A WAVELET-BASED STATISTICAL MODEL FOR IMAGE RESTORATION

Yi Wan and Robert D. Nowak

Rice University
Department of Electrical Engineering
6100 South Main Street, Houston, TX 77005, USA

ABSTRACT

In this paper we develop a wavelet-based statistical method for solving the image restoration problem. In this approach, a signal prior is set up by modeling the image wavelet coefficients as independent Gaussian mixture random variables. We first specify a uniform (non-informative) distribution on the mixing parameters, which leads to a simple and efficient iterative algorithm for MAP estimation. This algorithm is similar to the EM algorithm in that it alternates between a state estimation step and a maximization step. Moreover, we show that our algorithm converges monotonically to a local maximum of the posterior distribution. We next generalize the result to non-uniform priors and develop an efficient integer programming algorithm that enables a similar alternating optimization procedure.

1. INTRODUCTION

Image restoration belongs to the class of problems named as linear inverse problems, which are common in areas such as medical diagnostics, radar and sonar target estimation, seismology, radio astronomy, microscopy [1, 2, 3]. Such problems are mathematically described as

$$\mathbf{y} = \mathcal{L}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x}$ is the original signal or image of interest, $\mathcal{L}$ is a linear operator, $\mathbf{n}$ is the additive noise term, and $\mathbf{y}$ is the observed data. The goal of the inverse problem is to recover $\mathbf{x}$ from $\mathbf{y}$; *i.e.,* to find an operator $\mathcal{H}$ (not necessarily linear) such that the estimate

$$\widehat{\mathbf{x}} = \mathcal{H}\mathbf{y} \tag{2}$$

is as close to $\mathbf{x}$ as possible according to some prescribed measure.

In practice the operator $\mathcal{L}$ is often near-singular, making such inverse problems ill-posed. Many classical methods and some more recent methods have been proposed to solve such problems, including several multiscale and wavelet-based approaches [1, 4, 5, 6, 7]. In this paper we present an alternative wavelet-based Bayesian approach to linear inverse problems. Our approach was first described in [8]. Here we investigate its performance and theoretical properties in greater detail and develop an extension of it. The method makes use of the wavelet domain statistical model proposed in [8, 9], and we develop a joint MAP estimator through a clean, direct and mathematically sound development. The suitable choice of Gaussian mixture model for the wavelet coefficients and an alternating maximization method leads to a very simple and efficient iterative algorithm.

This work was supported by the National Science Foundation, grant no. MIP–9701692, the Army Research Office, grant no. DAAD19-99-1-0290, and the Office of Naval Research, grant no. N00014-00-1-0390.

2. BASIC FORMULATION

It is widely accepted that the wavelet coefficients of natural images tend to be decorrelated. In addition, the distribution of these coefficients usually has a peaky and heavy-tailed symmetric shape, centered at the origin [9]. This motivates the use of prior densities based on independent Gaussian mixture models (IGMMs) for the wavelet coefficients. Let

$$\boldsymbol{\theta} = \mathcal{W}\mathbf{x}$$

where $\mathcal{W}$ is the wavelet transform operator, then (1), formulated in terms of wavelet coefficients $\boldsymbol{\theta}$, becomes

$$\begin{aligned} \mathbf{y} &= \mathcal{L}\mathcal{W}^{-1}\boldsymbol{\theta} + \mathbf{n} \\ &= \widetilde{\mathcal{L}}\boldsymbol{\theta} + \mathbf{n}. \end{aligned} \tag{3}$$

We model each wavelet coefficient as a Gaussian two-mixture random variable,

$$p(\theta_i) = (1 - q_i)\mathcal{N}(0, \sigma_0^2) + q_i\mathcal{N}(0, \sigma_1^2), \quad \sigma_0^2 \ll \sigma_1^2 \tag{4}$$

where $\mathcal{N}(\mu, \sigma^2)$ represents a Gaussian distribution with mean μ and variance σ^2.

Let $\mathbf{q} = \{q_i\}$, then the joint MAP estimator of $\boldsymbol{\theta}$ *and* $\mathbf{q}$

is

$$
\begin{aligned}
(\widehat{\boldsymbol{\theta}},\ \widehat{\mathbf{q}}) &= \underset{\boldsymbol{\theta},\ \mathbf{q}}{\operatorname{argmax}}\ \ p(\boldsymbol{\theta}, \mathbf{q} \mid \mathbf{y}) \\
&= \underset{\boldsymbol{\theta},\ \mathbf{q}}{\operatorname{argmax}}\ \ p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{q})p(\mathbf{q}). \qquad (5)
\end{aligned}
$$

For simplicity let's first set $p(\mathbf{q}) = p(q_i) = 1$. In this case,

$$
\begin{aligned}
L(\boldsymbol{\theta}, \mathbf{q}) &= \log p(\boldsymbol{\theta}, \mathbf{q}|\mathbf{y}) \\
&= -\frac{1}{2}(\mathbf{y} - \widetilde{\mathcal{L}}\boldsymbol{\theta})^T \Sigma^{-1} (\mathbf{y} - \widetilde{\mathcal{L}}\boldsymbol{\theta}) + \\
& \sum_{i=1}^{M} \log \left[\frac{1-q_i}{\sigma_0} e^{-\frac{\theta_i^2}{2\sigma_0^2}} + \frac{q_i}{\sigma_1} e^{-\frac{\theta_i^2}{2\sigma_1^2}} \right] + C_1 \qquad (6)
\end{aligned}
$$

where C_1 is a constant.

Usually the maximization of L is difficult because of the presence of mixture densities. However, the following lemma provides a necessary condition on any local maximal solution and will later on lead to an iterative algorithm that monotonically converges to a local maximum.

Lemma 1 *Let $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{q}})$ be any local maximum of $L(\boldsymbol{\theta}, \mathbf{q})$, i.e.,*

$$L(\boldsymbol{\theta}, \mathbf{q}) \leq L(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{q}})$$

for all $(\boldsymbol{\theta}, \mathbf{q})$ in a sufficiently small neighborhood of $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{q}})$. Then

1. *Each element $\widehat{q}_i$ of $\widehat{\mathbf{q}}$ is equal to either 0 or 1, and is given by*

$$
\widehat{q}_i = \begin{cases} 0 & \text{if } \widehat{\theta}_i^2 \leq \frac{\log\left(\sigma_1^2/\sigma_0^2\right)}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}} \\ 1 & \text{if } \widehat{\theta}_i^2 > \frac{\log\left(\sigma_1^2/\sigma_0^2\right)}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}} \end{cases} . \qquad (7)
$$

2. *$\widehat{\boldsymbol{\theta}}$ is uniquely expressed through $\widehat{\mathbf{q}}$ as*

$$\widehat{\boldsymbol{\theta}} = (\widetilde{\mathcal{L}}^T \Sigma^{-1} \widetilde{\mathcal{L}} + K^{-1})^{-1} \widetilde{\mathcal{L}}^T \Sigma^{-1} \mathbf{y} \qquad (8)$$

where K is a diagonal matrix with diagonal elements $K_{ii} = \sigma^2_{\widehat{q}_i}$.

The immediate result of the lemma is the following iterative algorithm for the MAP estimator described in Table 1.

We point out two aspects regarding the algorithm. First, even though the total number of possible iterations to reach a local maximum can be as high as 2^M, in practice it usually converges to a fixed point in less than 10 iterations. Also, since the linear operator within the bracket to be inversed in

---

Step 0 Initialization: $\boldsymbol{\theta}^{(0)} = 0,\ i = 0$

Step 1 Estimate $\mathbf{q}^{(i+1)}$ using $\boldsymbol{\theta}^{(i)}$ as

$$
q_j^{(i+1)} = \begin{cases} 0 & \text{if } \left(\theta_j^{(i)}\right)^2 \leq \frac{\log\left(\sigma_1^2/\sigma_0^2\right)}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}} \\ 1 & \text{if } \left(\theta_j^{(i)}\right)^2 > \frac{\log\left(\sigma_1^2/\sigma_0^2\right)}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}} \end{cases} \qquad (9)
$$

Step 2 Using estimated $\mathbf{q}^{(i+1)}$ in step 1 to form the diagonal matrix $K^{(i+1)}$ with $K_{jj}^{(i+1)} = \sigma^2_{q_j^{(i+1)}}$, then compute $\boldsymbol{\theta}^{(i+1)}$ as

$$\boldsymbol{\theta}^{(i+1)} = \left[\widetilde{\mathcal{L}}^T \Sigma^{-1} \widetilde{\mathcal{L}} + \left(K^{(i+1)}\right)^{-1} \right]^{-1} \widetilde{\mathcal{L}}^T \Sigma^{-1} \mathbf{y} \qquad (10)$$

Step 3 If $\mathbf{q}^{(i)} \neq \mathbf{q}^{(i+1)}$, then $i = i + 1$ and repeat steps 1 and 2; otherwise exit.

Table 1. Bayesian IGMM model restoration algorithm.

(10) is positive definite, we use conjugate gradient descent method to compute its inverse. This inversion step is the most time-consuming part in the algorithm. Next, this iterative algorithm is highly nonlinear and signal dependent. Wavelet coefficients of large magnitude will tend to evolve into the high variance state, while wavelet coefficients of small magnitude will tend to be suppressed. These operations are carried out in a global way. The following theorem summarizes the convergence behavior of the algorithm.

Theorem 1 *Let L be as in (6). Given any initial estimate $\boldsymbol{\theta}^{(0)}$ and $\mathbf{q}^{(0)}$, the iterative algorithm described in Table 1 will stop after S iterations with $S \leq 2^M$.*

Let $L^{(i)} = L(\boldsymbol{\theta}^{(i)}, \mathbf{q}^{(i)})$, then

$$L^{(0)} \leq L^{(1)} \leq \cdots \leq L^{(S)} = L^{(S+1)} = \cdots .$$

Choose $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(S)}$ and $\widehat{\mathbf{q}} = \mathbf{q}^{(S)}$, then $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{q}})$ satisfies the coupling conditions (7) and (8) in Lemma 1. If in addition for all $\theta_i,\ i = 1, 2, \cdots, M$,

$$\theta_i^2 \neq \frac{\log\left(\sigma_1^2/\sigma_0^2\right)}{\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}},$$

then $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{q}})$ is a local maximum point for L.

3. EXTENSION TO NON-UNIFORM PRIORS

In this section we allow the prior $p(\mathbf{q})$ to be non-uniform. We assume that $\log p(\mathbf{q}) = \phi(T)$. The following theo-

	Bridge	Mandrill	Fruit	Camera man	Building
σ_0^2	0.0033	0.0136	0.0005	0.0002	0.0008
σ_1^2	0.4930	0.4911	0.6330	0.2841	0.7774

Table 2. σ_0^2 and σ_1^2 trained from EM algorithm. All image pixel intensities are linearly scaled to the range [0,1].

rem provides an efficient integer programming procedure to maximize H when ϕ is concave on $[0, M]$.

Theorem 2 *Let*

$$a_i = \log\left[\frac{1}{\sigma_0}e^{-\frac{\theta_i^2}{2\sigma_0^2}}\right], \quad b_i = \log\left[\frac{1}{\sigma_1}e^{-\frac{\theta_i^2}{2\sigma_1^2}}\right],$$

$$d_i = b_i - a_i,\ \triangle_i = \phi(i) - \phi(i-1),\ h_i = d_i + \triangle_i$$

for $1 \leq i \leq M$.

Without loss of generality, assume $d_1 \geq d_2 \geq \cdots \geq d_M$.

If $\phi(t)$ *is concave on* $[0, M]$*, i.e.,*

$$\phi((1-\alpha)t_1 + \alpha t_2)) \geq (1-\alpha)\phi(t_1) + \alpha\phi(t_2)$$

for any $\alpha \in [0, 1]$ *and* $t_1, t_2 \in [0, M]$*, then*

1. $h_1 \geq h_2 \geq \cdots \geq h_M$.
2. *A* $\mathbf{q}$ *that globally maximizes* H *is given by the following formula*

$$q_i = \left\lfloor \frac{sgn(h_i) + 1}{2} \right\rfloor, \quad 1 \leq i \leq M, \tag{11}$$

where $sgn(t)$ *is the sign function.*

4. EXPERIMENTS

In this section we apply the method developed in this paper to recover images distorted by the convolution operator. We first examine the wavelet coefficient distributions of a set of representative, natural images. In Table 2, we list the σ_0^2 and σ_1^2 values trained from some common test images. We see that σ_1^2 stays in a fairly stable range, while σ_0^2 has a large variation. σ_0^2 tends to be large for images with more texture content such as the "bridge" and "mandrill" test images. While it tends to be significantly smaller for less textural images such as the remaining 3 test images. So we can first roughly set $\sigma_0^2 = 0.01$ for textural images and $\sigma_0^2 = 0.0001$ for images such as "Fruit".

After using this set of parameters for the initial image recovery, we can use it to reestimate the parameters and then go through the process again to get a more accurate result. It is known that wavelet coefficients have exponential decay property across scales, so we can also estimate such property based on the roughly estimated image at hand. In Figure 1 we compare the parameters trained from the original image and those trained from the initially restored image. We see that the exponential decay of σ_1^2 is well preserved, while σ_0^2 has a large change even though it remains roughly constant.

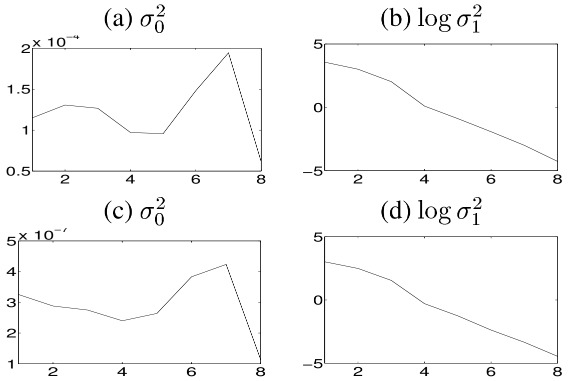

Fig. 1. *Re-estimation of* σ_0^2 *and* σ_1^2 *across scales for the "camera man" test image. (a), (b), from clean image. (c), (d), from restored image.*

Hence in the following experiments we only reestimate σ_1^2 across scales while keeping σ_0^2 unchanged.

One drawback of wavelet transform that limits its application is its translation dependence, which is also the cause of "blocky" artifacts in, for example, wavelet-based denoising. In [10] a translation invariant approach is taken to yield better performance. The idea is simple, for each circularly translated version of the observed data, apply the wavelet-based processing algorithm, then inverse translate the results back and take the average of them. This is equivalent to using different wavelet bases, all of which are translated versions of each other, to process the data and then take the average result. Here if we let $\mathcal{W}_v$ represent the wavelet transform using the standard wavelet basis translated by vector v, then we have

$$\mathbf{y} = \mathcal{L}\mathcal{W}_v^{-1}\boldsymbol{\theta}_v + \mathbf{n}$$

Notice that translation is also a linear operator. For each v, we apply the restoration process. In the end we take the average, i.e.,

$$\widehat{\mathbf{x}} = \frac{1}{|V|}\sum_{v \in V} \mathcal{W}_v^{-1}\widehat{\boldsymbol{\theta}}_v.$$

where V is the set of possible translations.

In Figure 2 we apply our method to the restoration of two test images. Harr wavelets are used. We first use a rough set of parameters to do an initial recovery. Then we use this recovery to estimate σ_1^2 scale by scale to get a better estimate of the parameters. After that we apply the algorithm for a collection of 25 translations (a 5×5 window of circular translations) and then take the average. we see that the texture regions are relatively well recovered, especially in difficult areas such as the facial region of the mandrill.

5. CONCLUSIONS

In this paper we develop a new wavelet-based method for solving linear inverse problem and apply it to image restoration from noisy and distorted observations. This Bayesian method employs a wavelet-domain independent GMM signal prior to produce a computationally efficient restoration algorithm that has produced good results. The extensions of the method developed in this paper are many. One is the direct application to problems involving any multi-mixture model, or use more realistic exponential decay across scale property on wavelet coefficients. Recently, graphical models, such as the wavelet-domain hidden Markov models proposed by Crouse *et al.*[9, 11], have been applied in image processing with very good results. In this paper, we only present the simplest independent wavelet coefficient prior to illustrate the key ideas and basic results, but the framework could be extended to handle more sophisticated HMM type priors.

6. REFERENCES

[1] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, New Jersey: Prentice Hall, 1977.

[2] F. M. J.-L Starck and A. Bijaoui, *Image Processing and Data Analysis*. Cambridge Univeristy Press, 1998.

[3] M. Bertero, "Linear inverse and ill-posed problems," *Advances in Electronics and Electron Physics*, vol. 75, no. 1, pp. 1–120, 1989.

[4] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy. Statist. Soc. Ser. B., 60, 725-749*, vol. 60, pp. 725–749, 1998.

[5] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *App. and Comp. Harmonic Analysis*, vol. 2, pp. 101–126, 1995.

[6] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavlet-based multiscale image restoration," *IEEE Trans. Image Processing*, vol. 5, no. 4, pp. 619–634, 1996.

[7] M. Belge, M. E. Kilmer, and E. L. Miller, "Wavelet domain image restoration with adaptive edge-peserving regularity," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 597–608, 2000.

[8] Y. Wan and R. Nowak, "A bayesian multiscale approach to joint image restoration and edge detection," in *Proceedings of SPIE Conf. 3813 — Wavelet Application in Signal and Image Processing VII*, Denver, CO, July, 1999.

[9] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, 1998.

[10] R. Coifman and D. Donoho, "Translation invariant denoising," in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. New York: Springer-Verlag, pp. 125–150, 1995.

[11] R. Nowak, "Multiscale hidden Markov models for Bayesian image analysis," in *Bayesian Inference in Wavelet Based Models*, Springer-Verlag, 1999. Editors B. Vidakovic and P. Müller.

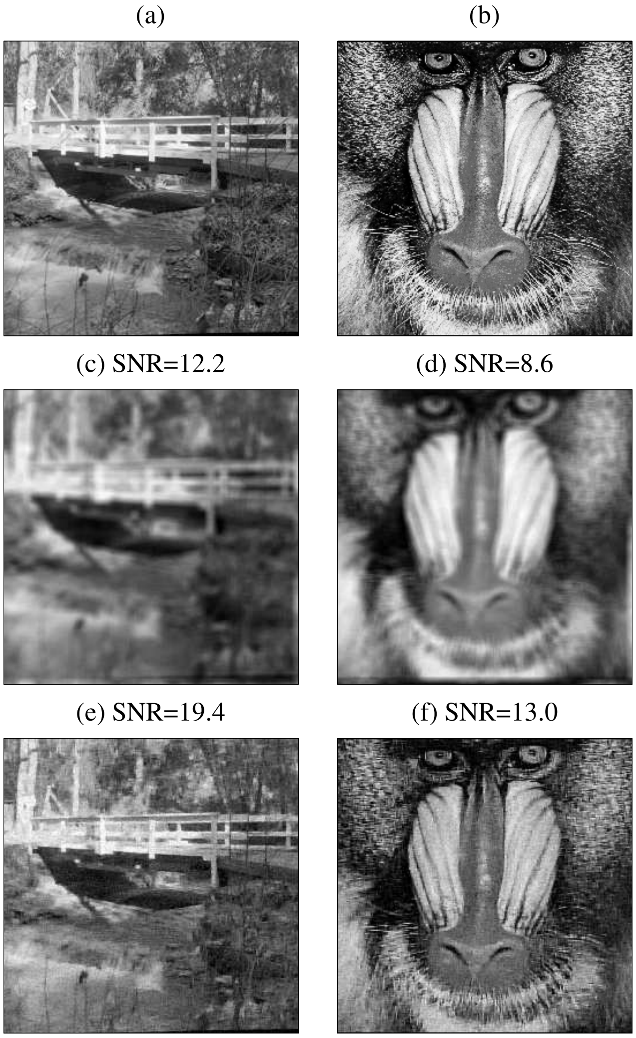

Fig. 2. *Experiments on "bridge" and "mandrill" test images. (a) (b), Clean images. (c) (d) Observed images after blurring by 7×7 box-car matrix with added white Gaussian noise (BSNR=30dB). (e) (f), Recovered images with initially chosen $\sigma_0^2 = 10^{-2}$ and $\sigma_1^2 = 10^{-1}$.*