Residual Correlation Regularization Based Image Denoising

Gulsher Baloch, Huseyin Ozkaramanli, and Runyi Yu, *Senior Member, IEEE*

***Abstract*—Patch-based denoising algorithms aim to reconstruct the clean image patch leaving behind the residual as contaminating noise. The residual should possess statistical properties of contaminating noise. However, it is very likely that the residual patch contains remnants from the clean image patch. In this letter, we propose a new residual correlation based regularization for image denoising. The regularization can effectively render residual patches as uncorrelated as possible. It allows us to derive an analytical solution for sparse coding (atom selection and coefficient calculation). It also leads to a new online dictionary learning update. The clean image is obtained through alternating between the two stages of sparse coding and dictionary updating. The performance of the proposed algorithm is compared with state-of-the-art denoising algorithms in terms of peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and feature similarity index (FSIM), as well as through visual comparison. Experimental results show that the proposed algorithm is highly competitive and often better than leading denoising algorithms. The proposed algorithm is also shown to offer an efficient complement to the benchmark algorithm of block-matching and 3D filtering (BM3D) especially.**

***Index Terms*—Correlation regularization, dictionary learning, image denoising, residual correlation, sparse representation.**

I. Introduction

THE main objective in patch-based image denoising algorithms that employ learned dictionaries is to make sure that atoms that best match clean image patch are picked. The K-SVD [4] denoising, for example, does this by projecting the noisy patch onto the dictionary atoms and picking the atom that gives the maximum orthogonal projection. As a result, at high noise levels, the residue usually contains structures from clean image patch; thus, it does not match the contaminating noise [2]. On the other hand, after the sparse coding stage is completed, the residual is expected to possess properties similar to those of contaminating noise. One such property is that the residues of different patches should be uncorrelated. This observation calls for processing patches in groups by considering local neighborhoods and making sure that the neighboring residuals are uncorrelated. Thus, in selecting atoms for a given patch, we determine sparse coefficient that leaves behind a residual that is as uncorrelated with the neighboring residuals as possible. The motivation of [2] is to highlight the importance of incorporating information about residual correlation in a dictionary learning process. Through a proposed strategy, it is shown that residual correlation reduction indeed can improve the performance of image denoising [2]. However, the sparse coding and dictionary update stages are not really based on residual correlation. Riot *et al.* [8] also proposed a variation in fidelity term to control the residual distribution. This is achieved by considering statistical moments of residual and the correlation on patches. It differs from our proposed algorithm in that it does not embed the framework of sparse representation via learned dictionaries. In [5], the correlation coefficient criterion is introduced to extract the meaningful structures from a noisy image. However, the approach adapted in [5] is different from the proposed algorithm here. There also exist other image denoising algorithms based on residual correlations such as [3], [7], [9], and [16]. None of these algorithms are based on a sparse-land model.

In this letter, we develop a new residual correlation regularization based image denoising algorithm. This regularization minimizes the correlation between neighboring residual patches. We derive an analytical solution for sparse coding (the atom selection and coefficient estimation). We also propose a new dictionary update that is based on the correlation regulation. The final clean image reconstruction is obtained via alternating between the sparse coding and dictionary update stages. Our experimental results show that the proposed algorithm is highly comparable and often superior to the state-of-the-art denoising algorithms.

The rest of the letter is organized as follows. The motivation and statement of the problem are given in Section II. The proposed algorithm is described in detail in Section III. Section IV presents experiments that compare the new algorithm with several state-of-the-art algorithms. And conclusions are drawn in Section V.

II. Motivation and Problem Statement

We consider the standard model for the image denoising problem: A clean image is corrupted by an additive uncorrelated noise. Let the image be partitioned into overlapping patches and each patch is arranged as column vector $\mathbf{x} \in \mathbb{R}^n$, which is modeled as

$$\mathbf{x} = \mathbf{x}^c + \mathbf{w} \tag{1}$$

where $\mathbf{x}^c$ is the clean patch and $\mathbf{w}$ is the noise. Given a dictionary $\mathbf{D}$ with atoms $\mathbf{d}_k$, $k = 1, 2, \ldots, K$. If $\mathbf{x}^c$ is approximately

Manuscript received October 31, 2017; accepted December 28, 2017. Date of publication January 1, 2018; date of current version January 16, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Glenn Easley. *(Corresponding author: Gulsher Baloch.)*

G. Baloch is with the Department of Electrical and Electronic Engineering, Eastern Mediterranean University, Gazimagusa 99628, Turkey, on leave from the Sukkur Institute of Business Administration, Sukkur 65200, Pakistan (e-mail: gulsher.baloch@cc.emu.edu.tr).

H. Ozkaramanli and R. Yu are with the Department of Electrical and Electronic Engineering, Eastern Mediterranean University, Gazimagusa 99628, Turkey (e-mail: huseyin.ozkaramanli@emu.edu.tr; yu@ieee.org).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2789018

represented by its code coefficients $\boldsymbol{\alpha} \in \mathbb{R}^K$ as $\hat{\mathbf{x}}^c = \mathbf{D}\boldsymbol{\alpha}$, then the residue is

$$\mathbf{r} = \mathbf{x} - \hat{\mathbf{x}}^c = (\mathbf{x}^c - \hat{\mathbf{x}}^c) + \mathbf{w}. \tag{2}$$

This implies that if the code is correctly determined so that $\mathbf{x}^c \approx \hat{\mathbf{x}}^c$, then $\mathbf{r} \approx \mathbf{w}$. As a result, the residue should possess the same statistical properties of the contaminating noise.

In maximum projection based algorithms, approximation of a noisy patch $\mathbf{x}$ is achieved by projecting it onto dictionary atoms and picking the atom that gives the maximum projection. Note that the performance of maximum projection based algorithms deteriorates as the noise level increases [12]. According to [2], the projection coefficient on atom $\mathbf{d}_k$ is

$$\mathbf{d}_k^T \mathbf{x} \approx ||\mathbf{x}^c|| \cos(\theta_{\mathbf{d}_k, \mathbf{x}^c}) + ||\mathbf{w}|| \cos(\theta_{\mathbf{d}_k, \mathbf{w}})$$

where $\theta_{\mathbf{a},\mathbf{b}}$ denotes the angle between vectors $\mathbf{a}$ and $\mathbf{b}$. At high noise levels where $\|\mathbf{w}\| >> ||\mathbf{x}^c||$, the noise $\mathbf{w}$ dominates the maximum projection and thus would dictate the atom selection process. And the atom that matches the contaminating noise is then picked. Consequently, the residual $\mathbf{r}$ contains remnants from clean signal and it would not possess properties of the noise.

In this letter, we develop a correlation-based regularization to ensure that the residuals of different patches are minimally correlated, hence behave like contaminating noise. Our problem to be solved in the sequel can be stated as follows. Given a patch $\mathbf{x}$ from a noisy image, we find a sparse code $\boldsymbol{\alpha}$ and dictionary $\mathbf{D}$, such that the representation $\hat{\mathbf{x}}^c = \mathbf{D}\boldsymbol{\alpha}$ gives a good approximation of the clean image $\mathbf{x}^c$ and the resultant residue $\mathbf{r} = \mathbf{x} - \hat{\mathbf{x}}^c$ is uncorrelated to the residues of neighboring patches of the noisy image.

III. Residual Correlation Regularization

A. Sparse Coding

Denote the current patch that is being processed by $\mathbf{x}$. The corresponding residue is initialized to be $\mathbf{r}_0 = \mathbf{x}$. Assume that M of its immediate neighbors have been processed, and the corresponding residuals are $\mathbf{r}^m$, $m = 1, 2, \ldots, M$.

Let us pick the sth atom for $\mathbf{x}$. Denote by $\mathbf{r}_{s-1}$ the residual formed after selection of $s-1$ atoms. If the atom picked is $\mathbf{d}_{k_s}$, and the corresponding coefficient is α_s, then the new residual is

$$\mathbf{r}_s = \mathbf{r}_{s-1} - \mathbf{d}_{k_s} \alpha_s. \tag{3}$$

We propose to select $\mathbf{d}_{k_s}$ and perform sparse coding by solving the following optimization problem:

$$(k_s^*, \alpha_s^*) = \arg \min_{k_s, \alpha_s} J_c(k_s, \alpha_s) \tag{4}$$

where the objective function is

$$J_c(k_s, \alpha_s) = \frac{1}{2} \|\mathbf{r}_s\|_2^2 + \sum_{m=1}^{M} \lambda_m \left| \mathbf{r}_s^T \mathbf{r}^m \right| \tag{5}$$

with $\lambda_m > 0, m = 1, 2, \ldots, M$, being regularization weighting parameters, which can be selected according to the level of the noise.

Note that the first term of J_c in (5) represents a fidelity term, whereas the second term realizes the regularization on residual correlation between the current residual $\mathbf{r}_s$ and neighboring residuals $\mathbf{r}^m$. Minimizing J_c enforces the residual of the current patch as uncorrelated with those of neighboring patches as possible. As a result, statistical properties of the residual would quite likely resemble those of the contaminating noise.

Expanding $J_c(k_s, \alpha_s)$ in terms of k_s and α_s, we have

$$\begin{aligned} &J_c(k_s, \alpha_s) \\ &= \frac{1}{2} (\mathbf{r}_{s-1} - \mathbf{d}_{k_s} \alpha_s)^T (\mathbf{r}_{s-1} - \mathbf{d}_{k_s} \alpha_s) \\ &\quad + \sum_{m=1}^{M} \lambda_m \left| (\mathbf{r}_{s-1} - \mathbf{d}_{k_s} \alpha_s)^T \mathbf{r}^m \right| \\ &= \frac{1}{2} \Big(\alpha_s^2 \mathbf{d}_{k_s}^T \mathbf{d}_{k_s} - 2\alpha_s (\mathbf{d}_{k_s})^T \mathbf{r}_{s-1} + \mathbf{r}_{s-1}^T \mathbf{r}_{s-1} \Big) \\ &\quad + \sum_{m=1}^{M} \lambda_m \left| (\mathbf{r}_{s-1} - \mathbf{d}_{k_s} \alpha_s)^T \mathbf{r}^m \right|. \end{aligned}$$

Taking derivative of $J_c(k_s, \alpha_s)$ with respect to (w.r.t.) α_s gives

$$\frac{\partial f}{\partial \alpha_s} = -\mathbf{d}_{k_s}^T (\mathbf{r}_{s-1} - \mathbf{d}_{k_s} \alpha_s) - \sum_{m=1}^{M} s_m \lambda_m \mathbf{d}_{k_s}^T \mathbf{r}^m$$

where

$$s_m = \mathrm{sgn}\big((\mathbf{r}_{s-1} - \alpha_s \mathbf{d}_{k_s})^T \mathbf{r}^m \big) \tag{6}$$

with sgn being the sign function. Letting the derivative be equal to zero and noting that $(\mathbf{d}_{k_s}^T) \mathbf{d}_{k_s} = 1$, we obtain the sparse coefficient

$$\alpha_s = \mathbf{d}_{k_s}^T \mathbf{r}_{s-1} + \sum_{m=1}^{M} s_m \lambda_m \mathbf{d}_{k_s}^T \mathbf{r}^m. \tag{7}$$

Note that if all $\lambda_m = 0$, function J_c in (5) becomes the standard least square error (LSE) expression. And when no patch correlation regularization is used, the solution in (7) coincides with the solution of the corresponding LSE problem [4]. Furthermore, (7) shows explicitly how the LSE solution is modified when any neighboring patch is used for regularization.

There are two issues in calculating α_s from (7): 1) α_s is dependent on atom $\mathbf{d}_{k_s}$, which is yet to be determined; and more importantly 2) α_s itself also appears in the signs s_m of (6). Fortunately, issue 1) can be resolved by examining all the atoms in the dictionary that have not been selected for patch $\mathbf{x}$. And for each atom considered, issue 2) can also be circumvented by testing all 3^M combinations of s_m and selecting the combination that yields the least $J_c(k_s, \alpha_s)$. As a results, (4) is then solved. That is, we complete the sth atom selection and its coefficient calculation.

The above-mentioned atom selection process is repeated until the maximum number of atoms to be used is reached or the residual power is reduced below the noise power.

B. Dictionary Updating

The sparse coding stage described in Section III-A assumes that the dictionary $\mathbf{D}$ is known and fixed. If we wish to learn the dictionary from noisy image patches, we can resort to the two-stage dictionary learning algorithm employed in [1], [5], [13], and [6]. Specifically, we alternate between the sparse coding stage and a new dictionary update stage.

We now consider the dictionary update stage. Consider the tth iteration, with dictionary $\mathbf{D}_{t-1}$ available. Let the current patch be $\mathbf{x}_t$ with sparse code $\boldsymbol{\alpha}_t$, then the corresponding residue is $\mathbf{r}_t = \mathbf{x}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}_t$.[1] Similarly, let $\mathbf{r}_i$ be the residuals of the neighboring patch that have been processed under $\mathbf{D}_{t-1}$ with the corresponding code $\boldsymbol{\alpha}_i$, $i = 1, \ldots, t-1$.

The new dictionary $\mathbf{D}$ can be updated by solving the following optimization problem:

$$D_t = \arg\min_{\mathbf{D}} J_d(\mathbf{D}) \tag{8}$$

where

$$\begin{aligned} J_d(D) &= \frac{1}{t}\left[\sum_{i=1}^{t} \|\mathbf{r}_i\|_2^2 + \sum_{i=1}^{t-1} \lambda_i \left|\mathbf{r}_t^T \mathbf{r}_i\right|\right] \\ &= \frac{1}{t}\left[\sum_{i=1}^{t} \|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i\|_2^2 + \sum_{i=1}^{t-1} \lambda_i \left|(\mathbf{x}_t - \mathbf{D}\boldsymbol{\alpha}_t)^T(\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i)\right|\right]. \end{aligned} \tag{9}$$

Function J_d above explicitly shows that the neighboring residual patches contribute to the dictionary update. Note that if no neighborhood patch $\mathbf{x}_i$ is used in (9), then the proposed dictionary update reduces to the method of optimal directions for frame design [14].

In terms of matrix trace, J_d can be expressed as

$$\begin{aligned} J_d(D) = \frac{1}{t}\Big[&\mathrm{Tr}(\mathbf{D}^T\mathbf{D}\mathbf{A}_t - \mathbf{D}^T\mathbf{B}_t + \mathbf{x}_t\mathbf{x}_t^T) \\ &+ \mathrm{Tr}(\mathbf{D}^T\mathbf{D}\mathbf{f}_{t-1}\boldsymbol{\alpha}_{t-1}^T - \mathbf{D}^T\mathbf{g}_{t-1}\boldsymbol{\alpha}_{t-1}^T - \mathbf{D}\mathbf{f}_{t-1}\mathbf{x}_t^T + \mathbf{g}_{t-1}\mathbf{x}_t^T)\Big] \end{aligned} \tag{10}$$

where

$$\mathbf{A}_t = \sum_{i=1}^{t} \boldsymbol{\alpha}_i\boldsymbol{\alpha}_i^T = \mathbf{A}_{t-1} + \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t^T \tag{11}$$

$$\mathbf{B}_t = \sum_{i=1}^{t} \mathbf{x}_i\boldsymbol{\alpha}_i^T = \mathbf{A}_{t-1} + \mathbf{x}_t\boldsymbol{\alpha}_t^T \tag{12}$$

$$\mathbf{f}_t = \sum_{i=1}^{t-1} \lambda_i s_i \boldsymbol{\alpha}_i = \mathbf{f}_{t-1} + \lambda_{t-1} s_{t-1}\boldsymbol{\alpha}_{t-1} \tag{13}$$

$$\mathbf{g}_t = \sum_{i=1}^{t-1} \lambda_i s_i \mathbf{x}_i = \mathbf{g}_{t-1} + \lambda_{t-1} s_{t-1}\mathbf{x}_{t-1} \tag{14}$$

with

$$s_i = \mathrm{sgn}\big((\mathbf{x}_t - \mathbf{D}\boldsymbol{\alpha}_t)^T(\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i)\big). \tag{15}$$

Taking the derivative of the objective function in (10) w.r.t. D and setting it to be zero, we have

$$\mathbf{D}\mathbf{A}_c - \mathbf{B}_c = 0 \tag{16}$$

where

$$\mathbf{A}_c = \mathbf{A}_t + \mathbf{f}_{t-1}\boldsymbol{\alpha}_{t-1}^T, \tag{17}$$

$$\mathbf{B}_c = \mathbf{B}_t + \mathbf{g}_{t-1}\boldsymbol{\alpha}_{t-1}^T + \mathbf{x}_t\mathbf{f}_{t-1}^T \tag{18}$$

which can be approximately solved by

$$\mathbf{D}_t = \mathbf{D}_{t-1} + (\mathbf{B}_c - \mathbf{D}_{t-1}\mathbf{A}_c)\mathbf{A}_c^{-1}. \tag{19}$$

Since the coefficient vectors $\boldsymbol{\alpha}$ are sparse, when λ_i are selected to be small, the coefficients of $\mathbf{A}_c$ are generally diagonal. As a result, the kth column of $\mathbf{D}$ can be approximately updated as

$$\mathbf{D} = \mathbf{D}_{t-1} + (\mathbf{B}_c - \mathbf{D}_{t-1}\mathbf{A}_c)[\mathrm{diag}(\mathbf{A}_c)]^{-1}. \tag{20}$$

And the kth column of dictionary $\mathbf{D}_t$ can be calculated by an approximation followed by a normalization:

$$\begin{aligned} \mathbf{u}_k &\leftarrow \frac{1}{\mathbf{A}_c(k,k)}\big(\mathbf{B}_c(:,k) - \mathbf{D}_{t-1}\mathbf{A}_c(:,k)\big) + \mathbf{D}_{t-1}(:,k) \\ \mathbf{D}_t(:,k) &\leftarrow \frac{1}{\max(\|\mathbf{u}_k\|_2, 1)}\mathbf{u}_k. \end{aligned} \tag{21}$$

Note that $\mathbf{A}_c$ and $\mathbf{B}_c$ are related to the signs s_i, $i = 1, \ldots, t-1$, which are dependent on $\mathbf{D}$. To circumvent this dependence, we can in principle consider all combinations of the possible signs. However, it would quickly become infeasible as t increases. Nevertheless, we can practically overcome this problem in two ways: 1) We include only a small number of correlation terms in the objective function (9) (e.g., let $M = 2$); and 2) we use $\mathbf{D}_{t-1}$ instead of $\mathbf{D}$ in determining the sign, that is,

$$s_i = \mathrm{sgn}\big(\mathbf{x}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}_t)^T(\mathbf{x}_i - \mathbf{D}_{t-1}\boldsymbol{\alpha}_i)\big), \quad i = 1, \ldots, t-1. \tag{22}$$

C. Computational Complexity

In this section, we discuss the computational complexity of the proposed algorithm.

Recall that the K-SVD performs $\mathcal{O}(nKLP)$ operations per pixel [4], where n, K, L, P are the size of the patch, the size of the dictionary atoms, the maximum number of atoms selected, and the number of iterations, respectively. Similar approach is adapted in the proposed algorithm, except that for each patch being processed, we need to test for different combinations of signs in order to select the atom and to update the dictionary. As a result, the computational cost of the proposed algorithm becomes $\mathcal{O}(cnK^2LP)$, where c is the number of sign combinations considered at each selection and updating. Therefore, the proposed algorithm is computationally more expensive than the K-SVD [4].

On the other hand, we shall show in Section IV that the correlation-based regularization can help produce much better results than the K-SVD, both qualitatively and quantitatively. Note that the computational burden can be alleviated by searching for suboptimal solutions of (4) and (8), for example, by selecting atoms that reduce the objective function below a certain threshold value. This shall significantly reduce the computational complexity at the cost of minor loss in performance.

IV. Experiment Results and Comparison

In this section, we compare the proposed algorithm with state-of-the-art algorithms of image denoising. The performance is evaluated in terms of the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and feature similarity index (FSIM). A visual comparison is also presented.

[1]Though there is a little abuse of notation, $\mathbf{r}_t$ is not to be confused with $\mathbf{r}_s$ in Section III-A.

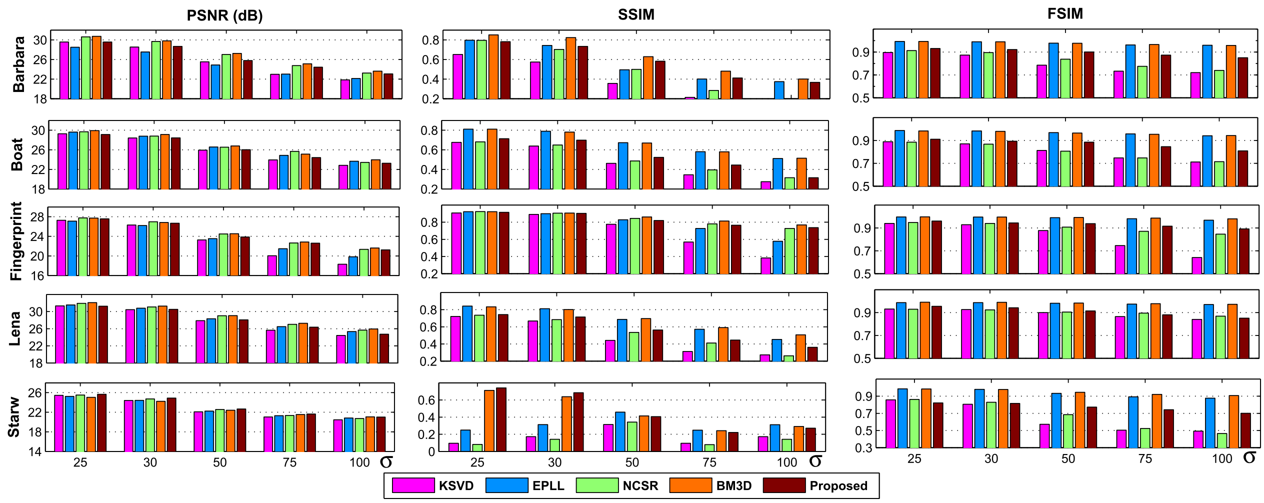

Fig. 1. PSNR, SSIM, and FSIM comparison.

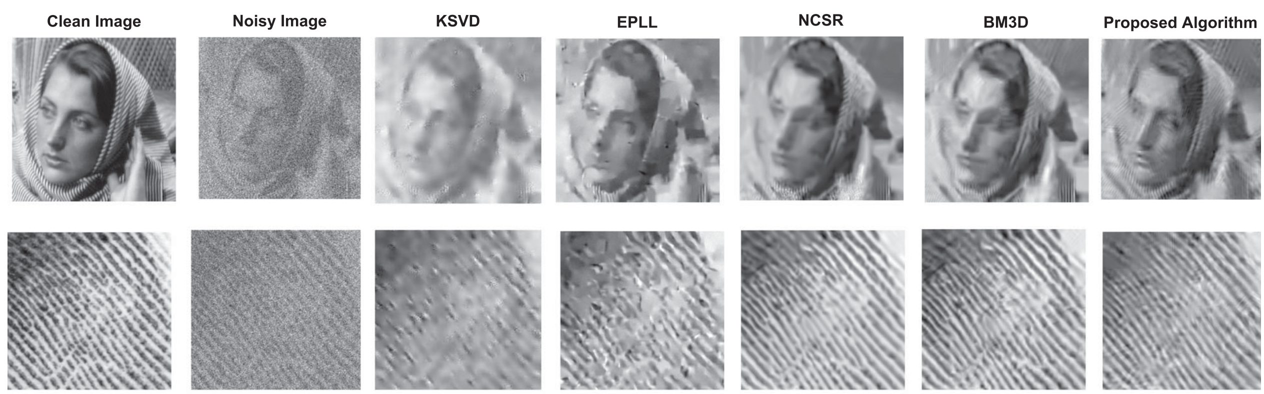

Fig. 2. Visual comparison: Barbara ($\sigma = 75$) and fingerprint ($\sigma = 100$).

We select some standard test images, all of size 512×512. Then image segments of size 80×80 are extracted from each image. These segments are further divided into 15×15 fully overlapping patches. The additive white Gaussian noise of power between 25 and 100 is generated. The number of immediate neighborhood patches is set to be $M = 2$. Sparse coefficient and dictionary update stages are iterated ten times. The regularization parameter λ_m varies from 0.5 to 1 with increasing noise levels. Note that we process (denoise) 80×80 segments of the image separately. Thus, we initialize dictionary of size 225×35 with prechosen and fixed atoms as given in [4]. The dictionary is updated after each iteration using (21). To reduce the computational cost, (22) is used to determine the signs.

We see from Fig. 1 that the proposed algorithm significantly outperforms the K-SVD [4] and the expected patch log likelihood (EPLL) [10] image denoising algorithms in terms of PSNR. In contrast to K-SVD [12], the performance of the proposed algorithm improves with increasing noise levels. However, the block-matching and 3D filtering (BM3D) [7] and the nonlocally centralized sparse representation (NCSR) algorithms [11] produces high PSNR results (except for Starw image). Also, the proposed algorithm performs better as compared to the NCSR and K-SVD algorithms in terms of SSIM and FSIM, as shown in Fig. 1.

For visual comparison, we show in Fig. 2 (first row) a portion of the Barbara image reconstructed for $\sigma = 75$. It is clear that the stripes on the scarf near the hand are well-reconstructed by the proposed algorithm. All the other algorithms except the BM3D [7] do a poor job in recovering such fine structures. Similarly, Fig. 2 (second row) shows that the K-SVD [4] and EPLL [10] algorithms are unable to recover ridges in a fingerprint image at high noise level of $\sigma = 100$, whereas the proposed algorithm does an excellent job to recover these structures of ridges.

We conclude that the BM3D and NCSR algorithms produce better PSNR results; however, for images with repeating structures and at high noise levels, the proposed residual correlation regularization based algorithm performs relatively better. Also, according to the visual results obtained, it is as good if not better than the BM3D and NCSR algorithms.

V. Conclusion

We presented a new residual correlation regularization that helps to make statistical properties of residual to be similar to those of the contaminating noise. This is achieved by a new method of sparse coefficient estimation and dictionary update stage based on the proposed regularization. Our experiment results demonstrate that the proposed algorithm is a good complement to state-of-the-art image denoising algorithms. Improvement toward reducing computational load of the proposed algorithm needs to be further investigated.

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[2] G. Baloch and H. Ozkaramanli, "Image denoising via correlation-based sparse representation," *Signal, Image Video Process.*, vol. 11, pp. 1501–1508, Nov. 2017.

[3] H. Yue, X. Sun, J. Yang, and F. Wu, "Image denoising by exploring external and internal correlations," *IEEE Trans. Signal Process.*, vol. 24, no. 6, pp. 1967–1982, Jun. 2015.

[4] 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.

[5] Y. He, T. Gan, W. Chen, and H. Wang, "Multi-stage image denoising based on correlation coefficient matching and sparse dictionary pruning," *Signal Process.*, vol. 92, pp. 139–149, 2012.

[6] E. M. Ender and O. Bayir, "K-SVD meets transform learning: Transform K-SVD," *IEEE Signal Process. Lett.*, vol. 21, no. 3, pp. 347–351, Mar. 2014.

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

[8] P. Riot, A. Almansa, Y. Gousseau, and F. Tupin, "Penalizing local correlations in the residual improves image denoising performance," in *Proc. 2016 24th Eur. Signal Process. Conf.*, 2016, pp. 1867–1871.

[9] D. Brunet, E. R. Vrscay, and Z. Wang, "The use of residuals in image denoising," in *Proc. Int. Conf. Image Anal. Recog.*, 2009, pp. 1–12.

[10] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 479–486.

[11] W. Dong, L. Zhang, Lei, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[12] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, vol. 1, 2006, pp. 895–900.

[13] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 689–696.

[14] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1999, pp. 2443–2446.

[15] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[16] S. Yang, W. Min, L. Zhao, and Z. Wang, "Image noise reduction via geometric multiscale ridgelet support vector transform and dictionary learning," *IEEE Trans. Image Process.*, vol. 22, no. 11, pp. 4161–4169, Nov. 2013.