Appears in: Proc. IEEE Int'l. Conf. on Image Processing, San Antonio, TX, Oct 2007.

STATISTICALLY DRIVEN SPARSE IMAGE APPROXIMATION

Rosa M. Figueras i Ventura and Eero P. Simoncelli

Howard Hughes Medical Institute
Courant Institute of Mathematical Sciences / Center for Neural Science
New York University

ABSTRACT

Finding the sparsest approximation of an image as a sum of basis functions drawn from a redundant dictionary is an NP-hard problem. In the case of a dictionary whose elements form an overcomplete basis, a recently developed method, based on alternating thresholding and projection operations, provides an appealing approximate solution. When applied to images, this method produces sparser results and requires less computation than current alternative methods. Motivated by recent developments in statistical image modeling, we develop an enhancement of this method based on a locally adaptive threshold operation, and demonstrate that the enhanced algorithm is capable of finding sparser approximations with a decrease in computational complexity.

Index Terms— Sparse image approximation, overcomplete representation, redundant dictionary, image statistics.

1. INTRODUCTION

Many problems in image processing can be formulated in terms of approximation with a relatively small set of basis functions. Having such a compact representation of a signal can lead to better signal modeling, which in turn leads to improvements in a variety of applications such as coding, denoising, source separation, classification or segmentation. The classical approach to sparse approximation is based on statistical properties of the ensemble. Specifically, principal components analysis provides a linear basis sorted according to variance, and independent component analysis methods have been developed to exploit higher-order statistics [1].

More recently, a variety of authors have begun to explore the idea of signal-specific sparse approximation. Instead of trying to learn a basis that efficiently approximates a class of signals (either probabilistic or deterministic), these methods try to find the sparsest approximation of a *single image* given a large (typically redundant) set of basis functions. The dictionary of functions may include functions that capture some specific geometric properties of images, such as orientation and scale, local geometry, or local texture structures. Ideally, one would like to have a decomposition that is zero everywhere except where there are meaningful features.

For a highly redundant base, finding the sparsest approximation is NP-hard, but there exist several approximation methods that attempt to reach a solution in a more affordable way. The most well-known of these are Matching Pursuit (MP) [2] (a greedy algorithm which has many cousins), and Basis Pursuit Denoising (BPDN) [3]. More recently, a new method known as Noise Shaping has been developed, based on sequentially alternating projections onto the transform space and the image space [4, 5, 6] This method converges (albeit to a local minimum), and appears to provide sparser image approximations than Basis Pursuit or Matching Pursuit [5] when applied to redundant bases that have reasonably efficient forward and inverse transforms (the efficiency of the algorithm is directly proportional to the efficiency of the transforms). Nevertheless, direct examination of the solutions suggests that they can be improved by careful consideration of the properties of images. In this paper, we develop a modified version of this method that incorporates some information about image statistics. We show through simulations that this method can outperform the original Noise Shaping method, in terms of image quality at a given sparsity level.

This paper is structured as follows: Sec. 2 gives an overview of the original Noise Shaping algorithm. Sec. 3 presents its locally adapted version, introducing image statistics in the algorithm. Sec. 4 presents some approximation results. Finally, 5 draws some conclusions.

2. NOISE SHAPING ALGORITHM

The Noise Shaping algorithm [4] iteratively computes a sparse decomposition of a signal as depicted in Fig. 1. The signal is first projected onto the full set of basis functions (represented by transformation T). In the transform domain, the coefficients are compared with a global threshold value, θ, and those that are less than this value are discarded. This threshold is chosen to achieve a desired level of sparsity (i.e., a desired number of coefficients). Next, the approximation error for this thresholded set of coefficients is computed, by applying the inverse transform T^{-1} to the thresholded coefficients and subtracting the resulting signal from the original image. This approximation error is again transformed using T, and these transform coefficients of the error are added to

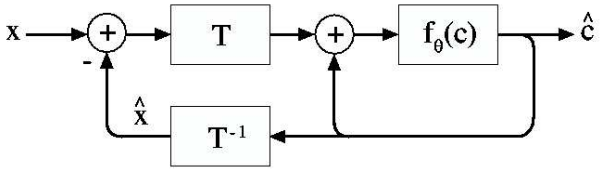


Fig. 1. Scheme of the Noise Shaping algorithm. Note that $\hat{\mathbf{c}}$ must be initialized to zero.

the thresholded ones. The resulting coefficients are guaranteed to provide an exact representation of the signal, since

$$\begin{aligned} & T^{-1}\left(T\left(X - T^{-1} f_\theta\left(T(X)\right)\right) + f_\theta\left(T(X)\right)\right) \\ = \;& X - T^{-1} f_\theta\left(T(X)\right) + T^{-1} f_\theta\left(T(X)\right) \\ = \;& X, \end{aligned}$$

where f_θ is the thresholding operator. This operation is performed iteratively until stability is reached (i.e., when adding the coefficients of the error and thresholding does not change the approximation obtained). The sparse representation corresponds to the thresholded coefficients. We refer to this version of the Noise Shaping algorithm, based on a Global Threshold (GT), as *one shot GT*. This algorithm guarantees convergence to a local minimum, but its final solution is highly dependent on the initialization parameters [4, 5].

It is possible to increase the performance of the algorithm by changing its initialization [4, 5]. In particular, one can start with a more highly sparsified representation, obtained by running the algorithm with higher threshold settings. More generally, the algorithm can be run in successive passes, adjusting the threshold settings on each step so as to increase the number of coefficients until the desired sparsity is achieved. We refer to this as the *evolutionary GT* algorithm.

The evolutionary algorithm provides better approximation performances than the one shot version, at the cost of substantial additional computation. If the algorithm takes K passes to achieve convergence, and every step costs an amount of time τ, the one shot algorithm takes only a time $K\tau$ to achieve the solution, while the evolutionary algorithm takes $SK\tau$, with S the number of steps taken to achieve the desired number of coefficients. Ideally, one would like to have a better way to achieve the solution, reducing either the number of steps S and/or the number of passes K, while preserving or improving the quality.

3. LOCALLY ADAPTIVE NOISE SHAPING

The Noise Shaping algorithm guarantees convergence to a local minimum [4, 5]. It is a well known fact that images contain localized structures, and that large-amplitude transform coefficients tend to cluster around such structures. This behavior can be modelled with a local Gaussian process whose standard deviation is spatially varying [7, 8]. If a Global Threshold (GT) is used to select a set of coefficients, they will all be chosen from regions containing high-contrast features, and medium- and low-contrast features will be ignored, decreasing the visual quality of the representation. This suggests that the threshold should be modulated by an estimate of the local energy or variance of the coefficients, as has been done in denoising [9]. Based on previous work [10, 11], we compute this estimate for a given coefficient i using a weighted ℓ^2-norm:

$$\Theta(i) = \beta \cdot \sqrt{\sum_{j \in \mathcal{N}(i)} w_j \cdot c_j^2 + \alpha}, \qquad (1)$$

with $\mathcal{N}(i)$ a generalized neighborhood containing coefficients surrounding the ith coefficient c_i, and w_j a weight vector over the neighborhood. The threshold may be efficiently computed using convolution operations in the transform domain. The parameter β controls the number of coefficients that remain after thresholding.

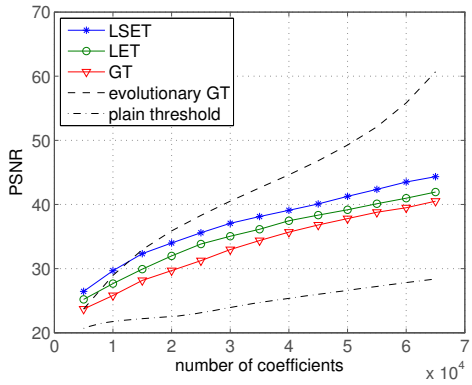
We refer to this modified algorithm, with the global threshold replaced by this locally adapted threshold, as the *Local Energy Threshold (LET)*. As with the original GT algorithm, this method may be used in its one-shot or evolutionary form. The threshold mask is computed once for the one shot algorithm, and once (based on the error-corrected coefficients) for every step of the evolutionary algorithm.

The LET algorithm can be made more effective by taking into account only those coefficients that have been retained after thresholding. The computation of $\Theta(i)$ follows exactly Eq. 1, but substituting c_j by $\hat{c}_j$. The first iteration of this computation considers all $\hat{c}_j$ to be zero, and is thus exactly equal to the GT algorithm. The threshold mask is re-computed on every iteration, to incorporate the newly chosen coefficients. We refer to this version of the algorithm as *Local Sparsified Energy Threshold (LSET)* algorithm. Again, it is possible to use an evolutionary LSET or a one shot LSET.

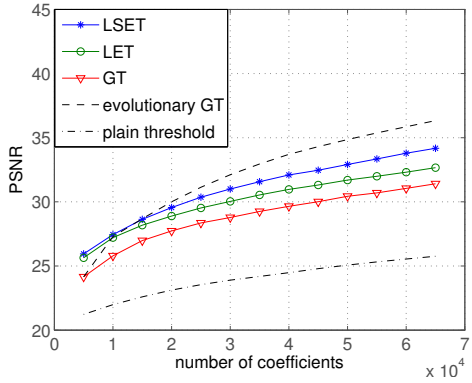
4. RESULTS

This section presents comparisons of the one shot and the evolutionary LET and LSET algorithms with the GT algorithm. A comparison of the evolutionary GT algorithm with other sparse approximation techniques (Basis pursuit, Matching pursuit) is provided in [5], where it is shown that it outperforms the other methods.

For a transform, T, we use the steerable pyramid transform [12], a redundant multiscale image representation, with oriented basis functions related by translation, rotation and dilation, that forms a tight frame. We expect the results to be similar for any multiscale oriented image basis of similar overcompleteness. For the steerable pyramid, the redundancy is controlled by changing the number of orientations used. The results presented in this paper are obtained with a steerable

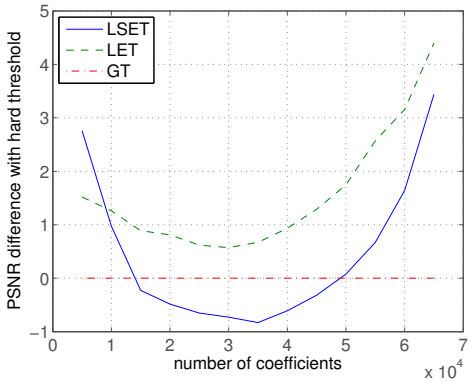


(a) Barbara 256x256

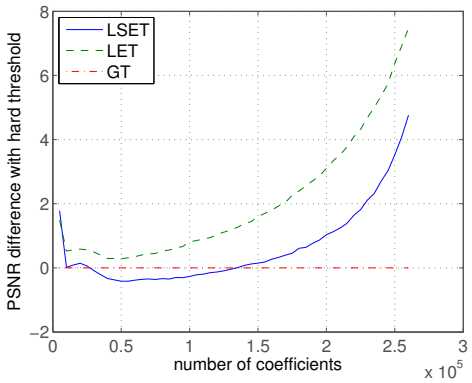


(b) Boats 512x512

Fig. 2. Comparison of the performance for the one shot GT, LET and LSET algorithms.



(a) Barbara (256x256)



(b) Boats (512x512)

Fig. 3. Comparison of the performance of the evolutionary GT, evolutionary LET and evolutionary LSET algorithms.

pyramid with 8 orientations and 5 scales, leading to a more than six-fold redundancy factor. We found that modest increments or decrements in the number of orientations or scales did not lead to noticeable changes in the relative behavior of the sparsification algorithms. LET and LSET results compute the energy mask using a simple neighborhood consisting of a 7×7 patch of spatial neighours drawn from all 8 pyramid orientations (total neighborhood size is thus 392). For the results presented here, w_i are uniform weights. The results could possibly be improved by taking into account specificities of the signal and the dictionary into the weighting.

Fig. 2 shows the convergence obtained with the one shot algorithms for the *Barbara* image (256×256) and the *Boats* (512×512), compared with the performance of the *evolutionary* GT algorithm. The difference in quality among the methods is substantial. The LSET algorithm is noticeably better than the LET algorithm, which is substantially better than the one shot GT algorithm. On the other hand, the evolutionary GT algorithm shows substantial improvements over all one shot methods for significant numbers of coefficients. This comes at a substantial computational cost, proportional to the number of steps needed in the evolutionary algorithm. Figure 4 shows a visual comparison of the three different one shot algorithms at 5000 coefficients. It can clearly be seen that the use of the local energy of the coefficients to modulate the thresholding operation improves the visual quality of the resulting image. Notice, for example, the texture details that appear on the shawl around Barbara's face in the image obtained with the LSET algorithm that are completely absent in the GT image. The PSNR value is also substantially better. The LET algorithm result is nearly as good as the LSET visually, but noticeably inferior in terms of PSNR.

Fig. 3 shows the performance of the evolutionary LET and LSET algorithms compared to the performance of the evolutionary GT algorithm for the same images as in Fig. 2. For these results, the evolution is done in steps of 5000 coefficients at a time (for example, a solution with 20000 coefficients is achieved in 4 steps, of 5000, 10000, 15000 and finally 20000 coefficients). Decreasing the step size (the jump on number of coefficients performed at every iteration) increases the performance of all algorithms by roughly equal amounts, at a substantial cost in computation time. The graphs of Fig. 3 show the relative approximation performance for different number of coefficients, starting at 5000 coefficients and finishing at the number of coefficients equivalent to the image

(a) GT method, PSNR=23.69dB (b) LET method, PSNR=25.18dB (c) LSET method, PSNR=26.45dB

Fig. 4. Visual comparison of the one shot approximation of Barbara (256x256) with 5000 coefficients.

dimension. All curves are subtracted from that of the evolutionary GT algorithm, whose performance thus corresponds to a horizontal line at zero. The performance of the LET algorithm is always superior to the performance of the GT algorithm, and this is especially noticeable for the low and high rates. Curiously, the performance of the LSET algorithm is generally inferior to that of the LET algorithm, and it is even inferior to the GT algorithm for intermediate rates. We also found that the LET and especially the LSET algorithm require fewer iterations to achieve stability than the GT algorithm: The increase of quality that can be seen in Fig. 3 is not due to an increase of computational time, but to a more efficient and effective algorithm.

5. CONCLUSIONS

We have presented a sparsification algorithm that can be applied to any redundant representations for which the inverse transform is easily computed. The algorithm is based on the recent Noise Shaping method [4], and the improvement comes from the choice of the nonlinear operation used to reduce the number of active coefficients. In [4, 5], the non-linearity is a global thresholding of the coefficients. Here, we use an adaptive threshold based on an estimate of the local variance of the coefficients. This simple modification improves the convergence of the algorithm both in number of iterations that are needed to achieve stability, and the achieved image quality (for both PSNR and visual appearance). The use of more complex window shapes, different size neighborhoods, and neighbors at other scales could enhance the adaptation of the dictionary to the statistical properties of images, and could lead to improvements over the results presented here.

6. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[3] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[4] N. G. Kingsbury and T. H. Reeves, "Redundant representation with complex wavelets: how to achieve sparsity," in *ICIP*, 2003, vol. 1, pp. 45–48.

[5] L. Mancera and J. Portilla, "L0-norm-based sparse representation through alternate projections," in *ICIP*, 2006, pp. 2089–2092.

[6] M. Elad, "Shrinkage from redundant representations," in *Workshop on Sig. Proc. with Adaptative Sparse Structured Representations (Spars05)*, 2005.

[7] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Adv. Neural Information Proc. Systems (NIPS*99)*. May 2000, vol. 12, pp. 855–861, MIT Press.

[8] J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden markov models," *IEEE Trans. on Image Proc.*, vol. 10, no. 7, pp. 1056–1068, 2001.

[9] G. S. Chang and M. Vetterli, "Spatial adaptive wavelet thresholding for image denoising," in *ICIP*, 1997, vol. 2, pp. 374–7.

[10] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. on Image Proc.*, vol. 8, no. 12, pp. 1688–1701, December 1999.

[11] S. Lyu and E. P. Simoncelli, "Statistically and perceptually motivated nonlinear image representation," in *Proc. SPIE, Conf. on Human Vision and Electr. Imag. XII*, 2007, vol. 6492.

[12] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *ICIP*, 1995, vol. 3, pp. 444–447.