Restoring Image Quality Through Structure Preserving De-noising

Krishna Ratakonda and Narendra Ahuja
Department of Electrical and Computer Engineering
Beckman Institute, University of Illinois, Urbana, IL 61801.
krishna@stereo.ai.uiuc.edu

Abstract

In many image transmission and acquisition situations, the image may become corrupted by additive noise. De-noising refers to the process of removing the noise while maintaining good visual quality. This problem has assumed major significance with the increase in image related communication that has accompanied the exponential growth of the internet. Traditionally, image quality is measured in terms of PSNR (Peak Signal to Noise Ratio) which may have limited relation, at best, to the perceptual quality of the image. In this paper we present a novel de-noising scheme which results in significantly improved performance in terms of both perceptual quality and PSNR. Furthermore, we show that the de-noising framework that we propose encompasses the usual linear transform based de-noising schemes as special cases.

1 Introduction

Classical methods for de-noising images treat the problem no differently from its one-dimensional counterpart. There is no explicit way of preserving important structural information which is crucial for interpretation, recognition and other computer vision tasks. In fact most of the usual de-noising schemes perform poorly with respect to the preservation of structure. This paper proposes an approach to de-noising that improves PSNR while preserving structure. This is accomplished by employing structure sensitive constraints to guide the usual PSNR-enhancing de-noising schemes. Instead of making the usual assumption that the image is a smooth signal, we view the image as a multi-scale, piecewise-smooth, 2-D signal and use the PSNR based approach on each of these smooth parts defined by a multi-scale image segmentation. Clearly, the fidelity of image segmentation used would have a direct impact on the performance of the algorithm. To this end we use a multi-scale image segmentation algorithm that extracts image regions, which are in agreement with what is percieved, in the presence of significant geometric, topological and spectral complexity, and regardless of the number and parameters of the natural scales present in a given image. Since the image varies smoothly over each segmented region, the PSNR also improves significantly compared to that achieved by

* This research was supported by the Advanced Research Projects Agency under grant N00014-93-1-1167 administered by the Office of Naval Research and the NSF grant IRI 93-19038.

the same de-noising method applied globally over the entire image. Thus, the net effect is the integration of high-fidelity image segmentation and better than usual PSNR gains in each segment, leading to an improvement in perceptual quality in both optical and spectral senses.

The model for image estimation in the presence of additive noise may be written as $Y_{ij} = X_{ij} + \epsilon_{ij}$, where i, j are indices into the image represented as a matrix, Y is the noisy signal, X is the uncorrupted signal and ϵ represents the additive noise.

Many of the previously proposed de-noising methods with good performance use an orthonormal transform such as a Wavelet or a Fourier transform to map the noisy image Y into a domain where the energy of X is compactly represented. Note that we may consider the contributions due to X and ϵ indepedently due to the linearity of the transform. The signal compaction properties of the transforms ensure that most of the energy of X is concentrated into a few coefficients in the transform domain. Since the energy of X is not changed by an orthonormal linear transform (Parseval's relation) these few coefficients, which recieve significant contribution from X, have relatively large values.

If the noise ϵ is assumed to be identically and independently distributed, additive and white Gaussian (i.i.d. AWGN), the contribution to the transform coefficients due to noise is still distributed across the transform domain and their values retain the standard deviation characterizing the noise in the image domain. The contribution due to noise in each coefficient is small but effects all the coefficients. This distribution of the noise energy may also be explained by the fact that an orthonormal linear transform only results in a rotation of the coordinate system representing the image. Hence, the probability distributions of the noise coefficients, which are symmetric with respect to axis rotation due to i.i.d. assumption, are not affected by such a coordinate transformation.

Thus in the transform domain we get a representation of the noisy image Y in which the actual image X is typically concentrated in a few large values and the noise contributes small values to all transform coefficients. This discussion would suggest that applying a threshold on the transform coefficients and setting to zero all values below the threshold might be a good de-noising scheme. Donoho and Johnstone [1] analyzed the scenario under the i.i.d. AWGN assumption[1] and derived the optimal threshold to be applied when a linear, orthonormal transform is used. It is well known that linear transforms (including Wavelets) result in a distribution of the energy contributed by the edge areas in the image all over the transform domain in contrast to the energy contributed by the non-edge areas in the image which is localized to low frequencies. Since the transitions across edges contribute mainly to the small values, which are removed by the thresholding operation, we obtain distortion in edge areas.

Liu and Moulin [2] propose a more complicated approach which uses a lossy coder to code the image; the rate-distortion factor of the lossy coder being determined adaptively depending on the noise level corrupting the image. They use a Wavelet transform based lossy coder [3], which benifits mainly from the different representations of the noise and the signal in the Wavelet domain. The basic assumption is that the loss incurred in the coder is mostly due to the removal of noise which is valid since i.i.d. Gaussian noise is hard to compress. Their method works well for high SNRs and produces fewer artifacts than the simpler thresholding scheme of Donoho [1]. However edge distortion invariably occurs at low SNRs since the lossy coder does not perform as well for the i

[1]It may be noted that the AWGN assumption can be readily extended in a robust way to other probability distributions provided that the i.i.d. assumption is satisfied.

rate-distortion factors corresponding to low SNRs.

We introduce the new framework for structure mediated de-noising for i.i.d. AWGN in the next section and outline the key new ideas proposed. Our framework incorporates two major but independent components which are discussed in sections 3 and 4. Implementation details and a few illustrative results are presented in section 5.

2 Proposed Framework for De-noising

The key idea behind the proposed approach comes from the discussion in the previous section on the rate-distortion (Liu-Moulin) and hard thresholding (Donoho-Johnstone) schemes. Both these schemes inherently depend on the fact that the non-noisy signal is *compressible and hence compactly representable* in the Wavelet domain. The important point to note is that these schemes can be applied *independently on different portions of the image*, provided that the pixel population within each region is sufficiently large (we will define this more concretely later) to support the asymptotic assumptions involved. Thus we might think of the following processing: (a) split the image into disjoint sets of pixels and (b) apply the usual de-noising scheme over each set independently.

Provided that we split the image into disjoint collections of pixels, each collection characterized by grey value homogeniety, the Wavelet transform would be able to compress each of the regions individually to a high degree. This follows because the Wavelet transform (and in fact any other compacting linear transform) has difficulty in compacting signals with both smooth as well as high frequency components [4]. By segmenting the regions, say into regions characterized by gray level similarity, and choosing an individual region over which to apply the transform, we *effectively eliminate most of the high frequency coefficients* which cause loss in compaction, resulting in high compaction. This maximizes the performance of the schemes that we have discussed in Section 1.

A segmentation scheme which would robustly segment an image in the presence of Gaussian noise is needed. The presence of Gaussian noise is not likely to cause much degradation in the segmentation scheme relative to the case of the noise free image since: (a) segmentation involves pixel population analysis in order to find regions and (b) the noise is assumed to be i.i.d. and hence does not superimpose significant structure on the image. Recently, a new multi-scale segmentation scheme has been proposed that extracts image regions at all natural geometric and photometric scales present. At each scale, the regions extracted are consistent with perception for a wide range of the geometric, topological and spectral complexities [5]. We use this algorithm, in our present work, for segmentation.

The idea of exploiting the advantages of the Wavelet (and other linear) transforms over regions immediately leads to another problem: regions have arbitrary shapes and the Wavelets and other linear transforms cannot be applied over a region of arbitrary support. These transforms are defined only over square (or rectangular) supports. In fact, Wavelets may be applied, without computational tricks, only to supports which are powers of two. We propose a solution which not only extends the definition of the Wavelet transform (and any other discrete linear transform) to arbitrary supports but also leads to maximal compaction. Section 4 and a previous paper [6] provide the details.

To summarize, the framework that we propose involves three key steps:

1. Segment the image to obtain a multi-scale partition of the image into homogeneous regions and post process the results (section 3).

2. Convert each region into a form so that the linear transform can be applied (section 4).
3. Apply the transform based processing over each region individually, invert the linear transform and combine the results from different regions.

It is to be noted that the only restriction on the baseline de-noising scheme is that it is linear transform based. Most of the schemes which perform well in practice belong to this category. Furthermore, by considering the entire image as a single region we get back to the usual version of the baseline de-noising scheme. Thus, our framework includes as special cases all baseline de-noising schemes which are based on signal compaction in linear transform domains.

In the above discussion we have justified why we expect to obtain an advantage over the base-line processing scheme using our framework. Results show significant improvement in visual quality due to both PSNR increase and better structure preservation. The proposed framework can be viewed as a multi-representational paradigm which combines the advantages of linear schemes which generally work well with smooth signals, and the structural information provided by multi-scale segmentation which complements the strengths of the linear methods. Thus we benefit from the complementarity of representations.

3 Segmentation

In this paper, we will be using the segmentation information provided by the multi-scale segmentation transform described in [5]. As explained in the previous section, the performance of the segmentation algorithm is not expected to degrade very much due to i.i.d. AWGN. However, since we will be considering low SNRs, there might occur spurious structures in the image introduced by the noise which will be (correctly) detected by the low level segmentation method. Such noise artifacts will tend to be small and therefore could be removed by postprocessing the result of segmentation. Another reason for post processing is that the population of the pixels in each region should be above the minimum required to validate the asymptotic assumptions behind de-noising algorithms. Since the segmentation provided by the algorithm in [5] detects regions at all spatial scales, including those regions smaller than can be used, the small regions must be replaced by their sufficiently large parent regions at coarser scales, or if such a parent region cannot be found merged with a suitable neighbor, before applying the transform-based de-noising algorithms.

3.1 Post processing of segmentation data

We consider regions at the finest photometric scale for de-noising independently. For those regions that are too small, the path to the root of the segmentation tree is searched to find the nearest parent region having sufficiently large size. If no such region is found, that is if the finest scale region does not merge with any other region, then the region is deleted. This helps ensure that the asymptotic assumptions are followed and any spurious regions introduced by noise are disregarded.

To arrive at the size threshold, let X_i, $i = 1 \cdots n$ be a sequence of i.i.d. random variables. As n tends to ∞, the value of $\sum_{i=1}^{n}(X_i/n)$ approaches the expected value of the individual random variables. Under the assumption that the noise is zero mean, we formulate the following criterion to specify an acceptable level of the validity of the asymptotic results $E(|\sum_{i=1}^{n}(X_i/n))|^2) < 1$.

Since the expectation above decreases with increasing n, we obtain a threshold for the minimum size of the region. With Gaussian noise of standard deviation σ, we will thus need the number of pixels in a region is at least σ^2. This assumption will ensure that if we replace all the pixels within a region with the average value of the region (A), A will be a random variable with standard deviation smaller than 1 and mean equal to the the average of the same pixels in the uncorrupted image. Normally one would expect that artifacts would appear at region edges since we process each region independently of the others. The minimum size requirement on regions is intended to reduce this artifact by ensuring that the region mean in the noisy image is not very different from the region mean in the uncorrupted image.

Given the above threshold for the minimum size of a region, we obtain a simple algorithm. We apply the linear transform based de-noising to all the regions at finest scale that have a size larger than the threshold. For any region at the finest scale with size below the threshold, we perform the following steps:

1. If the region gets merged with another region at a coarser scale in the segmentation tree, since the regions at coarser scales are preserved at finer scale, find the smallest acceptable region along the path to the root and use it to replace all descendent leaves of the region.

2. If the region at a fine scale persists in all coarser scales of segmentation, merge it with the neighboring region which has minimum disparity in mean gray value from the current region.

Figure 1 shows the segmentation before and after post processing for the image Lenna. The output of the segmentation algorithm shows a number of small regions which are generated by isolated large values of noise. These regions are removed from the segmentation after post processing.

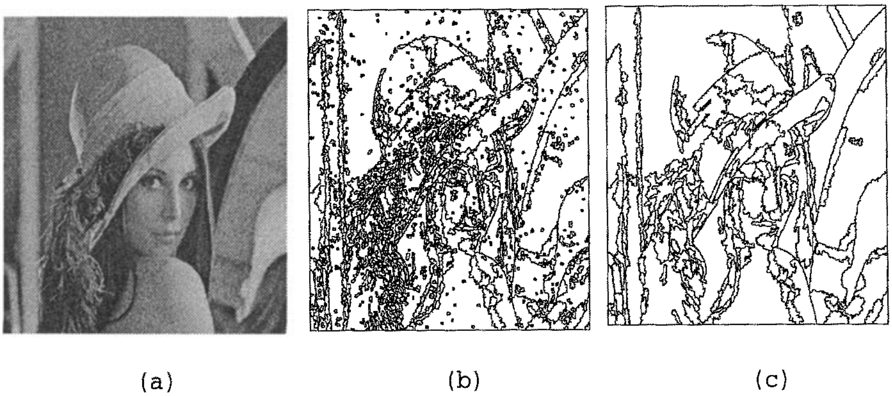


Figure 1: (a) Lenna corrupted by Gaussian noise of s.d. 15 (b) Segmentation at the finest scale and (c) Segmentation after post processing using multi-scale segmentation data.

4 Linear Transforms over Arbitrary Supports

Discrete linear transforms in two dimensions are in most cases defined over a rectangular support. The usual practice when we want to apply the transform over an arbitrarily shaped support is to fill out the rest of the support with zeros to make up the rectangle and then use the natural definition of the transform over a rectangle. This is an extension of the one dimensional case where we fill out an arbitrary length data set with zeros to form a data set of length 2^n either to increase the computational speed (through FFTs for Fourier Transforms) or to satisfy the definition of the transform (in the case of dyadic Wavelets). This, however, does not lead to a satisfactory definition of the linear transform in two or more dimensions for many applications.

The above discussion leads us to the following question: what should be the values attributed to the sample points which lie within the rectangle but not within the support of the function? With each possible choice of the values for the pixels which lie outside the support but within the rectangular region, we can associate a possible function-transform pair. We would like to define the free pixels (the pixels within the rectangular region but outside the support) so as to minimize the high frequency content in the transform domain. Thus we may view this as *minimizing a cost function which maximally shifts the energy in the low frequency values over all possible choices of the free pixels.* The choice of the free pixels should distribute the energy due to region boundary, as well as the energy within the rectangle but outside the region, to as low frequency values as possible.

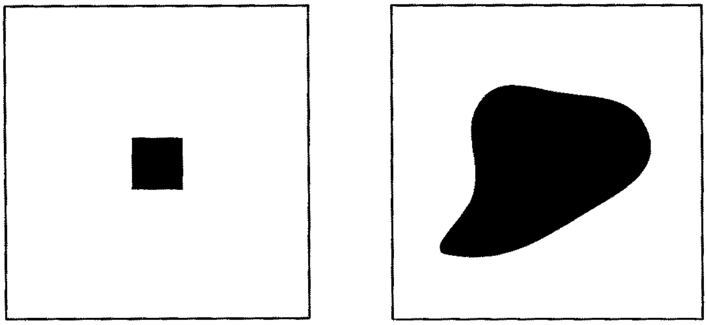
In section 4.1, we provide a re-definition of the discrete linear transform that packs the energy into the low frequency areas. The solution in constructed within the projection on convex sets formalism. Since all the constraint sets are convex, this framework guarantees convergence to a solution. This kind of formulation has been treated by us in a more general context in [6]. If speed is an issue, it is to be noted that this algorithm may be reformulated to achieve quadratic convergence [7].

4.1 Re-defining the linear transform

In this section we will define constraints that the "free pixels" should satisfy, inorder that we may pack most of the energy in the low frequency areas of the image. These constraints may be different depending on the meaning attributed to the term "high frequency". For this purpose, we define ζ as the side of a square shaped region containing the low frequency components in the transform domain (figure 2(a)). All components which lie outside the square region are high frequency components and are to be minimized. Note that we could avoid ζ by defining our cost criterion to be the minimization of the first (or second order moment) of the coefficients in the transform domain. This kind of cost criterion imposes larger penalties on higher frequency terms, thus resulting in packing of the energy in the low frequencies. However, we found the improvement in performance not significant enough to justify the increased complexity. Choice of a value for ζ will be discussed later.

Now we have obtained two constraints that we would like the signal to satisfy:

1. In the transform domain the image should have minimum energy outside the square region of side ζ as defined above. That is all components outside the square should be as close to zero as possible in the L^2 norm sense.
2. In the spatial (or image) domain, the image should satisfy the known values of the function within its support.



(a) (b)

Figure 2: Constraints for the DFT. (a) DFT domain: Values outside the shaded square should be minimized (in L^2 sense). (b) Spatial (or image) domain: Known values of the function should not be disturbed.

These two constraints are illustrated in figure 2. That they may be put in terms of convex sets and a solution obtained using the Projection on Convex Sets algorithm is a straight forward exercise. The reader is refered to [8, 6] for more details.

What should be the value of ζ? We found that the results that we obtain (for this application in de-noising) do not vary much with ζ provided that we choose ζ not too conservatively. If we choose the value of ζ too conservatively, we might end up with too little variation in the free pixels and hence might not obtain optimal packing of the energy in the low frequencies. Since each region may be enclosed in squares of different sizes, we chose ζ to be proportional to the side length of the square on which the transform is being applied. As mentioned earlier, a less adhoc solution would be to avoid the choice of ζ altogether, and maximize the energy at low frequencies.

5 Implementation and Results

We use the Donoho-Johnstone hard thresholding criterion as our baseline scheme. As described in section 1, the optimal performance of this criterion requires maximal packing of the image energy in as few coefficients in the transform domain as possible. Since each of the individual regions processed with the baseline scheme already approximately satisfies this requirement for the optimal performance of the Donoho-Johnstone criterion, we expect only marginal changes in performance if we apply more complicated processing. From [1], we obtain the following threshold to be applied on the Wavelet coefficients for Gaussian noise of s.d. σ: $(Threshold)^2 = \sigma^2(2ln(N))$, where N is the number of pixels within a particular region. We use the Daubechies 4-tap Wavelet filters inorder to perform the simulations. The POCS algorithm for calculating the optimal transform over the arbitrary shaped region converges within 2-3 iterations to within 10% of its final value on the average. As was noted before, convergence can be speeded up even more by formulating the optimization problem in the formalism provided in [7].

Illustrative results are shown for the image Goldhill. Results for the Goldhill image for a Gaussian noise of standard deviation 15 are shown in figure 3. In figure 3(c), the edges are clearly better preserved than figure 3(b). This may be seen from the bars on

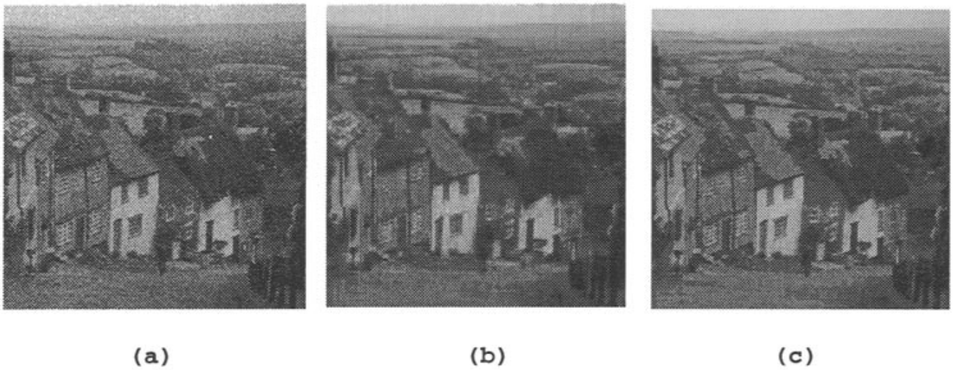


Figure 3: (a) Goldhill corrupted by i.i.d. AWGN of s.d. 15 (PSNR 24.6 dB) (b) Goldhill processed with Donoho-Johnstone criterion (PSNR 29.61 dB) (c) Goldhill processed with proposed approach (PSNR 31.70 dB)

the windows, the outline of the roofs and other portions of the image. Furthermore, the streaks that are found in figure 3(b) disappear in figure 3(c).

The results show that our algorithm preserves the structure as well as improves PSNR. Although results have been presented using the Donoho-Johnstone hard thresholding criterion as its baseline scheme, other transform based approaches could be easily used to de-noise the individual regions [2, 9].

References

[1] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by Wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[2] Juan Liu and P. Moulin. Denoising of images using a complexity regularization criterion. *To be published in ICIP-97.*

[3] S. Servetto, K. Ramchandran, and M. T. Orchard. Morphological representation of wavelet data for image coding. *Proceedings ICASSP*, 1995.

[4] C. Herley, K. Ramchandran, et al. Time-varying orthonormal tilings of the time-frequency plane. *ICASSP*, 2:891–5, 1993.

[5] N. Ahuja. A transform for the detection of multi-scale structure. *IEEE Transactions PAMI*, December 1996.

[6] K. Ratakonda and N. Ahuja. Discrete multi-dimensional linear transforms over arbitrarily shaped supports. *To be published in ICASSP*, 1997.

[7] S. S. Dharanipragada and K. S. Arun. A quadratically convergent algorithm for convex set constrained signal recovery. *IEEE Transactions on Signal Processing*, 44(2):248–66, February 1995.

[8] H. H. Chen et al. A block transform coder for arbitrarily shaped image segments. *IEEE International Conference on Image Processing*, 1:85–89, 1994.

[9] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:468–478, 1978.