DENOISING OF IMAGES USING DESIGNED SIGNAL DEPENDENT FRAMES AND MATCHING PURSUIT

Kjersti Engan, Karl Skretting and John Håkon Husøy

University of Stavanger
Dept. of Electrical and Computer Eng.
N-4036 Stavanger, Norway
{kjersti.engan,karl.skretting}@uis.no

ABSTRACT

The use of frames, or overcomplete dictionaries, for sparse signal representation has been given considerable attention in recent years. The major challenges are good algorithms for sparse approximations, and good methods for choosing or designing frames. We are concerned with the latter, and have developed algorithms for training frames for a class of data and a specific application. The application presented in this paper is denoising of images with additive Gaussian noise. We present a method for training of constrained overlapping frames to be used for denoising of images. Experiments show that the proposed method improves denoising results compared to adaptive Wiener filtering and wavelet denoising.

1. INTRODUCTION

Signal expansions using *frames*, or *overcomplete dictionaries*, can be regarded as generalizations of signal expansions based on transforms, filter banks, and wavelets. The use of transforms, filter banks, or wavelets to represent a signal corresponds to expressing signal vectors as linear combinations of basis vectors. Forming the hypothesis that the representation capabilities of sparse linear combinations will increase with the addition of more vectors than what is needed to form a basis, leads to the exploration of frames or overcomplete dictionaries.

In many application we are interested in a *sparse approximation* of a signal rather than an exact representation. The reconstructed N-dimensional signal vector (often an approximation of $\mathbf{x}$) can be written as:

$$\tilde{\mathbf{x}} = \mathcal{F}\mathbf{w} = \sum_{i=1}^{K} w_i \mathbf{f}_i, \tag{1}$$

where $K > N$, i.e. the number of dictionary vectors is larger than the dimension, thus the vectors $\mathbf{f}_j$ are not linearly independent. Finding the *best* sparse approximation is NP-hard [1], and suboptimal vector selection algorithms are used. We use Order Recursive Matching Pursuit (ORMP) throughout this work [2].

Whenever an image is converted from one form to another, e.g., copied, scanned, digitized, transmitted, printed, or compressed, many types of noise or degradations can be present in the image. Hence, an important subject is the development of image enhancement algorithms that remove (smooth) noise artifacts while retaining image structure. Classical image denoising techniques are based on filtering and tend to blur the image [3].

Adaptive Wiener filtering is a classical denoising technique and is used as a reference system in this work. More recently wavelet-based denoising techniques have been recognized as powerful tools for denoising. Donoho and Johnstone [3, 4] introduced the use of wavelets in denoising, and their method is also a reference system used in this work. In the Donoho and Johnstone method a Discrete Wavelet Transform (DWT) is performed on the noisy image, the scaling coefficients are left unchanged, while the wavelet coefficients are thresholded to remove the noise. For additive Gaussian noise the thresholding is usually done as soft thresholding. In this paper we propose thresholding already sparse frame coefficients, w_i of Eq. 1, using custom designed frames.

2. TRAINING OF OVERCOMPLETE DICTIONARIES/FRAMES

Design of frames to use for specific applications or datasets has been given some, but relatively modest, attention [5, 6, 7, 8], and frames are often chosen rather than optimized.

We have developed frame design algorithms inspired by the Generalized Lloyd Algorithm (GLA) for vector quantization codebook design to be able to train a frame to perform well for a class of data and for specific applications. The original algorithm was constructed for unconstrained block oriented frames [7] and was later extended to overlapping frames, with and without different kinds of constraint like predefined structure and symmetries[9].

Unconstrained overlapping frames can become unpractical for signals with more than one dimension. We impose constraint on the overlapping frames to make them more tractable for signals of more than one dimension. A brief overview of the design algorithm is given in the following. More details can be found in [9].

For a one dimensional signal and the block-oriented case the synthesis equation for a collection of signal blocks can be stated as $\tilde{\mathbf{x}} = \mathcal{F}\,\mathbf{w}$ or equivalently

$$\begin{bmatrix} \vdots \\ \tilde{\mathbf{x}}_l \\ \tilde{\mathbf{x}}_{l+1} \\ \tilde{\mathbf{x}}_{l+2} \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & & & \\ & \mathbf{F} & & \\ & & \mathbf{F} & \\ & & & \mathbf{F} \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ \mathbf{w}_l \\ \mathbf{w}_{l+1} \\ \mathbf{w}_{l+2} \\ \vdots \end{bmatrix}. \tag{2}$$

In the following, let $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1\tilde{\mathbf{x}}_2 \dots \tilde{\mathbf{x}}_L]$, $\mathbf{W} = [\mathbf{w}_1\mathbf{w}_2 \dots \mathbf{w}_L]$, and the training data $\mathbf{X} = [\mathbf{x}_1\mathbf{x}_2 \dots \mathbf{x}_L]$. The optimal frame will depend on the target sparseness factor and the class of signals we want to represent. We want to find the frame, $\mathbf{F}$, of size $N \times K$ where $K > N$, and the sparse coefficient vectors, $\mathbf{w}_l$, that minimize the sum of the squared errors. The objective function to be minimized is

$$J = J(\mathbf{F}, \mathbf{W}) = \|\mathbf{X} - \tilde{\mathbf{X}}\|^2 = \|\mathbf{X} - \mathbf{F}\mathbf{W}\|^2. \tag{3}$$

Finding the optimal solution to this problem is difficult if not impossible. We split the problem into two parts to make it more tractable. The iterative solution strategy presented below results in good, but in general suboptimal, solutions to the problem.

The algorithm starts with a user supplied initial frame $\mathbf{F}^{(0)}$ and then improves it by iteratively repeating two main steps:

1. $\mathbf{W}^{(i)}$ is found by vector selection using frame $\mathbf{F}^{(i)}$, where the objective function is $J(\mathbf{W}) = \|\mathbf{X} - \mathbf{F}^{(i)}\mathbf{W}\|^2$ and a sparseness constraint is imposed on $\mathbf{W}$.
2. $\mathbf{F}^{(i+1)}$ is found from $\mathbf{X}$ and $\mathbf{W}^{(i)}$, where the objective function is $J(\mathbf{F}) = \|\mathbf{X} - \mathbf{F}\mathbf{W}^{(i)}\|^2$. This gives:

$$\mathbf{F}^{(i+1)} = \mathbf{X}(\mathbf{W}^{(i)})^T \big(\mathbf{W}^{(i)}(\mathbf{W}^{(i)})^T\big)^{-1} \tag{4}$$

Then we increment i and go to step 1.

i is the iteration number. The first step is suboptimal due to the use of practical vector selection algorithms, while the second step finds the $\mathbf{F}$ that minimizes the objective function.

When we extend our design strategy to the general overlapping case, the large frame, $\mathcal{F}$, can be written as

$$\mathcal{F} = \begin{bmatrix} \ddots & \mathbf{F}_1 & & & \\ \ddots & \vdots & \mathbf{F}_1 & & \\ \ddots & \mathbf{F}_P & \vdots & \mathbf{F}_1 & \ddots \\ & & \mathbf{F}_P & \vdots & \ddots \\ & & & \mathbf{F}_P & \ddots \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_P \end{bmatrix}. \tag{5}$$

The synthesis vectors are the columns of $\mathcal{F}$ or $\mathbf{F}$. $\mathbf{F}$ (of size $NP \times K$) can be partitioned into P submatrices, $\{\mathbf{F}_p\}_{p=1}^{P}$ each of size $N \times K$.

In [9, 10] $\mathcal{F}$ was set to be the product of two matrices, $\mathcal{F} = \mathcal{G}\mathcal{H}$, corresponding to

$$\begin{bmatrix} \ddots & & \\ & \mathbf{F} & \\ & & \mathbf{F} \\ & & & \ddots \end{bmatrix} = \begin{bmatrix} \ddots & & \\ & \mathbf{G} & \\ & & \mathbf{G} \\ & & & \ddots \end{bmatrix} \begin{bmatrix} \ddots & & \\ & \mathbf{H} & \\ & & \mathbf{H} \\ & & & \ddots \end{bmatrix} \tag{6}$$

with $\mathcal{G}$ as the large matrix with $\mathbf{G}$, of size $NP \times N$, along the diagonal. $\mathcal{H}$ denotes the large matrix with $\mathbf{H}$, of size $N \times K$, along the diagonal. The structure of the first matrix, $\mathcal{G}$, corresponds to the synthesis matrix of a critically sampled FIR synthesis filter bank. The constituent matrices of $\mathcal{F}$, the $\mathbf{F}$ matrices, are each of size $NP \times K$ and defined by

$$\mathbf{F} = \mathbf{G}\mathbf{H} = \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_P \end{bmatrix} \mathbf{H} = \begin{bmatrix} \mathbf{G}_1\mathbf{H} \\ \vdots \\ \mathbf{G}_P\mathbf{H} \end{bmatrix}. \tag{7}$$

The signal representation is now $\tilde{\mathbf{x}} = \mathcal{F}\,\mathbf{w} = \mathcal{G}\mathcal{H}\,\mathbf{w}$.

The task of designing $\mathcal{F}$ can now be divided into two parts: Selecting a reasonable $\mathcal{G}$, which we then keep fixed, and finding an $\mathcal{H}$ (or equivalently its constituent matrices $\mathbf{H}$) using the method described for block oriented frames. By fixing $\mathcal{G}$ a restriction is imposed on $\mathcal{F}$, the synthesis vectors of $\mathbf{F}$ may not be freely selected vectors from $\mathcal{R}^{PN}$, but are restricted to be in the N dimensional subspace of $\mathcal{R}^{PN}$ spanned by the N columns of $\mathbf{G}$. The objective function will now be

$$J = J(\mathbf{H}) = \|\mathbf{x} - \mathcal{G}\mathcal{H}\mathbf{w}\|. \tag{8}$$

Suppose that the columns of $\mathcal{G}$'s constituent matrices, $\mathbf{G}$, are chosen as the synthesis vectors (filter responses) of an *orthogonal* perfect reconstruction filter bank, then $\mathcal{G}^{-1} = \mathcal{G}^T$ and the norm is conserved, $\|\mathbf{x}\| = \|\mathcal{G}\mathbf{x}\| = \|\mathcal{G}^{-1}\mathbf{x}\|$. This implies that

$$J = \|\mathbf{x} - \mathcal{G}\mathcal{H}\mathbf{w}\| = \|\mathcal{G}^{-1}(\mathbf{x} - \mathcal{G}\mathcal{H}\mathbf{w})\| = \|\mathcal{G}^T\mathbf{x} - \mathcal{H}\mathbf{w}\|, \tag{9}$$

and we can design $\mathcal{H}$ in exactly the same manner as we design a block-based frame. The only difference is that we use $(\mathcal{G}^T\mathbf{x})$ rather than $\mathbf{x}$ as the training signal. *That is, we do the approximation in the coefficient domain rather than in the signal domain.*

This technique can be generalized to two dimensions [10]. $\mathbf{G}$ is chosen as the synthesis vectors of an *orthogonal* perfect reconstruction filter bank. Examples can be Lapped Orthogonal Transform (LOT) [11] and Extended Lapped Orthogonal Transform (ELT), and different Discrete Wavelet Transforms (DWT). Finding the representation coefficients, $\mathbf{w}$, is a two step process. First the analysis corresponding to the synthesis, $\mathbf{G}$, is performed *separably*, and we obtain the analysis-coefficients $(\mathcal{G}^T\mathbf{x})_{2D}$. The second step is to find a sparse approximation of these coefficients using the block oriented frame $\mathcal{H}$, and this is done *non-separably*. Therefore the analysis-coefficients, $(\mathcal{G}^T\mathbf{x})_{2D}$, are rearranged into vectors before being represented by a sparse approximation, $\mathbf{w}$, using the frame $\mathcal{H}$. Experimentally we found it beneficial to let one coefficient from each frequency band be arranged lexiographically into a vector when $\mathbf{G}$ is chosen to be a lapped transform, LOT/ELT. When $\mathbf{G}$ is chosen to be a 3-level DWT the coefficients are arranged into vectors according to a wavelet tree, i.e. each vector consists of one coefficients from each of the 3-level decompositions, four from the three 2-level decompositions and 16 from the three 1-level decompositions. $\mathbf{H}$ can be designed and used in the coefficient domain by using the block-based frame design algorithm presented in the beginning of this section.

3. DENOISING USING OVERLAPPING FRAMES

Since DWT has proven to be successful in denoising of images [3], and since overlapping frames can be regarded as a generalization of DWT or filter banks, it was natural to try denoising of images using overlapping frames. The main idea of denoising by using an efficient signal representation, including both DWT and overlapping frames, is that hopefully the noise space is quite different from the signal space. By trying to efficiently represent the signal space, the parts that are left out are mainly noise.

Using DWT and soft thresholding, many of the small wavelet coefficients are set to zero, thus we get a sparse representation of the image. Soft thresholding also lowers the value of the wavelet coefficients that are kept to reduce the noise influence in the signal space.

Method	PSNR [dB] - Lena		
	$\sigma^2 = 100$	$\sigma^2 = 400$	$\sigma^2 = 900$
Wiener			
3×3	32.7	28.7	25.5
5×5	31.7	29.3	27.3
Wavelet	T=15	T=30	T=60
DWT db4	32.0	28.8	26.9
DWT db8	32.0	28.8	26.9
DWT coif5	32.2	28.9	27.1
Frame	T=15	T=30	T=60
MSE$_{lim}$:	80	170	240
$\mathbf{G}$=ELT	31.8	28.9	26.9
$\mathbf{G}$=db4	32.1	29.3	27.5
$\mathbf{G}$=db8	32.1	29.2	27.3
$\mathbf{G}$=coif5	32.2	29.3	27.5

Table 1. Denoising of Lena images with different σ^2. T is soft thresholding level. MSE$_{lim}$ is local quality criterion for each block of coefficients, deciding local sparsity.

There are two factors involved in the denoising using overlapping frames. First we find a sparse approximation of the image, and the degree of sparseness must be chosen. One possibility used in [12] is to decide how many coefficients to be nonzero in the representation of the entire image. Sparsity can also be imposed by deciding for a quality criterion (Mean Squared Error (MSE)) for the representation of a signal segment. Frame vectors are selected until the quality criterion is satisfied, letting the local sparseness be decided from the quality criterion. The latter performs somewhat better and is used in this work. A considerable part of the noise should be removed just by finding this efficient sparse representation of the image. The other factor is soft thresholding of the nonzero coefficients. This can put some of the chosen non-zero coefficients to zero, resulting in an even more sparse representation, but mostly it means reducing the value of the coefficients.

4. EXPERIMENTS

Experiments were done on images with additive Gaussian noise with σ^2 =100, 400, and 900. The frames used in these experiments were trained on a training set of 8 natural graylevel images with 8 bit per pixel (bpp), and they were tested on images outside the training set. In all experiments $\mathbf{H}$ was trained using the same choices for $\mathbf{G}$ as was done for testing.

Table 1 shows PSNR results for the best experiments on the test image Lena with the different denoising methods, and Table 2 shows PSNR values for denoising on two different images both with σ^2 =400. In denoising applications the visual results are more important than the PSNR values, so Figure 1 and 2 depicts details from the denoising experiments on Lena with $\sigma^2 = 100$ and $\sigma^2 = 400$, respectively. In figure 3 parts of the image Julie, from the denoising experiments with $\sigma^2 = 400$ are shown. The visual test as well as the PSNR test shows that the frame method, especially selecting $\mathbf{G}$ as a wavelet, performs better than the pure wavelet method. Wiener filtering with 3×3 mask performs best in terms of PSNR but a lot of the noise is still left in the image. Wiener filtering with 5×5 mask over-smooths the image, and edges and details are preserved considerably better using frames or DWT.

Method	PSNR [dB] - $\sigma^2 = 400$	
	Boat	Julie
Wiener		
3×3	28.38	26.25
5×5	28.33	25.14
Wavelet	T=30	T=30
DWT db4	28.04	25.41
DWT db8	28.00	25.26
DWT coif5	28.13	25.45
Frame MSE$_{lim}$=170	T=30	T=30
$\mathbf{G}$=ELT	28.28	25.10
$\mathbf{G}$=db4	28.51	25.34
$\mathbf{G}$=db8	28.35	25.18
$\mathbf{G}$=coif5	28.52	25.39

Table 2. Best results form the different methods on noisy images with $\sigma^2 = 400$. T is soft thresholding level. MSE$_{lim}$ is local quality criterion for each block of coefficients, deciding local sparsity.

Fig. 1. Details from image Lena. Top line from the left; original and noise added ($\sigma^2 = 100$). Second line; Wiener with 5×5 mask and DWT coiflet5. Last line; frame with G=db4 and frame with G=coiflet5.

Fig. 2. Details from image Lena. Top line from the left; original and noise added ($\sigma^2 = 400$). Second line; Wiener with 5×5 mask and DWT coiflet5. Last line; frame with G=db4 and frame with G=coiflet5.

5. SUMMARY AND CONCLUSIONS

We have developed algorithms for designing signal dependent frames for various applications. A version of the algorithm including constraint to make it computationally attractive for two-dimensional signals is presented, and its usefulness for image denoising is demonstrated. The denoising experiments show that our overlapping frames have a good potential in this important application. Sharp edges and textures are preserved better than using adaptive Wiener filtering or DWT with soft thresholding. Unfortunately some artifacts are added in the smooth regions with both DWT and overlapping frames. How to reduce such undesired effects is a topic for further investigations.

6. REFERENCES

[1] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, pp. 227–234, Apr. 1995.

[2] S. F. Cotter, J. Adler, B. D. Rao, and K. Kreutz-Delgado, "Forward sequential algorithms for best basis selection," *IEE Proc. Vis. Image Signal Process*, vol. 146, no. 5, pp. 235–244, Oct. 1999.

[3] Al Bovik, *Handbook of Image and Video Processing*, Academic Press, San Diego, USA, 2000.

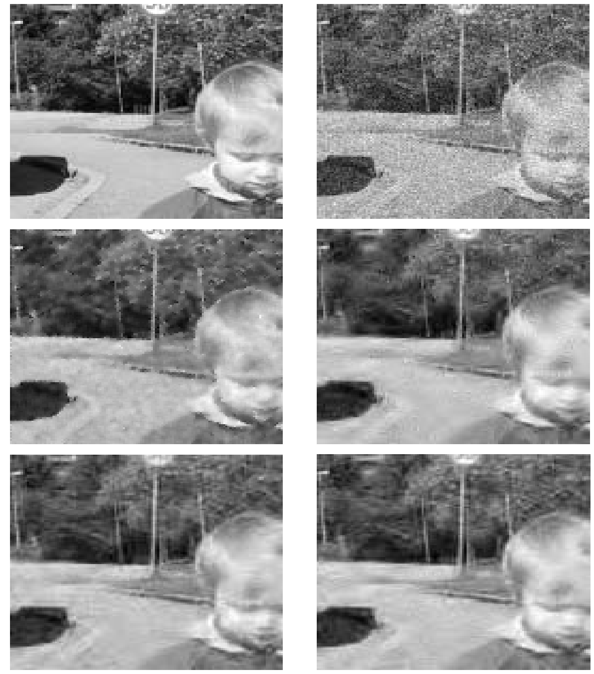

Fig. 3. Details from image Julie. Top line from the left; original and noise added ($\sigma^2 = 400$). Second line; Wiener with 3×3 mask and DWT coiflet5. Last line; frame with G=db4 and frame with G=coiflet5.

[4] D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no.3, pp. 613–627, May 1995.

[5] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, pp. 337–365, Feb. 2000.

[6] K. Kreutz-Delgado, B.D. Rao, K. Engan, T.-W. Lee, and T. Sejnowski, "Learning overcomplete dictionaries and sparse representations," *Neural Computation.*, 2000.

[7] K. Engan, S. O. Aase, and J. H. Husøy, "Multi-frame compression: Theory and design," *Signal Processing*, vol. 80, pp. 2121–2140, Oct. 2000.

[8] P. Schmid-Saugeon and A. Zakhor, "Dictionary design for matching pursuit and application to motion-compensated video coding," *IEEE Trans. Circuits, Syst. for Video Tech.*, vol. 14, no. 6, pp. 880–886, June 2004.

[9] Karl Skretting, *Sparse Signal Representation using Overlapping Frames*, Ph.D. thesis, Stavanger University College/NTNU, Stavanger, Norway, 2002, Available at http://www.ux.his.no/~karlsk/.

[10] K. Skretting, K. Engan, J. H. Husøy, and S. O. Aase, "Sparse representation of images using overlapping frames," in *Proc. SCIA'01*, Bergen, Norway, June 2001.

[11] H. S. Malvar and D. H. Staelin, "The LOT: Transform coding without blocking effects," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 4, pp. 553–559, Apr. 1989.

[12] K. Engan and K. Skretting, "A novel image denoising technique using overlapping frames," in *Proc. of 2nd. IASTED Int. Conf. on Visualization, Imaging, and Image Processing*, Malaga, Spain, Sept. 2002, pp. 265–270.