IMAGE DENOISING THROUGH MULTI-SCALE LEARNT DICTIONARIES

Jeremias Sulam, Boaz Ophir, Michael Elad

Department of Computer Science, Technion - Israel Institute of Technology
e-mail: {jsulam,boazo,elad}@cs.technion.ac.il

ABSTRACT

Over the last decade, a number of algorithms have shown promising results in removing additive white Gaussian noise from natural images, and though different, they all share in common a patch based strategy by locally denoising overlapping patches. While this lowers the complexity of the problem, it also causes noticeable artifacts when dealing with large smooth areas. In this paper we present a patch-based denoising algorithm relying on a sparsity-inspired model (K-SVD), which uses a multi-scale analysis framework. This allows us to overcome some of the disadvantages of the popular algorithms. We look for a sparse representation under an already sparsifying wavelet transform by adaptively training a dictionary on the different decomposition bands of the noisy image itself, leading to a multi-scale version of the K-SVD algorithm. We then combine the single scale and multi-scale approaches by merging both outputs by weighted joint sparse coding of the images. Our experiments on natural images indicate that our method is competitive with state of the art algorithms in terms of PSNR while giving superior results with respect to visual quality.

***Index Terms*—** K-SVD, sparse, dictionary, multiscale, denoising.

1. INTRODUCTION

The problem of recovering an underlying image, or any other data, from measurements contaminated with noise is one of the most studied problems in signal processing. In the traditional set-up, a signal $\mathbf{z} \in \mathbb{R}^n$ is contaminated by additive noise $\boldsymbol{\eta}$ such that $\mathbf{y} = \mathbf{z} + \boldsymbol{\eta}$. The objective is then to recover the original signal by removing the noise from the corrupted data $\mathbf{y}$. In this work, and as it is mostly assumed, we will consider the noise to be white and iid, i.e. $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2)$.

Sparsity-based models have had a growing importance in signal processing in general, and have led to efficient algorithms in image denoising, in particular [1]. This class of methods assume that a natural signal can be expressed as a linear combination of only a few atoms from a redundant dictionary $\mathbf{D} \in \mathbb{R}^{n\times m}, n < m$. Looking for such a sparse representation accounts for solving the following problem:

$$\min_{\mathbf{x}} ||\mathbf{x}||_0 \text{ subject to } ||\mathbf{y} - \mathbf{D}\mathbf{x}||_2^2 \leq \epsilon^2, \qquad (1)$$

where $\mathbf{x} \in \mathbb{R}^m$ is the sparse representation vector for $\mathbf{y}$ within an accuracy of ϵ, and $||\mathbf{x}||_0$ counts the number of non-zeros in $\mathbf{x}$. Obtaining such a sparse representation is NP-hard in general, but several greedy algorithms and other relaxations methods are at our disposal to tackle this problem under certain conditions [1]. Methods such as the OMP [2], MP [3], FOCUSS [4] and others enable us to approximate the solution to the sparse coding problem.

A central issue in this approach is the choice of the dictionary. Transforms that are analytically defined might serve the purpose, providing also an important advantage in terms of their implementation by avoiding an explicit matrix representation. On the other side, *learning* the dictionary itself from the real data has proven to be more effective, at the expense of managing explicit matrices and more complex algorithms. This task can be written as:

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{D}\mathbf{X}||_F^2 \text{ subject to } ||\mathbf{x}_i||_0 \leq T, \ \forall i, \qquad (2)$$

where $\mathbf{Y} \in \mathbb{R}^{n\times N}$ is a matrix containing N signal examples, and $\mathbf{X} \in \mathbb{R}^{m\times N}$ are the corresponding sparse vectors, both ordered column wise. Several authors have proposed iterative methods to undertake this problem [5, 6]. Among them, the K-SVD algorithm [7] has been widely used for different applications in image processing. In denoising in particular, one could seek for a sparse representation over a dictionary trained to perform optimally for some group or kind of images. Interestingly, the same denoising task can be dealt with *while* training a dictionary on the noisy image itself achieving better performance [8].

Adopting a broader view, a common feature in most state of the art denoising methods is a patch-based concept: when dealing with high dimensional data, the motif is to work on overlapping patches of size $\sqrt{n}\times\sqrt{n}$, and then tile and average the results. Some of these include the NLM [9], BM3D [10], LSSC [11], and the K-SVD is no exception to this common approach. While this strategy provides a practical solution under this framework, several problems arise from working in a single and small scale scheme. As we will show later, no matter what the core algorithm is, this affects the visual quality of

Acknowledgements: The research leading to these results has received funding from the European Research Council under European Union's Seventh Framework Programme, ERC Grant agreement no. 320649.

the denoised image. The resulting artefacts are more noticeable in large smooth areas, typical of commonplace images such as scenery and landscapes. What is there to gain by these algorithms from a more global approach? Could a multi-scale framework provide an improvement to these methods?

We propose to merge the K-SVD denoising algorithm [8] with a wavelet analysis, in a similar way to the approach taken in [12]. This will lead to an effective sparse decomposition of the image content using different scale atoms in a natural way. As a result, the potential of the K-SVD denoising algorithm is exploited beyond the single scale limitations, reaching state of the art results. In this paper we describe this idea in details, tie it and contrast it to existing work, and demonstrate the effectiveness of the proposed scheme.

2. RELATED WORK

The idea of combining dictionary learning with a multi-scale analysis framework is not new. In [13], the authors proposed to train wavelet coefficients with a sparsity inducing prior on a wavelet pyramidal decomposition structure, achieving slightly better results for compression. Later, the authors in [14] used different size patches taken from a quadtree structure to train a multi-scale dictionary, in a first extension of the K-SVD algorithm to a multi-scale scheme.

The work presented in [12] introduced the construction of true multi-scale dictionaries by learning patch based atoms in the analysis domain of the wavelet transform. In this case, the resulting dictionary appears as the multiplication of a wavelet synthesis matrix with a learnt dictionary in the wavelet domain, i.e., $\tilde{\mathbf{D}} = \mathbf{W_S D}$, where $\mathbf{W_S}$ is the synthesis matrix of a wavelet (inverse) transform. However, choosing an orthogonal wavelet with periodic extension enables the authors to work in the analysis domain instead, by solving the following optimization problem:

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{W_A Y} - \mathbf{DX}||_F^2 \text{ subject to } ||\mathbf{x}_i||_0 \leq T,\ \forall i, \tag{3}$$

where $\mathbf{W_A}$ is the analysis operator (wavelet transform) matrix. This expression suggests to adapt the atoms to sparsely represent the wavelet coefficients of the different training examples. Moreover, the authors proposed to train different sub-dictionaries $\mathbf{D_b}$ per band by employing K-SVD on 8×8 patches of the wavelet sub images. This simple scheme allows to work with different sized atoms, since a patch in a first decomposition level implies an effective patch of four times its area in the image domain. Once the collection of sub-dictionaries is trained, the authors in [12] use a *global* framework for the sparse coding stage, where the patches from different scales *compete* for additional coefficients selecting the one that gives the most profit in terms of the residual energy, with a global variant of the OMP algorithm.

All these approaches have looked for a better representation of some class of data or images in terms of some dictionary. As such, they fail to treat the denoising task competitively, as indeed demonstrated in [12]. In [8], the K-SVD denoising algorithm was formally derived by proposing a global image prior that forces patch-based local sparsity over patches in every location of the image. The problem is solved iteratively using an error threshold for the sparse coding which depends on σ, treating each patch *independently*. We will make use of this concept and extend it to a multi-scale framework.

3. OUR CONTRIBUTION

In this paper we propose to continue and extend the work in [12], and tackle specifically the denoising problem. In [12] the authors have shown an example of a naive denoising through *M-term* approximation, using a global pursuit. The results reported in this method were not competitive with the single-scale K-SVD. In this paper we propose to adapt the multi-scale sub-dictionaries to the noisy image itself and treat the pursuit locally. This resembles the work in [8], but in a multi-scale scenario. Each band of the decomposition is treated separately, training a subdictionary for each band, which is then used to denoise the corresponding wavelet coefficients. In a final stage, the multi-scale K-SVD and the traditional (single-scale) K-SVD denoised images are combined though a weighted joint sparse coding in order to benefit from the advantages that each bring. This last step allows us to maximize the information shared between the two images, and obtain a better estimate for the original signal.

3.1. Multi-scale K-SVD denoising

Consider a noisy image $\mathbf{Y}$, its wavelet transform as a collection of band images $\mathbf{Y}_b^W = (\mathbf{W}_A \mathbf{Y})_b$, and its estimated denoised version $\hat{\mathbf{Z}}_b^W$, $b = 1, ..., L = 3S + 1$, with S decomposition levels. Generalizing the work in [8], we propose a global maximum *a posteriori* (MAP) estimator for denoising the image in the wavelet domain as

$$\begin{aligned} \forall b, \{\mathbf{x}_{ij,b}, \mathbf{D}_b, \hat{\mathbf{Z}}_b^W\} = \underset{\mathbf{x}_{ij,b},\mathbf{D}_b,\mathbf{Z}_b^W}{\arg\min}\ \lambda ||\mathbf{Y}_b^W - \mathbf{Z}_b^W||_2^2 \\ + \sum_{ij} \mu_{ij,b} ||\mathbf{x}_{ij,b}||_0 + \sum_{ij} ||\mathbf{D}_b \mathbf{x}_{ij,b} - \mathbf{R}_{ij,b} \mathbf{Z}_b^W||_2^2, \end{aligned} \tag{4}$$

where $\mathbf{x}_{ij,b}$ is the sparse vector for the $(i,j)-$patch in the decomposition band b, $\mathbf{R}_{ij,b}$ is a matrix that extracts that patch from the sub-image $\mathbf{Z}_b^W$, and λ is a Lagrange multiplier. This optimization problem can be solved iteratively by first considering a fixed set of dictionaries $\mathbf{D}_b$ and obtaining the vectors $\mathbf{x}_{ij,b}$ by any pursuit method. Then the sub dictionaries are updated using a K-SVD step. These steps are repeated for a fixed number of iterations. Finally, we update $\mathbf{Z}_b^W$ by

$$\hat{\mathbf{Z}}_b^W = \left(\lambda \mathbf{I} + \sum_{ij} \mathbf{R}_{ij,b}^T \mathbf{R}_{ij,b}\right)^{-1} \left(\lambda \mathbf{Y}_b^W + \sum_{ij} \mathbf{R}_{ij,b} \mathbf{D}_b \mathbf{x}_{ij,b}\right). \tag{5}$$

After the different sub band images have been denoised in the wavelet domain, the multi-scale denoised image is obtained by applying the inverse wavelet transform. Note that by working on patches of the same size in all decomposition

Fig. 1. One image from the dataset in [15], and its denoising results (noise level $\sigma_\eta = 35$). Top left: original image. Top right: K-SVD (PSNR = 31.01, SSIM = 0.857). Bottom left: BM3D (PSNR = 33.16, SSIM = 0.923). Bottom right: Fused K-SVD algorithm (PSNR = 33.16, SSIM = 0.940). Note the artefacts in the single scale patch based methods.

levels, we consider different-scale *effective* patches in the image domain. This gives our algorithm a more global outlook than that of the regular K-SVD denoising algorithm, and involves essentially the same computational complexity, plus the forward and backward wavelet transform. The complexity analysis detailed in [12] is still valid here.

3.2. Fusing Single and Multi-Scale Results

After this multi-scale K-SVD denoising stage, we go one step further. While working on the wavelet coefficients on the different scales $1, 2, \ldots, S$, we miss considering the *scale 0*. Following this motivation, we propose to merge the outcome of the original (*single-scale*) K-SVD denoised image $\hat{Z}_{ss}$ with the output of the *multi-scale* K-SVD algorithm proposed here, $\hat{Z}_{ms}$. Both of these have some remaining noise and different artefacts, but correspond to the same underlying image. We aim to recover the information common to both of them by a weighted joint sparse coding, as motivated by [16] and shown in Fig. 2. We concatenate corresponding patches of both images with a weighting factor β as $\tilde{\mathbf{y}} = [\, \mathbf{y}_{ms}^T \sqrt{1+\beta}\ ,\ \mathbf{y}_{ss}^T \sqrt{1-\beta}\,]^T \in \mathbb{R}^{2n}$. We may then use the dictionary given by $\mathbf{A} = [\, \mathbf{D}^T \sqrt{1+\beta}\ ,\ \mathbf{D}^T \sqrt{1-\beta}\,]^T \in \mathbb{R}^{2n \times m}$ to obtain the sparse vector $\boldsymbol{\alpha} \in \mathbb{R}^m$ by the OMP algorithm. Finally, the denoised patch will be given by $\hat{\mathbf{z}} = \mathbf{D}\boldsymbol{\alpha}/\sqrt{2}$.

As we will see later, the multi scale K-SVD algorithm outperforms the single scale K-SVD specially in the presence of high noise due to the increasing patch-like artefacts, which the multi scale approach is more robust to. This indicates that

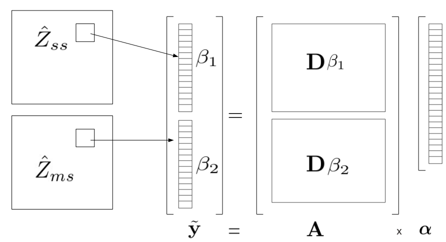

Fig. 2. Joint sparse coding stage of the Fused K-SVD denoising algorithm. $\beta_1 = \sqrt{1+\beta}$, $\beta_2 = \sqrt{1-\beta}$.

β should be close to 1 in such cases, and close to 0 when the noise level is lower. One could just propose a function $\beta = f(\sigma_\eta)$ accordingly, or choose an adaptive method that optimizes this parameter for each patch. For the sake of simplicity we consider here a linear function of the initial noise level, from $\beta = 0$ for $\sigma = 0$ to $\beta = 0.9$ for $\sigma = 50$. Certainly other choices are possible, and the implications of this choice will be commented later on.

4. EXPERIMENTS

In this section we present the results of a denoising experiment on landscape images from the online NOAA library [15]. We chose these images as they contain large scenery areas that are poorly treated by typical patch-based denoising methods. One of this images is depicted in the top left corner of Fig.1. Fifteen images from this dataset, size 870×1360, were contaminated by white Gaussian noise with zero mean and variable standard deviation σ. For the multi-scale decom-

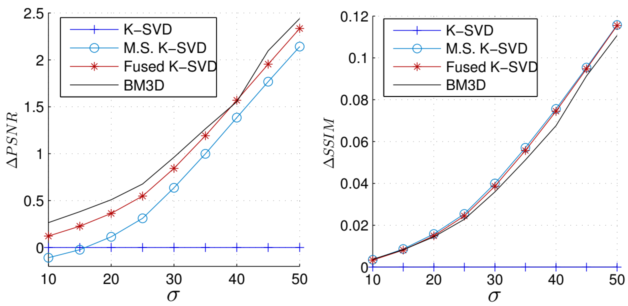

Fig. 3. Denoising results by the K-SVD [8], BM3D [10], multi-scale K-SVD and Fused K-SVD algorithms, averaged over 15 testing images from the NOAA database [15]. Left: PSNR gain with respect to K-SVD. Right: SSIM gain with respect to K-SVD.

position we used a discrete Meyer wavelet, with 2 decomposition levels. By choosing a unitary transform, the stoping criteria for the sparse coding stage in the denoising algorithm is simply $\epsilon = c \cdot \sigma$, where $c = 1.15$ following [8].

We evaluate our denoising results with two image quality measures: the popular Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) [17]. While simple and practical, the PSNR relies only on the absolute difference pixel by pixel, and does not provide a good signal fidelity measure [18]. As such, its ability to compare images from a human perception point of view is poor. The SSIM is somehow a more complete image quality measure, which builds upon the idea that human perception is highly adaptive to structural information from images and visual scenes [17]. We include in the results those obtained by the BM3D algorithm [10], computed with the code made available by the authors, and with their recommended parameters. We also compare our performance against the regular single-scale K-SVD. Note that all three methods use 8×8 patches.

We may also benefit from choosing an appropriate initial dictionary [8]. To this end, we trained a single-scale and a multi-scale dictionary on 20 natural images (outside the above set of test images), for the single-scale and multi-scale versions of the K-SVD algorithm, respectively. The same single scale initial dictionary was later used to merge the final outcome of the *Fused K-SVD* algorithm, as described in the previous section. In this case we use OMP with an error threshold of $\epsilon = 0.1 \cdot \sigma_\eta$, where this factor has been chosen empirically, accounting not only for the remaining noise but also for the difference in the artefacts of the two images.

In Fig. 3 we present the averages over all testing images for the different algorithms, relative to that of K-SVD. The multi-scale K-SVD outperforms the single-scale K-SVD in almost the whole range of noise variance, and the Fused K-SVD and BM3D present the best results, with the latest being slightly higher in terms of PSNR. Note that the last weighted joint sparse coding stage enables an extra boost, and the full fused algorithm improves the results by 0.2 - 0.3 dB compared to the plain multi-scale K-SVD. Turning to the SSIM results, the artefacts on the smooth areas in the regular K-SVD denoised images are strongly penalized by this measure. Multi-scale K-SVD and Fused K-SVD seem to be the best, with our methods slightly outperforming BM3D. Fusion gives no gain with respect to this measure. In Fig. 1 we depict the results of the K-SVD, BM3D and Fused K-SVD on the example image.

The reason for this difference in both measures should not be surprising. While BM3D makes little mistakes in terms of absolute value, these errors are more *noticeable* when there are large smooth areas, which causes the annoying texture artefacts that can be seen in the images in Fig. 1. It is in these areas where our method shows its greatest benefits. The coding of the deeper decomposition levels implies choosing big atoms yielding nicely codded smooth patches. These atoms are treated considering a more global approach than just looking at a 8×8 patch in the image domain. This makes the method more robust to higher noise levels, where the texture artefacts become stronger. However, this advantage comes at the cost of losing some details in the sharp edges of the image. The absolute error at these points are slightly higher that those made by BM3D, as noted by the PSNR results.

To finish this section, we have a word about the standard images such us *Lena, Barbara, etc*. In these cases -that we do not reproduce here due to the lack of space- the performance of the Fused K-SVD algorithm is between 0.3-0.4 dB (PSNR) and 0.002-0.01 (SSIM) lower than BM3D. Note that these images are small (512×512) and hardly present any smooth areas of considerable size. Even in these images, however, there is a notable improvement over the regular K-SVD in both measures (up to 0.55 dB in PSNR and 0.035 in SSIM).

5. CONCLUSION

We have presented a multi-scale extension of the K-SVD denoising algorithm by proposing a global MAP estimator for the denoised image in the wavelet domain. We solve this minimization problem iteratively in terms of the K-SVD algorithm per band, applying a multi-scale patch denoising of the image. We then boost the results by fusing the single scale and multi-scale K-SVD outcome images by a weighted sparse coding step. The results obtained by this method show the potential benefits of working within a multi-scale framework. We are able to combine bigger effective atoms that give rise to clear smooth areas, in which most current methods fail.

The combination of the regular and multi-scale K-SVD denoised images could be improved by proposing a patch-based adaptive weight instead of a global one, and the joint sparse coding alternative is effective, but not necessarily the only one. An orthogonal wavelet transform enables a simple multi-scale analysis, but other multi-scale transforms might yield improvements on this framework and are worth exploring. Moreover, this multi-scale approach is not restricted to the K-SVD algorithm, and the question posed in the introduction still holds for other methods. We have shown that there is still a lot to gain from patch-based methods, and similar extensions for these could be proposed.

6. REFERENCES

[1] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[2] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal Matching Pursuit: Recursive Function Approximat ion with Applications to Wavelet Decomposition," *Asilomar Conf. Signals, Syst. Comput. IEEE.*, pp. 40–44, 1993.

[3] S. Mallat and Z. Zhang, "Matching Pursuits With Time-Frequency Dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[4] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[5] K. Engan, S. O. Aase, and J. H. Husoy, "Method of Optimal Directions for Frame Design," in *IEEE Int. Conf. Acoust. Speech, Signal Process.*, 1999, pp. 2443–2446.

[6] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Dictionary Learning for Sparse Coding," in *26th Int. Conf. Mach. Learn.*, Montreal, Canada, 2009.

[7] M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[8] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries.," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–45, Dec. 2006.

[9] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," *Conf. Comput. Vis. Pattern Recognition, IEEE.*, pp. 60–65, 2005.

[10] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," *Proc. SPIE-IS&T Electron. Imaging*, vol. 6064, pp. 1–12, 2006.

[11] J. Mairal, F. Bach, and G. Sapiro, "Non-local Sparse Models for Image Restoration," *Int. Conf. Comput. Vis., IEEE*, vol. 2, pp. 2272–2279, 2009.

[12] B. Ophir, M. Lustig, and M. Elad, "Multi-Scale Dictionary Learning Using Wavelets," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 5, pp. 1014 – 1024, Sept. 2011.

[13] P. Sallee and B. A. Olshausen, "Learning Sparse Multiscale Image Representations," *Adv. Neural Neural Inf. Process. Syst.*, vol. 15, pp. 1327–1334, 2002.

[14] J. Mairal, G. Sapiro, and M. Elad, "Multiscale Sparse Image Representation with Learned Dictionaries," *Int. Conf. Image Process.*, vol. 3, pp. 105–108, 2007.

[15] National Oceanic and Atmospheric Administration U.S. Department of Commerce, "Noaa photo library," `http://www.photolib.noaa.gov`.

[16] N. Yu, T. Qiu, and F. Ren, "Denoising for Multiple Image Copies through Joint Sparse Representation," *J. Math. Imaging Vis.*, vol. 45, pp. 46–54, 2013.

[17] 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. 600–12, Apr. 2004.

[18] Z. Wang and A. C. Bovik, "Mean Squared Error : Love It or Leave It?," *IEEE Signal Process. Mag.*, vol. 26, no. January, pp. 98–117, 2009.