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is the denoised version of $\mathbf{y}$. Restricted to the additive Gaussian case, the work in [14] suggested to assess denoising quality based on visually inspecting the method-noise: The less image structures seen in it, the better the denoising performance. Indeed, even state-of-the-art denoising techniques contain traces of the image content in the method-noise due to imperfect denoising. Fig. 1 demonstrates this for the K-SVD result.

Motivated by this fact, the work reported in [15–18] relies on the method-noise in order to obtain improved denoising performance. The work in [15] introduces image quality measures that are based on the method-noise without using a reference-image. They suggest an iterative technique to extract the residual information from the method-noise based on an adaptive Wiener filter [19]. An improvement of the NLM [5] is suggested in [16] by modifying the weight function and adopting similar ideas to the ones developed in [15]. The work in [17] proposes a user-interactive approach which applies a variant of the bilateral filter [20] that exploits the cleaned image and the method-noise in order to extract the residual information. The work reported in [18] builds on the BM3D [4], replacing its second (Wiener filter) layer by an adaptive spatial filtering similar to the one practiced in [17].

In this work we suggest a denoising improvement tailored specifically to the K-SVD algorithm [3]. K-SVD determines

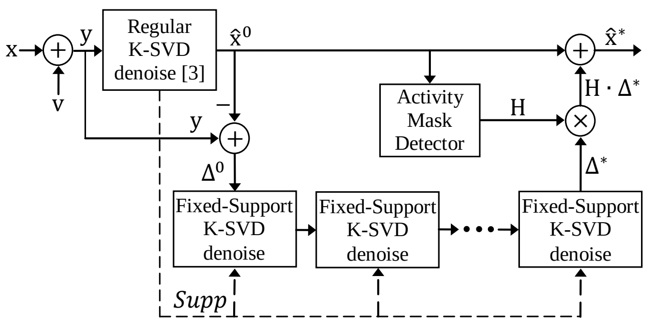

Fig. 2. Flowchart of the proposed iterative technique.

for each patch a small set of atoms that participate for its representation. Nevertheless, after patch-averaging, orthogonality of the overall residual to these atoms is lost. We propose to leverage this property in order to salvage "stolen" image content that resides in the method-noise.

This paper is organized as follows: In Section 2 we present some background material on K-SVD denoising, in Section 3 we present our novel denoising-boosting technique, and in Section 4 we demonstrate the improvements obtained for several test images and noise levels.

2. BRIEF BACKGROUND ON K-SVD

In this section, we briefly recall the K-SVD method for image denoising, based on sparse representation of its patches. The reader is referred to [2, 3] for more details.

Sparse and redundant representation modeling and dictionary learning using the K-SVD, lead to a highly effective denoising method. Such an algorithm divides the noisy image into $\sqrt{n} \times \sqrt{n}$ maximally overlapping patches, where n (e.g 64) is a fixed a-priori. Then, it trains a dictionary $\mathbf{D} \in \mathbb{R}^{n \times m}$ using the corrupted image itself, where $\mathbf{D}$ is composed of $m > n$ atoms. The next stage is to represent each noisy patch by a sparse composition over the trained dictionary using the Orthonormal Matching Pursuit (OMP). *This implies that the corresponding patch residual after denoising is necessarily orthogonal to the chosen atoms.* Finally the denoised image is obtained by averaging the cleaned patches. Note that this averaging ruins the orthogonality mentioned above. The K-SVD algorithm can be described as the outcome of a minimization problem:

$$\{\alpha_{ij}, \mathbf{D}, \hat{\mathbf{x}}\} = \arg \min_{\alpha_{ij}, \mathbf{D}, \mathbf{x}} \lambda \|\mathbf{x} - \mathbf{y}\|_2^2 + \sum_{i,j} \mu_{ij} \|\alpha_{ij}\|_0 + \sum_{i,j} \|\mathbf{D}\alpha_{ij} - \mathbf{R}_{ij}\mathbf{x}\|_2^2,$$

where $\hat{\mathbf{x}}$ is the denoised image, $\mathbf{D}$ is the trained overcomplete dictionary, α_{ij} represents the sparse representation vector for the (i, j)-patch in $\hat{\mathbf{x}}$, and $\mathbf{R}_{ij}$ is a matrix that extracts the (i, j)-patch from the image. The notation $\|\alpha_{ij}\|_0$ stands for the number of the nonzeros in α_{ij}.

3. THE PROPOSED TECHNIQUE

Our scheme starts by applying K-SVD denoising on the noisy image $\mathbf{y}$, giving (i) $\hat{\mathbf{x}}^0$ the pre-refined denoised image, (ii) $\Delta^0 = \mathbf{y} - \hat{\mathbf{x}}^0$ the method-noise, and (iii) $Supp\{\alpha_{ij}\}$ the atoms participating in the representation of the first-stage denoised (i, j)-patch. The main idea behind our approach is the belief that the method-noise contains residual image content *that can be represented as a linear combination of the very same atoms used in the initial denoising stage.* This belief relies on the gap that exists between the local patch treatment and the final global outcome - patches are processed separately and then merged by averaging over their overlaps. Therefore, the eventual (global) residual is not orthogonal w.r.t. to the atoms used for the local filtering. The proposed process exploits this property to recover image content from the method-noise. Our algorithm iteratively cleans the method-noise by alternating between local patch processing and global averaging. Fig. 2 presents a block diagram of the proposed technique; we repeatedly scrub the method-noise image by applying a modified K-SVD, where the supports from the first denoising stage are used, such that the method-noise patches are projected onto these pre-chosen subspaces[1], followed by averaging the overlapped patches. Thus, this modified K-SVD is different from the regular K-SVD [3] in one major respect - no pursuit is needed. The following is a pseudo-code description of the proposed algorithm:

Init Stage:

(i) $\{Supp\{\alpha_{ij}\}, \mathbf{D}, \hat{\mathbf{x}}^0\} \leftarrow$ K-SVD($\mathbf{y}$), i.e. denoise the given noisy image using the K-SVD algorithm [3].

(ii) Set $k = 0$ and $\Delta^0 = \mathbf{y} - \hat{\mathbf{x}}^0$, $\hat{\mathbf{x}}^* = \hat{\mathbf{x}}^0$.

(iii) Compute H, a mask obtained by detecting the active regions in $\hat{\mathbf{x}}^0$.

Repeat several times:

Update Residual Stage:

(i) For each residual (i, j)-patch, $\Delta_{ij} = \mathbf{R}_{ij}\Delta^k$, compute:

$\beta_{ij} \leftarrow$ OMP($\Delta_{ij}, \mathbf{D}, Supp\{\alpha_{ij}\}, \delta$), i.e. allow a slight modification to the given α_{ij} support (more details below) in representing this error patch, and assign the sparse-coding β_{ij} to it.

(ii) Compute the global Δ^{k+1} residual image by averaging all the overlapping $\mathbf{D}\beta_{ij}$ patches by

$$\Delta^{k+1} = \left(\sum_{ij} \mathbf{R}_{ij}^T \mathbf{R}_{ij}\right)^{-1} \left(\sum_{ij} \mathbf{R}_{ij}^T \mathbf{D}\beta_{ij}\right)$$

Stopping Criterion:

(i) Set $\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^0 + \mathrm{H} \cdot \Delta^{k+1}$.

(ii) if $\left|Corr\left(\mathrm{H} \cdot \hat{\mathbf{x}}^{k+1}, \mathrm{H} \cdot (\mathbf{y} - \hat{\mathbf{x}}^{k+1})\right)\right|$ increases compared to the previous step, conclude.

[1] Actually, we slightly modify these supports - see hereafter.

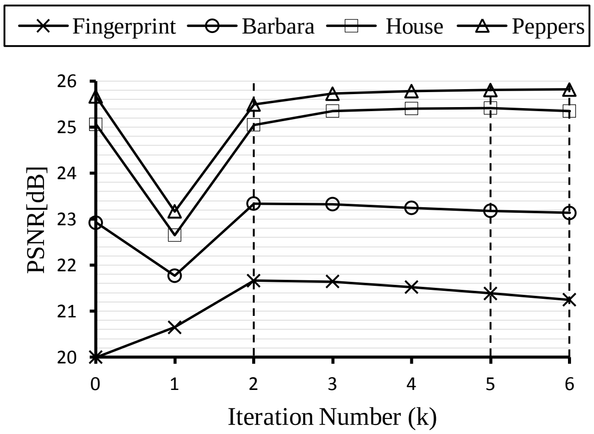

Fig. 3. PSNR [dB] of our iterative technique versus the iteration number for Fingerprint, Barbara, House and Peppers with $\sigma = 75$. Note that $k = 0$ stands for the regular K-SVD result.

The output of this algorithm is $\hat{\mathbf{x}}^*$. $Corr$ stands for a simple correlation, and $|\cdot|$ stands for absolute value (see details below), H is a matrix that indicates the image texture/active regions, and $\cdot$ denotes term-by-term matrix multiplication. The need for H comes from the fact that our method is mostly effective in texture and edge areas, and less so in smooth regions. As to the computational complexity of this algorithm, in each iteration the added complexity to the regular K-SVD is due to (i) pursuit of δ atoms ($\delta = 1$ in our experiments), and (ii) projection of the residual patches onto the chosen subspaces. These are far simpler than the core K-SVD algorithm. We now give more details on the modification of the support and the stopping criterion mentioned above.

3.1. Modifying the Initial Supports

While our proposed method can use the very same supports as found by the initial K-SVD algorithm, we found that allowing a slight modification to these supports may lead to further denoising improvement. For each patch, we suggest to expand the chosen set of atoms by additional δ ones. Assuming $\|\alpha_{ij}\|_0 = \mathrm{T}_{ij}$, we consider the following procedure:

(i) For each patch, create the matrix $\mathbf{D}_{ij}$ of size $n \times \mathrm{T}_{ij}$, its columns are atoms from $\mathbf{D}$ that correspond to the nonzero entries of α_{ij}. Then, find the representation vector by $z_{ij} = \left(\mathbf{D}_{ij}^T \mathbf{D}_{ij}\right)^{-1} \mathbf{D}_{ij}^T \Delta_{ij}$ where Δ_{ij} is the residual (i, j)-patch.

(ii) Expand z_{ij} support by sparse-coding additional δ atoms (from the remaining $m - \mathrm{T}_{ij}$ dictionary columns) by setting z_{ij} vector as initial solution for the OMP.

With $\beta_{ij} \leftarrow$ OMP $(\Delta_{ij}, \mathbf{D}, Supp\{\alpha_{ij}\}, \delta)$ we denote the above-mentioned algorithm for expanding α_{ij} support, where β_{ij} is the updated vector that represents Δ_{ij} patch over $\mathbf{D}$.

3.2. Stopping Criterion

In this section we propose a stopping criterion for the above algorithm in order to obtain the best denoising improvement. Fig. 3 demonstrates the need for such an estimation, as the quality may deteriorate at first, then improve, and finally converge back to zero improvement (since infinitely many steps lead to the original K-SVD outcome). It shows the Peak Signal to Noise Ratio (PSNR) versus the iteration number of the proposed technique. As can be seen, the iteration that obtains maximum PSNR value varies from one image to another despite of similar noise standard deviation. As the PSNR of the initial residual image is very low, projection of its patches onto the modified first-stage subspaces results in a PSNR deterioration. The multi-stage filtering scrubs the noise carefully and leads to a final PSNR improvement.

The original image $\mathbf{x}$, and the additive noise $\mathbf{v}$ are independent, and therefore we can assume that the less dependency between the denoised ($\hat{\mathbf{x}}$) and method-noise ($\mathbf{y} - \hat{\mathbf{x}}$) images, the better is the effective denoising performance (or image-noise separation). Following [15], we use the "Pearson's correlation test", which measures the dependency between two variables, x and v by $Corr(x, v) = \frac{\sigma_{xv}}{\sigma_x \sigma_v}$, where σ_x and σ_v are the standard deviations of x and v respectively, and σ_{xv} is their covariance. The closer $Corr(x, v)$ to zero, the less dependent x and v are. We propose the minimum absolute value of $Corr\left(\mathrm{H} \cdot \hat{\mathbf{x}}^k, \mathrm{H} \cdot (\mathbf{y} - \hat{\mathbf{x}}^k)\right)$ as our stopping criterion, where $\mathrm{H} \cdot \hat{\mathbf{x}}^k$ and $\mathrm{H} \cdot (\mathbf{y} - \hat{\mathbf{x}}^k)$ are the active regions of the k-iteration denoised and method-noise image, respectively.

4. EXPERIMENTAL RESULTS

In this section, detailed results of the proposed algorithm for various grayscale images and different noise standard deviation σ are presented. The parameters for both the regular and our iterative technique are based on the ones reported in [3]. We repeat our technique for 8 iterations and choose $\delta = 1$ atom. H is obtained by (i) applying "Canny edge detector" [21] on $\hat{\mathbf{x}}^0$, and then (ii) dilating the result. Fig. 4 (g) demonstrates H mask for the image Barbara.

Table 1 lists the PSNR of the K-SVD and our technique, the average improvement, ("Boost" column) and the stopping criterion estimation error (ϵ column). This error is defined as ϵ =PSNR($\hat{\mathbf{x}}^*_{\mathbf{ref}}$)−PSNR($\hat{\mathbf{x}}^*_{\mathbf{no-ref}}$), where $\hat{\mathbf{x}}^*_{\mathbf{ref}}$ is obtained by "scrubbing" the residual image until reaching maximum PSNR value, and $\hat{\mathbf{x}}^*_{\mathbf{no-ref}}$ is obtained by using the no-reference stopping criteria estimator, as described above. From Table 1 we can see that our method improves the K-SVD consistently, and especially for σ in the range of 25 to 100. Notice the robustness and accuracy of the proposed stopping criteria estimator - the estimation error is almost negligible. Table 2 shows that with the choice of adding one atom to each patch, we get the best improvement. Note that

σ/PSNR	Barbara		Couple		Fingerprint		House		Boats		Average		Boost	ϵ
15/24.61	32.51	**32.59**	31.52	**31.59**	30.05	**30.17**	34.38	**34.39**	31.81	**31.88**	32.06	**32.12**	0.06	0.03
25/20.18	29.58	**29.88**	28.91	**29.13**	27.24	**27.52**	32.20	**32.21**	29.34	**29.51**	29.45	**29.65**	0.20	0.01
50/14.16	25.41	**25.89**	25.27	**25.69**	23.22	**23.89**	28.02	**28.22**	25.94	**26.20**	25.57	**25.98**	0.41	0.03
75/10.63	22.92	**23.13**	23.56	**23.79**	19.99	**21.65**	25.05	**25.40**	24.02	**24.20**	23.11	**23.64**	0.53	0.06
100/8.14	21.85	**21.89**	22.61	**22.68**	18.30	**20.00**	23.64	**23.78**	22.85	**22.93**	21.85	**22.26**	0.41	0.04

Table 1. Summary of the denoising PSNR [dB] results. In each cell two denoising results are reported. Left: results of regular K-SVD, Right: results of the proposed technique using stopping criteria statistical estimator. We highlighted the best results for each pair. The last three columns present the average results, denoising improvement ("Boost"), and estimation error (ϵ), over all images.

δ	-1	0	1	2	3	4
Avg. Boost	0.06	0.12	0.32	0.32	0.28	0.25

Table 2. The K-SVD boosting as a function of δ.

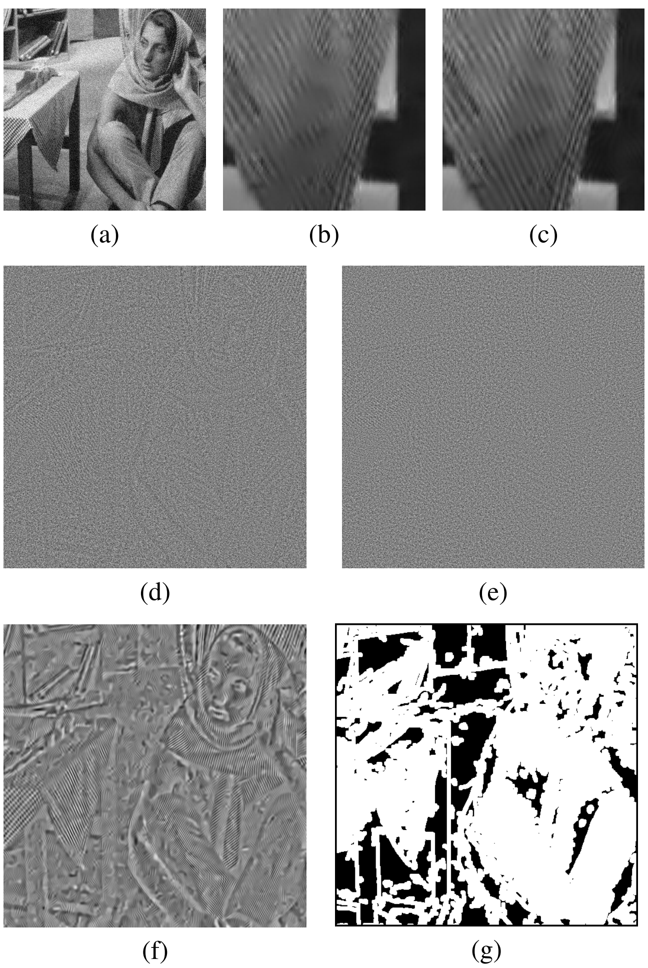

Fig. 4. Comparison of the regular K-SVD and our technique with $\sigma = 25$. (a) Noisy image, (b) Zoom on the regular K-SVD result, (c) Zoom on our result, (d) Regular K-SVD method-noise, (e) The method-noise after our processing, (f) The extracted residual information from (d), and (g) H the active region mask. (includes 75.2% of the image area). Note that we apply linear contrast stretch on (d)-(g) images.

in contrast to the above, adding a single atom to the regular K-SVD deteriorates the performance, with an average PSNR degradation of 0.65 dB. Fig. 4 demonstrates a visual improvement for the image Barbara: The regular K-SVD method-noise image contains residual structures (d). Multiplying Δ^* image (f) by H (g) and then adding the result back to $\hat{\mathbf{x}}^0$ (b) restores some of the tablecloth texture (c), and less image residual content appears in the final method-noise (e).

We performed several comparative tests between our scheme and the ones reported in [15–18]. We found that applying an adaptive Wiener filter on the K-SVD method-noise, as in [15, 16], is ineffective due to the very low Signal to Noise Ratio (SNR) of this image[2]. We tested the modified bilateral technique in [17, 18] using our activity map H as a replacement for the interactive part in [17] and the edge-detection done in [18]. The parameters of this K-SVD post-processing algorithm were optimized, leading to an average improvement of 0.15 dB (to be compared to the 0.32 dB that our method gets).

5. CONCLUSION AND FURTHER WORK

In this work we present a simple and efficient method for boosting the performance of the K-SVD image denoising algorithm. The proposed technique exploits the gap that exists between the local processing of the image patches and the eventual global averaging that generates the final outcome. Our algorithm boosts the overall denoising performance by extracting image-content information from the method-noise, leveraging on the very same atoms chosen for the denoising of the patches in the original algorithm. We applied a variant of the proposed technique on the method-noise images of the NLM and the first step of BM3D and obtained an effective boosting. These efforts are inspired by the seminal work on the BM3D [4], and the recently published work by Talebi, Zhu, and Milanfar [23], both offering systematic ways to improve denoising results by a second-layer processing stage that leverages on the initial denoised result.

ACKNOWLEDGEMENTS

We thank the authors of [15] for sharing their code with us.

[2]The work in [15, 16] tested this idea with the Total-Variation [22] and NLM [5] method-noise respectively, which leave more image-content.

6. REFERENCES

[1] M. Elad, "Sparse and redundant representation modeling - what next?," *IEEE Signal Processing Letters*, vol. 19, pp. 922–928, 2012.

[2] M. Elad, "Sparse and redundant representations: From theory to applications in signal and image processing," Springer, 2010.

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

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[5] A. Buads, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE CVPR*, 2005, vol. 2, pp. 60–65.

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proc. IEEE ICCV*, 2009, pp. 2272–2279.

[7] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, pp. 1531–1549, Apr. 2010.

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

[9] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, Feb. 2007.

[10] C. Kervrann and J. Boulanger, "Optimal spatial adaptation for patch-based image denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 2866–2878, Oct. 2006.

[11] J. Boulanger, C. Kervrann, and P. Bouthemy, "Space-time adaptation for patch-based image sequence restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1096–1102, Jun. 2007.

[12] M. Van De Ville and D. Kocher, "Nonlocal means with dimensionality reduction and sure-based parameter selection," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2683–2690, Sep. 2011.

[13] A. Buades, M. Lebrun, and J.-M. Morel, "Implementation of the non-local bayes image denoising algorithm," *Image Processing On Line*, 2012.

[14] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," vol. 4, no. 2, pp. 490–530, 2005.

[15] D. Brunet, E. R. Vrscay, and Z. Wang, "The use of residuals in image denoising," in *Proc. International Conference on Image Analysis and Recognition*, 2009, pp. 1–12.

[16] B. Xiong and Z. Yin, "Structural similar patches for nonlocal-means with modified robust m-estimator and residual images," in *Proc. IEEE ICMA*, 2011, pp. 709–714.

[17] J. Chen, C. K. Tang, and J. Wang, "Noise brush: Interactive quality image-noise separation," *ACM Transactions on Graphics*, vol. 28, no. 5, Dec. 2009.

[18] Y. Pyo, R-H Park, and S. Chang, "Noise reduction in high-iso images using 3-d collaborative filtering and structure extraction from residual blocks," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 2, pp. 687–695, 2011.

[19] J.-S. Lee, "Digital image enhancement and noise ltering by use of local statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 165–168, 1980.

[20] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE ICCV*, 1998, pp. 839–846.

[21] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.

[22] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, pp. 259–268, 1992.

[23] H. Talebi, X. Zhu, and P. Milanfar, "How to saif-ly boost denoising performance," *IEEE Transactions on Image Processing*, 2012.