Improving Denoising Algorithms via a Multi-scale Meta-procedure

Harold Christopher Burger and Stefan Harmeling

Max Planck Institute for Intelligent Systems, Tübingen, Germany

Abstract. Many state-of-the-art denoising algorithms focus on recovering high-frequency details in noisy images. However, images corrupted by large amounts of noise are also degraded in the lower frequencies. Thus properly handling all frequency bands allows us to better denoise in such regimes. To improve existing denoising algorithms we propose a meta-procedure that applies existing denoising algorithms across different scales and combines the resulting images into a single denoised image. With a comprehensive evaluation we show that the performance of many state-of-the-art denoising algorithms can be improved.

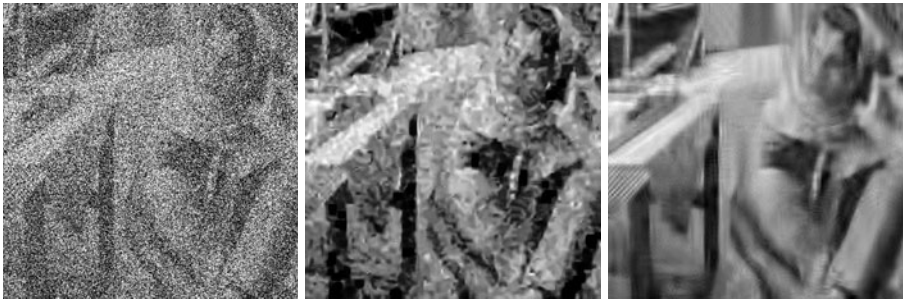
1 Introduction

The problem of removing noise from natural images has been extensively studied, so methods to denoise natural images are numerous and diverse. [4] classifies denoising algorithms into three categories: The first class of algorithms rely on smoothing parts of the noisy image [12,16,14] with the aim of "smoothing out" the noise while preserving image details. The second class of denoising algorithms exploit learned image statistics. A natural image model is typically learned on a noise-free training set (such as the Berkeley segmentation dataset) and then exploited to denoise images [11,17,5]. In some cases, denoising might involve the careful shrinkage of coefficients. For example [13,1,8,9] involve shrinkage of wavelet coefficients. Other methods denoise small images patches by representing them as sparse linear combinations of elements of a learned dictionary [3,7,6]. The third class of algorithms exploits the fact that different patches in the same image are often similar in appearance [2,4].

Denoising algorithms are usually evaluated on their ability to remove additive white Gaussian noise (AWGN). Standard test images exist for this purpose. The most popular performance measure is arguably the peak signal to noise ratio (PSNR), which is related to the mean squared error (MSE).

Hypothesis: We speculate that most denoising algorithms focus on removing noise on the higher frequencies and thus are often best suited for recovering fine-scale information. Wiener filtering, bilateral filtering [14], but also the fields of experts approach [11] rely on relatively small filters to denoise images. The small size of these filters causes these approaches to ignore larger-scale information. Denoising approaches based on dictionaries such as [3] typically decompose the

R. Mester and M. Felsberg (Eds.): DAGM 2011, LNCS 6835, pp. 206–215, 2011.



noisy
$\sigma = 200$, PSNR: 7.59dB

BM3D [2]
PSNR: 18.88dB

MS-BM3D (our approach)
PSNR: 20.96dB

Fig. 1. In high noise settings, our approach improves the results achieved with BM3D

image into small patches and then denoise the patches separately and independently. Larger-scale structure is lost when the image is decomposed into small patches. So we hypothesize that many denoising algorithms can be improved by employing a multi-scale approach.

Assumption: Our approach assumes that the statistics of natural images are invariant to changes in spatial scale. An intuitive justification for this assumption is that scenes are about equally likely to be viewed from different distances. This assumption has been successfully exploited by others [9].

Contributions: We present a meta-procedure than can be used in combination with existing denoising methods, yet often improves the results. We choose algorithms from all three categories to show that our procedure is versatile. We evaluate the PSNR on a set of 13 standard test images with varying amounts of added noise. In most cases, we use commonly available implementations of these algorithms.

Related work: Besides the denoising method mentioned above, there is a procedure that is relatively similar to ours. In [4], the authors introduce the "stochastic denoising" procedure and propose an extension (called "multi-pass denoising") in order to handle "larger-scale" noise (i.e. noise that is not uncorrelated across neighboring pixels). The extension is similar to our meta-procedure in that in addition to the original image, a single down-sampled version is denoised. The down-sampled denoised image is up-sampled and combined with the denoised image of the original size. Different from our method is that the authors combine the images using a pixel-specific linear blend between the two images. The ratio of the blend is controlled by the gradient of the image at that pixel. No quantitative evaluation was provided in [4], but we include it in our evaluation. [10] also considers a multiscale approach for image denoising by thresholding coefficients in different frequency bands.

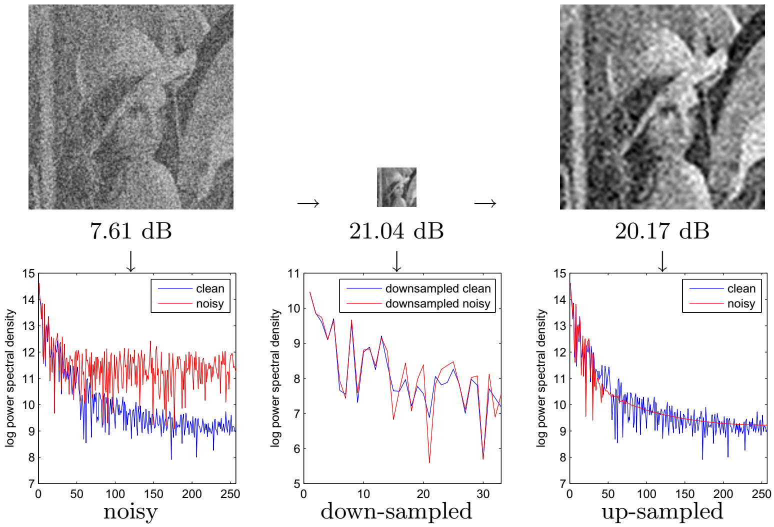


Fig. 2. Noisy Lena becomes less noisy by down- and up-sampling (top row) with power spectra (bottom row)

2 Down-Scaling Has a Denoising Effect

When an image that has been corrupted with AWGN is down-scaled, the image becomes more recognisable. The effect is illustrated in Fig. 2: adding a large amount of Gaussian noise leaves the "Lena" image barely recognisable (upper left). Nonetheless, the down-scaled version (upper middle) seems to contain much less noise.

Down-scaling an image effectively averages neighboring pixel values, causing the uncorrelated values of the noise to become smaller. Since neighboring pixels in natural images are often highly correlated, the down-scaling process is not that damaging to the image information. Another explanation is that natural images have the most energy in the low frequencies whereas AWGN is uniformly spread over the whole spectrum. Down-sampling an image keeps mainly the low frequencies, which are precisely the frequencies where the image information is strongest (bottom row in Fig. 2). Nonetheless, if the amount of noise is very large, frequencies in the middle of the spectrum are also affected, so the image information in lower frequencies should also be denoised.

3 How to Denoise Lower Frequencies

We imagine a hypothetical scenario in which we wish to recover the low frequencies of a noisy image as best as possible. To evaluate how well we recovered the low frequencies, we compare the resulting image to a down-scaled version of the ground truth image. We compare two approaches:

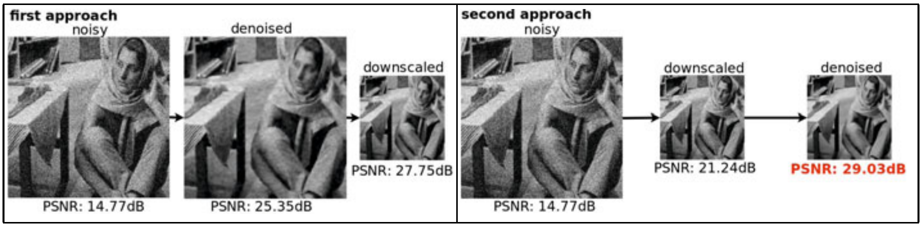


Fig. 3. Which approach better recovers the low frequencies? First down-scaling, then denoising is better than the other way around when the noise is strong

1. First denoise, then down-scale the result.
2. First down-scale, then denoise.

Which approach is better? In the first approach, the denoising algorithm has more information available, while in the second approach the denoising algorithm is applied to the down-sampled version. Denoising a down-scaled image should be an easier task, which would suggest that the second approach is better. If the second approach achieves better results, we could conclude that denoising algorithms are not good at recovering large-scale (i.e. low frequency) information, confirming the hypothesis we advanced in the introduction.

Fig. 3 compares the two approaches using KSVD as the denoising procedure. Comparing the achieved PSNRs we see that the second approach is preferable to the first. This effect also holds for other denoising algorithm for a variety of different noise settings (see supplementary material for details).

Thus we can conclude that if we wish to recover low-frequency information with a denoising algorithm that is not designed to recover low frequencies, down-scaling the image might help. Effectively the down-scaling transforms the low-frequency information into high-frequency information which can be accessed by the denoising algorithm. In the following we show how this insight can be exploited with a multi-scale procedure such that the high-frequencies are recovered from the given noisy image, while we get the low frequencies from a down-scaled version of it.

4 Multi-scale Denoising

We propose a meta-procedure that relies on denoising not only the original noisy image, but also down-scaled versions of that image. This meta-procedure is formulated such that it can be combined with any existing denoising algorithm. The last step of our procedure consists in combining the denoised images at the different scales. The combination is motivated by Laplacian pyramids. Fig. 4 summarizes our method graphically.

We will denote by $d_\alpha(x)$ a procedure that down-samples the image x by the factor α. Similarly, we denote by $u_\alpha(x)$ the procedure that up-samples the image x by the factor α. In practice, we applied Matlab's `imresize` function with the Lanczos-3 kernel. Other kernels do not lead to significantly different results.

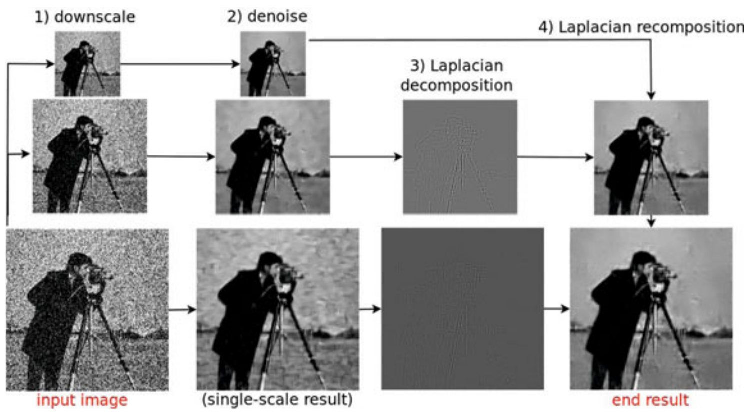


Fig. 4. Our procedure denoises a noisy image at different scales and then combines these images similarly to Laplacian pyramids

Note that resizing is a linear operator which can be represented as a matrix D. The covariance matrix of downsampled Gaussian noise is proportional to DD^T which is approximately the identity matrix for most resampling kernels (e.g. Lanczos). This fact implies that the AWGN assumption also holds for downsampled images.

Denoising at different scales. As parameters to our procedure we initially choose a denoising algorithm and scaling factors $\alpha_1, \ldots, \alpha_n$ (sorted in ascending order). Given a noisy image x_0, we create n down-sampled versions $x_1, \ldots, x_n$,

$$x_1 = d_{\alpha_1}(x_0); \quad \cdots \quad x_n = d_{\alpha_n}(x_0). \tag{1}$$

The images $x_0, \ldots, x_n$ are subsequently denoised using the same denoising procedure:

$$y_0 = \text{denoise}(x_0); \quad \cdots \quad y_n = \text{denoise}(x_n). \tag{2}$$

Next we combine the $n+1$ denoised images $y_0, \ldots, y_n$ in a Laplacian-pyramid fashion to obtain the best possible denoised image z_0 (which will have the same size as the input image x_0).

Recombining the images on the different scales. For this we decompose the image y_i into low and high frequency components l_i and h_i:

$$l_i = d_{\alpha_i/\alpha_j}(y_i) \qquad\qquad h_i = y_i - u_{\alpha_j/\alpha_i}(l_i). \tag{3}$$

Next, the low frequency information l_i is discarded and replaced by y_{i+1}, which has the same size as l_i. We do so because y_{i+1} contains more accurate low-frequency information. Combining y_{i+1} and h_i we obtain a reconstruction z_i at level i:

$$z_i = h_i + u_{\alpha_j/\alpha_i}(y_{i+1}) \tag{4}$$

which combines the best of y_i and y_{i+1}, i.e. the high frequencies of y_i and the low frequencies from y_{i+1}.

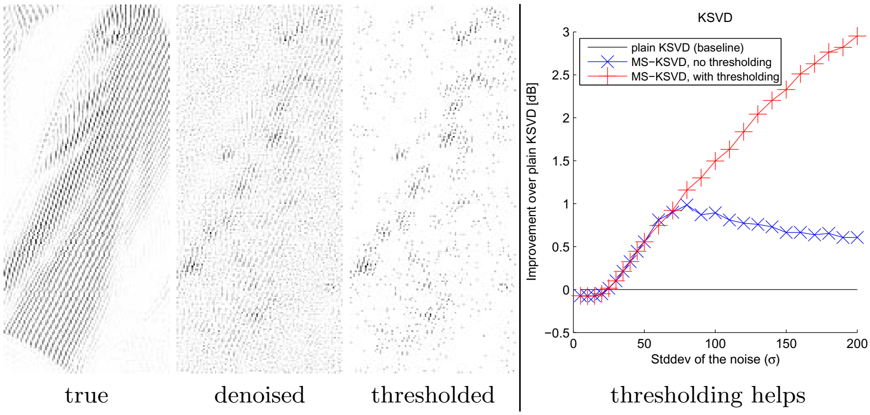


Fig. 5. Left: high frequencies of the clean image (lower-right corner of "Barbara"). Center: high frequency image of the denoised image (recovered from noisy image with $\sigma = 100$). The image contains mostly noise, but Barbara's pants are discernible. Right: thresholded high frequency image. Structure from the pants is kept. Panel on the right shows that thresholding helps.

As common for Laplacian pyramids, we start the multi-scale reconstruction with the two smallest images y_{n-1} and y_n and proceed through all scales until we reconstruct the image z_0 which is the denoising result of our method.

Shrinking high frequency coefficients. The right panel in Fig. 5 shows the benefit of using the proposed multi-scale meta-procedure with two scales in combination with the KSVD denoising algorithm. At noise levels above $\sigma = 25$, the meta-procedure (MS-KSVD, no thresholding, line '—x—') improves the results over the plain denoising algorithm (solid line). At first, the improvement grows with growing noisiness. However, when the noise becomes very strong, this effect is reversed: The multi-scale meta-procedure helps less and less. This effect is due to the fact that the high-frequency components z_i are beneficial in lower noise settings, but detrimental at higher noise levels. At very high noise levels, the denoising algorithm becomes incompetent at recovering high-frequencies. A possible solution to the problem is to attenuate the values in the high-frequency image z_i in such a way as to keep only the strongest components. We replace Eq. (4) by:

$$z_i = \mathcal{T}(h_i, \lambda) + u_{\alpha_j/\alpha_i}(y_j) \,, \tag{5}$$

where $\mathcal{T}(h_i, \lambda)$ is the hard-thresholding operator with threshold λ. Other attenuation methods lead to similar results.

The three images on the left of Fig. 5 show the effect of the hard-thresholding operator on a high-frequency image: The smaller values in the high-frequency are mostly due to errors in the denoising procedure and are successfully removed by the thresholding operation. The larger values however are unlikely to be due to errors in the denoising procedure and are therefore kept.

5 Experimental Evaluation and Results

Our meta-procedure is sensitive to the threshold parameter λ as well as to the sizes and numbers of scales used in the Laplacian pyramid. We tuned those hyperparameters for each considered denoising algorithm and for each noise level σ on a training set of 20 images from the Berkeley segmentation training dataset, see supplementary material. The smallest number of scales is 1 (no multi-scale approach) and the largest is 4. The scale sizes we chose are $(1/2)^k$ with $0 \leq k \leq 3$. This corresponds to repeatedly down-scaling by a factor of two.

As the test set, we used the 13 standard gray-scale images commonly known as: “Barbara”, “Boat”, “Cameraman”, “Couple”, “Fingerprint”, “Flintstones”, “Hill”, “House”, “Baboon”, “F16”, “Lena”, “Man” and “Peppers” (see supplementary material for images).

We applied our meta-procedure to nine state-of-the-art denoising algorithms whose implementations are commonly available. (1) Wiener filtering using Matlab's `wiener2` function with the default neighborhood size of 3. (2) Bilateral filtering [14][1] with three hyper-parameters that need to be set. Empirically, we found 10 to be a good value for the half-size of the Gaussian bilateral filter window. We chose $\sigma_1 = 3$ and set σ_2 between 10^{-4} and 2.2 depending on the noisiness of the image. (3) Bayesian least-squares Gaussian scale mixtures (BLS-GSM) [9][2], (4) Stochastic denoising [4][3], (5) Block-matching 3D (BM3D) [2][4], (6) Fields of Experts (FoE) [11][5], (7) Basis roation fields of experts (BRFoE) [17][6], and (8) Total variation denoising (TV) [12][7] all have implementations publicly available online. We used the default parameters for all methods except for FoE, where we were able to improve results over the publicly available implementation by adapting the number of iterations to the amount of noise in the image. We used our own implementation for (9) KSVD [3]. We found 10 iterations for training the dictionary to be sufficient.

Improvements for varying noise levels. Fig. 6 reports for various noise levels σ the difference between the results obtained in the single scale setting (denoted “baseline ...”) compared to our multi-scale meta-procedure (denoted “MS-...”). We also included results obtained with the “multi-pass” procedure proposed in [4] (denoted “Estrada-...”). The integer values from one to four along the line of our multi-scale procedure (“—x—”) indicate the number of scales applied.

When the noise level is low, in most cases our multi-scale meta-procedure does not improve the results of the baseline algorithm. In fact, the results are in those cases identical to the baseline algorithm. This happens when our multi-scale

[1] http://www.mathworks.com/matlabcentral/fileexchange/12191
[2] http://decsai.ugr.es/~javier/denoise/software/
[3] http://www.cs.utoronto.ca/~strider/Denoise/
[4] http://www.cs.tut.fi/~foi/GCF-BM3D/index.html#ref_software
[5] http://www.gris.informatik.tu-darmstadt.de/~sroth/research/foe/index.html
[6] http://www.cs.huji.ac.il/~yweiss/BRFOE.zip
[7] http://visl.technion.ac.il/~gilboa/PDE-filt/tv_denoising.html

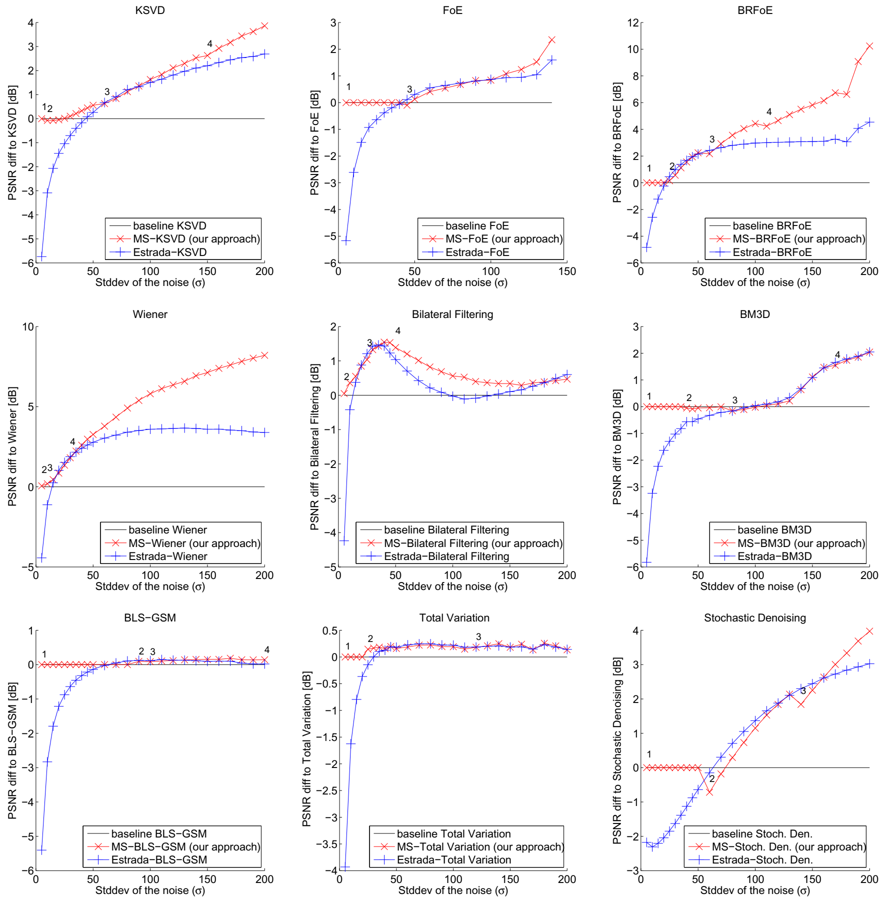


Fig. 6. Improvements achieved by combining our meta-procedure with nine different denoising algorithms. Results are averaged over 13 test images.

approach employs only the original scale (leading to the original denoising algorithm), indicating that in those noise regimes, it was not beneficial to use more scales in the training set.

The largest improvement achieved with our multi-scale meta-procedure occurs when the noise becomes stronger, which corrupts the low frequencies more and more. The improvement is particularly dramatic for Wiener and BRFoE (more than 8dB), which are patch-based methods that ignore the lower frequencies. Also KSVD is a method that is based on small patches, which also makes it blind to low frequencies, explaining the improvements obtained. However, some algorithms cannot be improved, such as BLS-GSM and Total Variation. This can be explained by the fact that BLS-GSM is a wavelet method and therefore already a multi-scale algorithm. So we see that a limitation of our meta-procedure is that it is only useful to apply it to denoising methods which are not already considering lower frequencies.

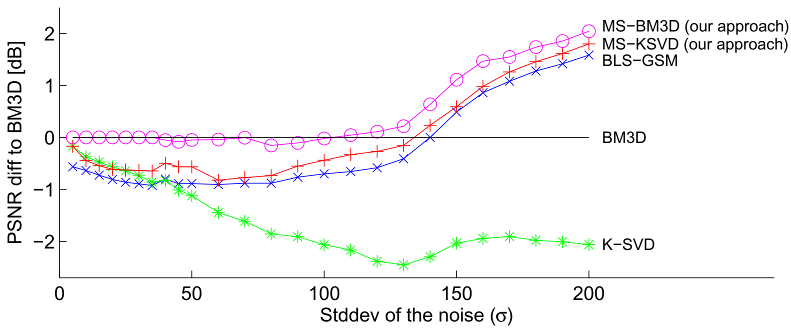


Fig. 7. The overall best methods compared to the baseline BM3D. For high noise setting our multi-scale approach applied to BM3D leads to the best results.

Note that our proposed meta-procedure outperforms the procedure by Estrada et al. [4] in almost all cases. Furthermore, our approach almost never deteriorates the denoising results, which sometimes happens for Estrada's method, especially when the noise is low. The improvements are reported in terms of PSNR, but we observed similar improvements in the structural similarity index [15] (see supplementary material).

KSVD vs. BLS-GSM revisited. In [3], the KSVD denoising algorithm is compared to BLS-GSM, described in [9]. It was noted that on the images "Peppers", "House" and "Barbara", KSVD outperforms BLS-GSM as long as the noise is below $\sigma = 50$. When the noise level is increased, BLS-GSM outperforms KSVD. We repeat the experiment on our images, but this time also report the results achieved with the multi-scale extension applied to KSVD (Fig. 7). We indeed observe that baseline KSVD outperforms BLS-GSM when the noise is low. However, the multi-scale version of KSVD outperforms BLS-GSM on all noise settings, see Fig. 7.

Multi-scale KSVD vs. BM3D. BM3D is often considered to be the best denoising algorithm currently available, even though Fig. 7 shows that for high noise levels BLS-GSM is superior. Also the multi-scale extensions of KSVD is better when the noise is very high.

Multi-scale BM3D vs. all others. Our multi-scale extension combined with BM3D delivers results that outperform all other denoising algorithms especially on the high noise levels, see Fig. 7.

6 Conclusion

For high noise levels, not only the high frequencies but also the low frequencies are corrupted. However, most image denoising algorithms are not always good at recovering low-frequency information. To improve such algorithms we devised a strategy to improve the denoising results using a multi-scale approach.

In comprehensive experiments we have shown that several state-of-the-art image denoising algorithms can be improved using this approach. Even though BM3D is arguably one of the best currently existing denoising algorithms, our method was able to improve its results on images that have been corrupted by high noise levels.

References

1. Chang, S., Yu, B., Vetterli, M.: Adaptive wavelet thresholding for image denoising and compression. IEEE Transactions on Image Processing 9(9), 1532–1546 (2002)
2. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing 16(8), 2080–2095 (2007)
3. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing 15(12), 3736–3745 (2006)
4. Estrada, F., Fleet, D., Jepson, A.: Stochastic image denoising. In: Proceedings of the British Machine Vision Conference, BMVC (2009)
5. Jain, V., Seung, H.: Natural image denoising with convolutional networks. Advances in Neural Information Processing Systems (NIPS) 21, 769–776 (2009)
6. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: IEEE 12th International Conference on Computer Vision (ICCV). pp. 2272–2279 (2009)
7. Mairal, J., Elad, M., Sapiro, G., et al.: Sparse representation for color image restoration. IEEE Transactions on Image Processing 17(1), 53 (2008)
8. Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M.: A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Transactions on Image Processing 11(5), 545–557 (2002)
9. Portilla, J., Strela, V., Wainwright, M., Simoncelli, E.: Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Transactions on Image processing 12(11), 1338–1351 (2003)
10. Rajashekar, U., Simoncelli, E.: Multiscale denoising of photographic images. In: The Essential Guide To Image Processing, pp. 241–261. Academic Press, London (2009)
11. Roth, S., Black, M.: Fields of experts. International Journal of Computer Vision 82(2), 205–229 (2009)
12. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(1-4), 259–268 (1992)
13. Simoncelli, E., Adelson, E.: Noise removal via Bayesian wavelet coring. In: Proceedings of International Conference on Image Processing (ICIP), pp. 379–382 (1996)
14. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of the Sixth International Conference on Computer Vision (ICCV), pp. 839–846 (1998)
15. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
16. Weickert, J.: Anisotropic diffusion in image processing. ECMI Series. Teubner-Verlag, Stuttgart (1998)
17. Weiss, Y., Freeman, W.: What makes a good model of natural images? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8 (2007)