



Image Denoising Using Collaborative Patch-Based and Local Methods

Vittoria Bruni^{1,2}(✉) and Domenico Vitulano²

¹ Department of SBAl, Sapienza - Rome University, Rome, Italy
vittoria.bruni@sbai.uniroma1.it

² Institute for the Applications of the Calculus (IAC) - C.N.R., Rome, Italy
d.vitulano@iac.cnr.it

Abstract. In this paper local and non-local denoising methods are jointly employed in order to improve the visual quality of the final denoised image. Based on the evidence that the output images of non local denoising methods are not pointwise better everywhere than the outputs images of local methods and than the noisy image itself, the cascade of two improvement steps is applied to the output image of a non local denoising method. The first step aims at correcting the output image by recovering the lost information directly from the noisy one. The second step aims at recovering those good estimations provided by a local regularization method. A pointwise weighted average between the involved image pair is used at each step. The weights are estimated from the noisy data using adaptive and automatic procedures. Experimental results show that the proposed approach allows us to greatly improve the results of patch based non local denoising in terms of both peak signal to noise ratio (PSNR) and structural similarity index (SSIM).

Keywords: Image denoising · Collaborative filtering
Residual method · BM3D · Wiener filtering

1 Introduction

Denoising is a longstanding and investigated topic. Despite the huge number of approaches proposed for its solution, it still remains a challenging and open research problem. Linear and non linear filtering, bayesian methods, shrinkage methods, variational models, local and non local filtering are only few examples of the different approaches and ideas underlying the existing solutions — see [10] for a complete and recent review. The most recent and performing denoisers are based on non local patch regularization. This kind of regularization allows us to reach high standard results, which are considerably better than those achieved by traditional local denoisers. Noteworthy examples are Non Local Means (NLM) [3] and Block Matching 3D (BM3D) [5], especially for the removal of additive white Gaussian noise. They mainly consist of grouping similar patches in the image and then removing noise by exploiting replicates of the same patch. NLM

is based on a weighted mean of similar patches in the whole image; BM3D is based on the concept of collaborative filtering. It mainly consists of a proper and joint use of non local similarities search (through the block matching algorithm) and standard transform-based denoising applied to 3D data, i.e. the stack of similar blocks.

Despite the very high performance of such approaches, patch-based non local denoising is not perfect. In fact, the average of non local but similar patches leads to an unavoidable smoothing of different parts of the noisy image under study, especially in case of patches misalignments due to the presence of noise or due to the lack of point-wise correspondence between patches. In addition, in order to reduce the prohibitive computational effort required by the search of similarities in the whole image, this search is limited in a neighborhood of each pixel; as a result not all image similarities contribute to the denoising process. A lot of research effort has been devoted to make non local methods fast and usable in real applications; in particular, methods for making faster the search of similarities have been proposed as well as for improving the visual quality of denoised images, see for example [2, 4, 6, 7, 9, 11, 12]. With regard to the last point, a classical way for improving the denoising result is the residual method [1]. It mainly consists of applying an edge preserving denoising filtering to the residual image, i.e. the one which is obtaining by subtracting the denoised image from the noisy one.

The aim of this paper is to show that there are several image pixels where local, even traditional, denoising performs better than non local one. In particular, it is possible to show that there is a class of pixels where classical Wiener filter performs better than the most performing non local denoiser, i.e. BM3D. The interesting aspect is that pixels belonging to this class are not necessarily located in correspondence to image edges. In fact, the performance of non local denoising ‘locally’ depends on how many and how similar patches are within the image. For a similar reason, somewhat more anti intuitive, it is possible to show that there is also another class of pixels for which it is better to leave the noisy image unchanged rather than performing a patch based denoising. Since it is very hard to predict which are the pixels belonging to the two aforementioned classes, in this paper a strategy based on two different convex combinations of local and non local information is proposed. The main aim is to embed local information into a non local denoising result. The first convex combination involves the output image of BM3D and the noisy image itself. In cascade, the second convex combination involves the output of the first step and the output image of a simple Wiener filter when applied to the noisy image. The parameters employed in the convex combinations are properly estimated from the data using two different approaches and they involve all image pixels, not only the ones belonging to each specific class.

Experimental results achieved on various test images show that the proposed approach is able to increase the visual quality of BM3D image of 1 db on average at low levels of noise and of about .5 db at higher noise levels.

2 The Proposed Model

In the denoising problem, the noisy image Y is modelled as follows

$$Y(\mathbf{i}) = X(\mathbf{i}) + N(\mathbf{i}), \quad \mathbf{i} \in \Omega \quad (1)$$

where X is the original (clean) image, N is an i.i.d. Gaussian white noise with zero mean and variance σ^2 and Ω is the image domain. \mathbf{i} stands for pixel location. Denoising goal is to get an estimate \hat{X} of X from Y . The recent literature has shown that non local patch-based denoisers, like BM3D (whose output image will be denoted by \hat{X}) are more effective in terms of SNR than traditional local denoisers, like Wiener filtering (whose output will be denoted by \bar{X}). However, this global result is not pointwise satisfied: there are some pixels for which \bar{X} is better than \hat{X} and others where Y is a better candidate than \hat{X} . As a result, we can define two distinct classes of pixels as follows

$$\Omega_1 = \{\mathbf{i} \in \Omega : |Y(\mathbf{i}) - X(\mathbf{i})| < |\hat{X}(\mathbf{i}) - X(\mathbf{i})|\} \quad (2)$$

$$\Omega_2 = \{\mathbf{i} \in \Omega : |\bar{X}(\mathbf{i}) - X(\mathbf{i})| < |\hat{X}(\mathbf{i}) - X(\mathbf{i})|\}. \quad (3)$$

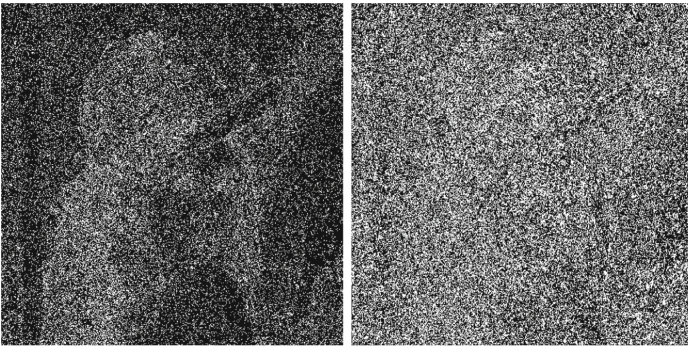


Fig. 1. $512 \times 512 \times 8$ bits Lena image corrupted by a Gaussian noise $N(0, 20)$. (Left) Map of pixels belonging to Ω_1 (percentage of white points 19%). (Right) Map of pixels belonging to Ω_2 (percentage of white points 38%).

An example of Ω_1 and Ω_2 is shown in Fig. 1. More in general, this means that some image high frequencies have been lost in the patch based non local method due to the regularization process; on the other hand, some low frequencies have been lost since some artifacts are introduced due to misalignments or not good similarities matching. That is why it is necessary to reintroduce in the image both some high and low frequency content. To this aim, we should construct the following sequence of images

$$X_n = X_{n-1} + \alpha_n(Y_n - X_{n-1}), \quad n \geq 1 \quad (4)$$

with $X_0 = \hat{X}$. α_n is a parameter to be estimated; Y_n is the sequence of images having decreasing high frequency content from which recovering the lost information in X_0 . $Y_1 = Y$ since it contains all image high frequencies, while Y_n , $n > 1$, is a more and more regularized version of the noisy image as n increases. α_n quantifies the amount of local frequency content that has to be recovered at each step. Eq. (4) can be rewritten as follows

$$X_n = \alpha_n Y_n + (1 - \alpha_n) X_{n-1}, \quad n \geq 1. \quad (5)$$

As a result, for $\alpha_n \in [0, 1]$, X_n is a convex combination of two regularized versions of the noisy image, i.e. two distinct estimations of the original one X . Based on these considerations, since it is not trivial to find out Ω_1 and Ω_2 in Eqs. (2) and (3), in this paper we propose to use just two iterations of the sequence in Eq. (5). The first one involves the noisy image Y and the output image \hat{X} of a patch based non local denoiser (BM3D), i.e.

$$X_1 = \alpha_1 Y + (1 - \alpha_1) \hat{X}. \quad (6)$$

The second one involves X_1 and the output image of a local denoiser (Wiener filter) \bar{X} , i.e.

$$X_2 = \alpha_2 \bar{X} + (1 - \alpha_2) X_1. \quad (7)$$

α_1 and α_2 can be defined as the solution of a joint least squares minimization problem, i.e. $(\alpha_1, \alpha_2) = \operatorname{argmin}_{\alpha_1, \alpha_2 \in [0, 1]} \|X - X_2\|_2^2$, or as the solutions of two separated minimization processes independently applied at each step of the proposed procedure, i.e. $\alpha_1 = \operatorname{argmin}_{\alpha_1 \in [0, 1]} \|X - X_1\|_2^2$, and $\alpha_2 = \operatorname{argmin}_{\alpha_2 \in [0, 1]} \|X - X_2\|_2^2$.

Unfortunately, as shown in the Appendix, these estimates depend on the original image through terms that cannot be considered negligible. As a matter of facts, in the estimation of α_1 , the independence of noise from the original image allows us to get the following reliable estimate for α_1 :

$$\alpha_1 \simeq 1 - \frac{\sigma^2 |\Omega|}{\|Y - \hat{X}\|_2^2}. \quad (8)$$

On the contrary, it is not the case for α_2 , where the term $\frac{\sum_{\Omega} N(X - X_1)}{\|Y - \hat{X}\|_2^2}$ is not negligible and its robust estimation is not trivial. In the following an estimation method based on the Minimum Description Length (MDL) is proposed for α_2 .

2.1 MDL for α_2 Estimation

MDL is based on the rationale that during a compression phase, the distortion between the compressed signal and the original one should be kept low; at the same time, the description of the compressed signal should use as few bits as possible. Since these two objectives are conflicting, a suitable criterion for reaching a compromise is needed, as for example Rissanen's MDL principle [8]. Let fix a class of models from which selecting the one which best represents the

data. According to the MDL principle, given a sequence of observations, the *best* model is the one that yields the shortest description length for describing the data using the model — the description length can be interpreted as the number of bits needed for encoding it. This description can be done through a two-part code: the first one describes the model, the other describes the data using the model. In our case, X_1 represents the data (observations) and we wish to find a model $M(\alpha_2)$ that describes it according to Eq. (7). $M(\alpha_2)$ minimizes the following two-part code-length:

$$\alpha_2 = \underset{\alpha_2 \in [0,1]}{\operatorname{argmin}} L(X_1, M(\alpha_2)) \quad (9)$$

where $L(X_1, M(\alpha_2)) = L(X_1|M(\alpha_2)) + \lambda L(M(\alpha_2))$, λ is a parameter that should balance the two terms. Using PSNR as error measure, $L(X_1|M(\alpha_2)) = 10 \log_{10} \frac{255^2}{\sum \sum (X_1 - X_{2,\alpha_2})^2}$, while X_{2,α_2} is Eq. (7) for a given α_2 . $L(M(\alpha_2))$ gives the bits necessary to encode X_2 , i.e. the bits necessary to code X_1 , $\alpha_2 \bar{X}$ and the parameter α_2 . Note that the bits budget required by α_2 is constant and can be neglected in the minimization. In order to make automatic the minimization process (choice of λ), PSNR is given with a 2 decimal digits precision; B is the bits budget for X_1 while just $\alpha_2 B$ bits are required for $\alpha_2 \bar{X}$. Hence, also for $L(M(\alpha_2))$ two decimal digits have been considered. The parameter λ is then set as the ratio between the maximum of $L(X_1|M(\alpha_2))$ and the maximum for $L(M(\alpha_2))$ with respect to $\alpha_2 \in [0, 1]$.

3 Experimental Results and Concluding Remarks

The proposed model, namely WN-BM3D, has been tested on several images corrupted by zero mean Gaussian noise with standard deviation ranging from 5 to 100. In this paper we will show the results on two $512 \times 512 \times 8$ bits test images: Lena and Fingerprint. Results have been evaluated in terms of Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity index (SSIM). Table 1 refers to Lena image and shows the value of the parameters α_1 and α_2 when two distinct oracle minimization procedures are performed. Oracle means that the original image is known and each minimization reaches the minimum MSE. PSNR after the first process (Eq. 6) and the second one (Eq. 7) as well as PSNR increase with respect to BM3D are shown. As it can be observed, PSNR increase is up 1.38 db and it grows as noise standard deviation decreases. Table 1 also gives α_1 and α_2 which have been estimated respectively using Eqs. (8) and (9). As it can be observed, the proposed method allows us to reach PSNR values which are very close to the oracle ones, confirming the robustness of the proposed minimization procedures. It is worth observing that as noise standard deviation increases, the second step is the one which contributes more to the refinement of BM3D image. Table 2 compares WN-BM3D and BM3D in terms of PSNR and SSIM for both Lena and Fingerprints images. WN-BM3D outperforms BM3D for both images and for all σ s. In particular, on more difficult images (where edges density is higher) like Fingerprint, the increase in terms of PSNR provided

by WN-BM3D is even higher. This may be explained considering that on such images patch based non local denoiser is not able to find 'faithful' patches for a given point, as the image is very irregular and sharp. In this case, an injection of local regularization can help to outperform the final result. Figure 2 allows us to evaluate the visual quality of the final images. As it can be observed, the proposed method provides images that show a reduced over-smoothing in flat regions. Finally, it is worth outlining that each step of WN-BM3D has been optimized with respect to PSNR. That is why SSIM may be worst than the BM3D for higher levels of noise. However, as Table 2 shows, by selecting Wiener filter iterated twice as local denoiser, the proposed method, namely 2WN-BM3D, is able to provide better results even with respect to SSIM for higher level of noise. The aim of future research will be twofold. On the one hand, the properties of the local methods to use in the proposed iterative residual procedure will be studied; on the other hand, starting from the empirical observation that the pointwise solution of a non local method can be successfully substituted for the one provided by a traditional local method or the noisy datum itself, methods for the selection of the best pointwise solution will be investigated. The latter goal will enable the proposed method to not only improve denoising results but to also provide a considerable computational saving with respect to non local methods.

4 Appendix

Joint Minimization. By putting Eq. (6) into Eq. (7) we get $\|(X_2 - X)\|_2^2 = \|(\hat{X} + \alpha_1(1 - \alpha_2)(Y - \hat{X}) + \alpha_2(\bar{X} - \hat{X}) - X)\|_2^2$. By imposing $\frac{\partial \|(X_2 - X)\|_2^2}{\partial \alpha_1} = 0$ and $\frac{\partial \|(X_2 - X)\|_2^2}{\partial \alpha_2} = 0$ and denoting by $\langle *, * \rangle$ the scalar product, using some algebra we get

$$\alpha_1 = -\frac{\langle \hat{X} - X + \alpha_2(\bar{X} - \hat{X}), \bar{X} - \hat{X} \rangle}{(1 - \alpha_2)\langle Y - \hat{X}, \bar{X} - \hat{X} \rangle} \quad \text{and} \quad \alpha_2 = \frac{\frac{\langle \bar{X} - X, Y - \hat{X} \rangle}{\|Y - \hat{X}\|_2^2} - \frac{\langle \bar{X} - X, \bar{X} - \hat{X} \rangle}{\langle Y - \hat{X}, \bar{X} - \hat{X} \rangle}}{\frac{\|\bar{X} - \hat{X}\|_2^2}{\langle Y - \hat{X}, \bar{X} - \hat{X} \rangle} - \frac{\langle \bar{X} - \hat{X}, Y - \hat{X} \rangle}{\|Y - \hat{X}\|_2^2}}$$

α_1 estimation. By using Eq. (1) and setting $\hat{X} = X + N_0$, where $N_0 = \hat{X} - X$ is BM3D approximation error, we have $\frac{d\|(X - X_1)\|_2^2}{d\alpha_1} = 0 \Leftrightarrow \alpha_1 = \frac{\langle X - \hat{X}, Y - \hat{X} \rangle}{\|Y - \hat{X}\|_2^2} = \frac{\langle Y - \hat{X} - N, Y - \hat{X} \rangle}{\|Y - \hat{X}\|_2^2} = 1 - \frac{\|N\|_2^2}{\|Y - \hat{X}\|_2^2} + \frac{\langle N, N_0 \rangle}{\|Y - \hat{X}\|_2^2} \simeq 1 - \frac{\sigma^2|\Omega|}{\|Y - \hat{X}\|_2^2}$, where the term $\frac{\sum_{\Omega} N N_0}{\|Y - \hat{X}\|_2^2}$ has been neglected since it is close to zero, as N is independent of X and nearly globally independent of \hat{X} .

α_2 estimation. $\frac{d\|(X - X_2)\|_2^2}{d\alpha_2} = 0 \Leftrightarrow \alpha_2 = \frac{\langle X - X_1, \bar{X} - X_1 \rangle}{\|\bar{X} - X_1\|_2^2} = \frac{\langle Y - X_1, \bar{X} - X_1 \rangle}{\|\bar{X} - X_1\|_2^2} - \frac{\langle N, \bar{X} - X_1 \rangle}{\|\bar{X} - X_1\|_2^2}$. Unfortunately, the last term is neither known nor negligible.

Table 1. Lena image. (*Top*) Oracle values for α_1 and α_2 in Eqs. (6) and (7). From left to right: noise standard deviation σ and PSNR of the noisy image; α_1 value; α_2 value; PSNR of the output image X_1 of the first step; PSNR of the output image X_2 of the second step; PSNR increment (ΔP) of WN-BM3D method with respect to BM3D. (*Bottom*) Same results with α_1 and α_2 estimated using Eqs. (8) and (9). Values are given with two decimal digits.

σ	5	10	20	30	40	50	60	70	80	90	100
PSNR	34.12	28.11	22.12	18.69	16.34	14.61	13.27	12.22	11.38	10.70	10.14
α_1	0.26	0.14	0.08	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02
α_2	0.10	0.18	0.14	0.12	0.10	0.08	0.07	0.06	0.07	0.07	0.07
PSNR 1st step	40.00	36.58	33.38	31.49	30.04	29.19	28.38	27.67	27.06	26.53	26.03
PSNR 2st step	40.07	36.82	33.60	31.70	30.24	29.35	28.52	27.78	27.22	26.68	26.18
ΔP	1.39	0.91	0.58	0.44	0.38	0.30	0.25	0.21	0.25	0.23	0.22
σ	5	10	20	30	40	50	60	70	80	90	100
PSNR	34.12	28.11	22.12	18.69	16.34	14.61	13.27	12.22	11.38	10.70	10.14
α_1	0.26	0.15	0.07	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
α_2	0.11	0.12	0.12	0.13	0.14	0.14	0.14	0.15	0.15	0.15	0.15
PSNR 1st step	40.00	36.58	33.38	31.44	29.86	29.05	28.27	27.57	26.97	26.45	25.95
PSNR 2st step	40.07	36.78	33.62	31.78	30.34	29.42	28.59	27.83	27.19	26.65	26.15
ΔP	1.38	0.88	0.57	0.50	0.48	0.37	0.32	0.26	0.23	0.20	0.19

Table 2. Lena and Fingerprints images. PSNR and SSIM comparison between WN-BM3D and BM3D. Results of 2WN-BM3D for Lena image have also been included.

Lena												
Noisy	σ	5	10	20	30	40	50	60	70	80	90	100
	PSNR	34.12	28.11	22.12	18.69	16.34	14.61	13.27	12.22	11.38	10.70	10.14
	SSIM	0.96	0.87	0.68	0.54	0.44	0.37	0.31	0.27	0.24	0.22	0.19
BM3D	PSNR	38.68	35.91	33.04	31.26	29.86	29.05	28.27	27.57	26.97	26.45	25.95
	SSIM	0.98	0.97	0.94	0.91	0.88	0.87	0.84	0.83	0.80	0.79	0.77
WN-BM3D	PSNR	40.07	36.78	33.61	31.76	30.34	29.42	28.59	27.83	27.19	26.65	26.15
	SSIM	0.99	0.97	0.95	0.92	0.89	0.87	0.84	0.82	0.79	0.77	0.75
2WN-BM3D	PSNR	39.98	36.72	33.63	31.81	30.36	29.50	28.70	28.00	27.38	26.85	26.35
	SSIM	0.99	0.97	0.95	0.92	0.89	0.88	0.85	0.83	0.81	0.79	0.77
Fingerprints												
Noisy	σ	5	10	20	30	40	50	60	70	80	90	100
	PSNR	34.13	28.12	22.15	18.75	16.42	14.69	13.36	12.31	11.46	10.77	10.20
	SSIM	0.99	0.99	0.95	0.90	0.84	0.79	0.73	0.67	0.62	0.58	0.54
BM3D	PSNR	36.49	32.45	28.80	26.82	25.30	24.52	23.75	23.12	22.56	22.06	21.61
	SSIM	1.00	0.99	0.97	0.95	0.93	0.90	0.88	0.86	0.84	0.81	0.79
WN-BM3D	PSNR	38.43	33.98	29.95	27.80	26.19	25.20	24.44	23.78	23.18	22.64	22.15
	SSIM	1.00	0.99	0.98	0.96	0.94	0.92	0.90	0.88	0.86	0.84	0.82

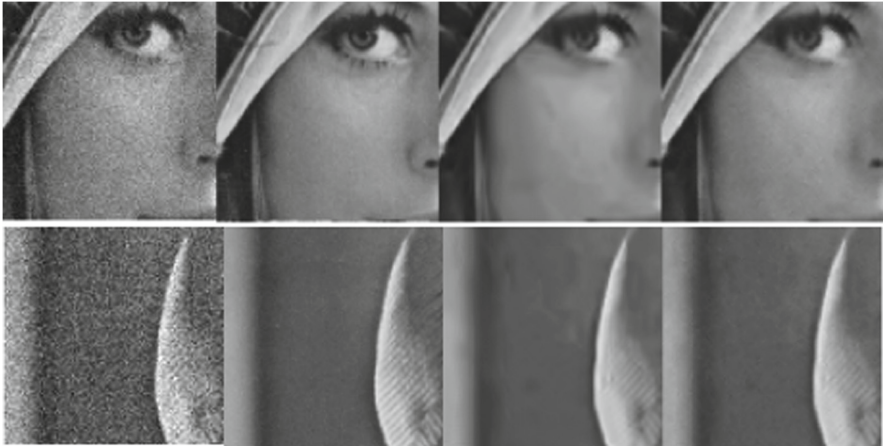


Fig. 2. Zoom of Lena in two different regions. (From left to right) Noisy; original; denoised using BM3D; denoised using WN-BM3D — noise standard deviation $\sigma = 20$.

References

1. Brunet, D., Vrscay, E.R., Wang, Z.: The use of residuals in image denoising. In: Kamel, M., Campilho, A. (eds.) ICIAR 2009. LNCS, vol. 5627, pp. 1–12. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02611-9_1
2. Bruni, V., Panella, D., Vitulano, D.: Non local means image denoising using noise-adaptive SSIM. In: Proceedings of IEEE EUSIPCO (2015)
3. Buades, A., Coll, B., Morel, J.: A review of image denoising algorithms, with a new one. *Multiscale Mod. Simul.* **4**(2), 490–530 (2005)
4. Chaudhury, K.: Acceleration of the shiftable $o(1)$ algorithm for bilateral filtering and non-local means. *IEEE Trans. Image Proc.* **22**, 1291–1300 (2012)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3D transform-domain collaborative filtering. *IEEE Trans. Image Proc.* **16**(8), 2080–2095 (2007)
6. Deledalle, C., Duval, V., Salmon, J.: Non-local methods with shape-adaptive patches (NLM-SAP). *J. Math. Imaging Vis.* **43**(2), 103–120 (2012)
7. Feng, X., Li, X., Wang, W., Jia, X.: Improvement of BM3D algorithm based on wavelet and directed diffusion. In: Proceedings of International Conference on Machine Vision and Information Technology (CMVIT) (2017)
8. Grunwald, P.D.: A tutorial introduction to the minimum description length principle. In: Grunwald, P.D., Myung, J., Pitt, M.A. (eds.) *Advances in Minimum Description Length: Theory and Applications*. MIT Press, Cambridge (2004)
9. Hou, Y.K., Zhao, C.X., Yang, D.Y., Cheng, Y.: Comments on image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **20**(1), 268–270 (2011)
10. Pizurica, A.: Image denoising algorithms: from wavelet shrinkage to nonlocal collaborative filtering. In: *Wiley Encyclopedia of Electrical and Electronics Engineering* (2017)
11. Vignesh, R., Oh, B.T., Kuo, C.C.: Fast non-local means (NLM) computation with probabilistic early termination. *IEEE Signal Proc. Lett.* **17**(3), 277–280 (2010)
12. Zhong, H., Ma, K., Zhou, Y.: Modified BM3D algorithm for image denoising using nonlocal centralization prior. *Signal Process.* **106**, 342–347 (2015)