AN EMPIRICAL BAYES EM-WAVELET UNIFICATION FOR SIMULTANEOUS DENOISING, INTERPOLATION, AND/OR DEMOSAICING

Keigo Hirakawa and Xiao-Li Meng

Harvard University
Department of Statistics
1 Oxford Street, Cambridge, MA 02138
{hirakawa,meng}@stat.harvard.edu

ABSTRACT

We present a unified framework for coupling the EM algorithm with the Bayesian hierarchical modeling of neighboring wavelet coefficients of image signals. Within this framework, problems with missing pixels or pixel components, and hence unobservable wavelet coefficients, are handled simultaneously with denoising. The hyper-parameters of the model are estimated via the marginal likelihood by the EM algorithm, and a part of the output of its E-step automatically provide optimal estimates, given the specified Bayesian model, of the noise-free image. This unified empirical-Bayes based framework, therefore, offers a statistically principled and extremely flexible approach to a wide range of pixel estimation problems including image denoising, image interpolation, demosaicing, or any combinations of them.

Index Terms— wavelets, missing data, denoising, interpolation

1. INTRODUCTION

The statistical modeling of image signals in the wavelet domain is primarily based on the observations that a wavelet coefficient and its adjacent/parent coefficients of image signals are correlated. Here we use Bayesian hierarchical modeling to capture such correlations. While similar approaches have been suggested [1, 2, 3], wavelet-based imaging algorithms commonly use heuristically determined parameters, which are often restrictive or inaccurate. To overcome these short comings, we assume a general model form and couple the EM algorithm framework with the Bayesian models to estimate the hyper-parameters via the marginal likelihood, that is, we adopt the empirical Bayes approach. One key benefit of this approach is that a part of the output provides automatically provides the optimal Bayesian estimate (given the specified model) of the noise-free image, as specified. Under the assumption of additive noise, a similar EM algorithm was designed in [4, 5], but the parameter estimation is limited to that of noise correlation matrix.

Another key benefit of this approach is that it can easily be extended to *simultaneously* deal with problems with missing or incomplete pixel values, either because of mechanical designs (e.g., demosaicing) or because distortion (e.g., picture impainting) or both. In the context of wavelet-based image processing, missing or incomplete pixel is a difficult problem because wavelet transform takes a linear combination of image signal, and thus many, or even *all* of the noisy wavelet coefficients are unobserved. Using the EM algorithm, however, the considerations for the missing pixel imputation and the model parameter estimation can be combined into a unified theoretical framework. The E-step imputation can be viewed as a solution to a model-based image denoising problem with incomplete data, or a model-based image interpolation given noisy data. This generalization has wide ranges of useful applications that include image denoising, image interpolation, demosaicing, or *combinations of them*.

Authors thank Dr. J. Portilla and Dr. E.P. Simoncelli for their helpful comments, and NSF for partial funding.

Given the space limitation and the considerable generality of our approach, we assume the familiarity with the EM algorithm [6] and focus on the basic theory under additive Gaussian noise. However, neither the additivity nor the normality is essential to our general approach–they just simplify both the presentation and implementation. We present the Bayesian hierarchical model in Section 2, and its EM algorithm in Section 3. We illustrate the method with experimental results in Section 4, and conclude in Section 5.

2. BAYESIAN HIERARCHICAL MODELING

Suppose there are N pixels in an image $\boldsymbol{f}^\top = [f_1, \ldots, f_N]$, where f_i is the true pixel value at location $i \in \{1, \ldots, N\}$. In an image denoising problem, the observed image, $\boldsymbol{y}$ is corrupted by noise. Assuming an independent additive Gaussian noise, let $\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{e}$, where $\boldsymbol{e} \sim \mathcal{N}(0, \Sigma_e)$. The objective here is to estimate $\boldsymbol{f}$ given $\boldsymbol{y}$. Suppose only some of the pixels in the image $\boldsymbol{y}$ are observed. Define $\boldsymbol{y}_{\text{obs}}^\top = [y_{\text{obs}1}, \ldots, y_{\text{obs}N_\text{o}}]$ and $\boldsymbol{y}_{\text{mis}}^\top = [y_{\text{mis}1}, \ldots, y_{\text{mis}N_\text{m}}]$ as observed and missing pixel values, respectively, where $N_\text{o} + N_\text{m} = N$ and, without loss of generality, $\boldsymbol{y}^\top = [\boldsymbol{y}_{\text{obs}}^\top, \boldsymbol{y}_{\text{mis}}^\top]$. We represent $\boldsymbol{f}$ and $\boldsymbol{y}$ by wavelet transform, $\boldsymbol{d} = \boldsymbol{W}\boldsymbol{f}$ and $\boldsymbol{w} = \boldsymbol{W}\boldsymbol{y}$, respectively. Here, $\boldsymbol{W}$ is the two-dimensional wavelet transform matrix operator with M subbands. Define w_{km} as the wavelet coefficient at the k-th pixel location in the m-th subband ($m = 0$ represents the scaling coefficients). Define vector $\boldsymbol{w}_{km} = \{w_{k'm'}\} \in \mathbb{R}^n$, where $w_{k'm'}$ are coefficients in $\boldsymbol{w}$ that are in the *neighborhood* of w_{km} (to be made precise in the sequel), and let $\boldsymbol{d}_{km}$ be defined similarly. Let $\boldsymbol{W}_{km}$ correspond to the appropriate rows of $\boldsymbol{W}$ such that $\boldsymbol{w}_{km} = \boldsymbol{W}_{km}\boldsymbol{y}$. Here we assume orthogonal wavelet transform (i.e. $\boldsymbol{W}\boldsymbol{W}^\top = \boldsymbol{I}$), but the theoretical framework extends to non-orthogonal and over-complete transforms in a straightforward manner (see Section 4).

The Bayesian hierarchical model of the image wavelet coefficients considered in this paper is based on ideas exploited in [1]. Assuming signal-independent additive Gaussian noise, wavelet coefficients are modeled hierarchically. Motivated by the observation that wavelet coefficients of natural images tend toward heavy-tailed distributions (see [1]), we model $\boldsymbol{d}_{km}$ by a t-distribution with ν_m degrees of freedom. That is, $\boldsymbol{w}_{km}$ and $\boldsymbol{d}_{km}$ take the following form:

$$\boldsymbol{w}_{km}|\boldsymbol{d}_{km} \sim \text{indep} \quad \mathcal{N}(\boldsymbol{d}_{km}, \boldsymbol{\Sigma}_{wm}) \tag{1}$$

$$\boldsymbol{d}_{km}|q_{km} \sim \text{indep} \quad \mathcal{N}(\boldsymbol{\xi}_m, \boldsymbol{\Sigma}_{dm}/q_{km}) \tag{2}$$

$$q_{km} \sim \text{indep} \quad \chi^2_{\nu_m}/\nu_m. \tag{3}$$

The use of the unobserved mean chi-squared variables $\boldsymbol{q} = \{q_{km}\}$ to represent the t_{ν_m} as part of the *augmented data* is a well-known strategy for easy EM implementation [7]. Here out goal is to estimate the (hyper-)parameter $\boldsymbol{\theta} = \{\boldsymbol{\Sigma}_{wm}, \boldsymbol{\Sigma}_{dm}, \nu_m, \boldsymbol{\xi}_m\}$ by maximizing the marginal log-likelihood $\ell(\boldsymbol{\theta}|\boldsymbol{y}_{\rm obs}) = \log p(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta})$. The direct maximization is very difficult because of the missing pixel values. The EM algorithm circumvents this problem by iteratively maximizing the much easier augmented-data log-likelihood $\ell(\boldsymbol{\theta}|\boldsymbol{x}) = \log p(\boldsymbol{x}|\theta)$, where $\boldsymbol{x} = \{\boldsymbol{y}, \boldsymbol{f}, \boldsymbol{q}\}$, or equivalently $\{\boldsymbol{w}, \boldsymbol{d}, \boldsymbol{q}\}$ are the *augmented data*. Because only $\boldsymbol{y}_{\rm obs}$ is observed, the E-step computes the imputation of augmented-data sufficient statistics, needed by the M-step, which includes $\hat{\boldsymbol{d}}^{[t+1]} = E\left[\boldsymbol{d}\middle|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}\right]$. This can be used to compute the "best" estimate, at the $(t+1)st$ iteration, of $\boldsymbol{f}$, the noise-free image, as $\hat{\boldsymbol{f}}^{[t+1]} = \boldsymbol{W}^T\hat{\boldsymbol{d}}^{[t+1]}$.

Before we proceed, we point out a potential theoretical incompatibility inherited in (1)-(2) when the neighborhoods are allowed to have overlaps. As we have no space to discuss this issue in any details, in particularly its connection with the over-complete expansions, we simply point out that assuming independence between overlapping neighborhoods provides a computationally efficient approximation for practical implementation.

3. EM ALGORITHM DERIVATION

Given $\boldsymbol{\theta}^{[t]}$, the (t+1)st iteration of the EM algorithm first calls for

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[t]}) = E\left[\log p(\boldsymbol{x}|\boldsymbol{\theta})\middle|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}\right]$$
$$= \sum_{k,m} E\left[\log p(\boldsymbol{w}_{km}|\boldsymbol{d}_{km}; \boldsymbol{\Sigma}_{wm}) + \log p(\boldsymbol{d}_{km}|q_{km}; \xi_m, \boldsymbol{\Sigma}_{dm}) + \log p(q_{km}|\nu_m)\middle|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}\right].$$

A celebrated result of EM algorithm [6] states that the choice of $\boldsymbol{\theta}$ that maximizes $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[t]})$, that is, the next iterate $\boldsymbol{\theta}^{[t+1]}$ increases our objective function: $\ell(\boldsymbol{\theta}^{[t+1]}|\boldsymbol{y}_{\rm obs}) \geq \ell(\boldsymbol{\theta}^{[t]}|\boldsymbol{y}_{\rm obs})$. The advantage of maximizing $Q\left(\boldsymbol{\theta}; \boldsymbol{\theta}^{[t]}\right)$ instead of directly the targeted $\ell(\boldsymbol{\theta}|\boldsymbol{y}_{\rm obs})$ is that it is often the case that $\log p(\boldsymbol{x}|\boldsymbol{\theta})$ depends on the missing part of $\boldsymbol{x}$ linearly via a set of *augmented-data sufficient statistics* $\boldsymbol{S}(\boldsymbol{x})$. Consequently, maximizing $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[t]})$ is the same as maximizing $\ell(\boldsymbol{\theta}|\boldsymbol{x})$, but with the (partially) missing $\boldsymbol{S}(\boldsymbol{x})$ imputed via the E-step: $\boldsymbol{S}^{[t+1]}(\boldsymbol{x}) = E\left[\boldsymbol{S}(\boldsymbol{x})\middle|\boldsymbol{y}_{\rm obs}; \boldsymbol{\theta}^{[t]}\right]$.

For our problem, this is most clearly seen when ν_m is known, and thus the $\log p(q_{km}|\nu_m)$ part is unneeded for $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[t]})$. We therefore present our basic algorithm first under this assumption. It is then easy to verify that $\boldsymbol{S}(\boldsymbol{x}) = \{\boldsymbol{S}_1(\boldsymbol{w}, \boldsymbol{d}), \boldsymbol{S}_2(\boldsymbol{q}, \boldsymbol{d})\}$, where

$$\boldsymbol{S}_1(\boldsymbol{w}, \boldsymbol{d}) = \{\boldsymbol{w}_{km}\boldsymbol{w}_{km}^\top,\ \boldsymbol{d}_{km}\boldsymbol{d}_{km}^\top,\ \boldsymbol{w}_{km}\boldsymbol{d}_{km}^\top\},$$
$$\boldsymbol{S}_2(\boldsymbol{q}, \boldsymbol{d}) = \{q_{km},\ q_{km}\boldsymbol{d}_{km},\ q_{km}\boldsymbol{d}_{km}\boldsymbol{d}_{km}^\top\}.$$

Indeed, since $\boldsymbol{\theta}$ is reduced to $\boldsymbol{\theta}_1 = \{\boldsymbol{\Sigma}_{wm}, \boldsymbol{\Sigma}_{dm}, \boldsymbol{\xi}_m\}$, the maximizer of $Q(\boldsymbol{\theta}_1; \boldsymbol{\theta}_1^{[t]})$ is the weighted least-square estimate [6]:

$$\boldsymbol{\xi}_m^{[t+1]} = \frac{\sum_k E[q_{km}\boldsymbol{d}_{km}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}]}{\sum_k E[q_{km}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}]}, \tag{4}$$

$$\boldsymbol{\Sigma}_{wm}^{[t+1]} = n_m^{-1}\sum_k E[\boldsymbol{w}_{km}\boldsymbol{w}_{km}^\top - \boldsymbol{w}_{km}\boldsymbol{d}_{km}^\top - \boldsymbol{d}_{km}\boldsymbol{w}_{km}^\top + \boldsymbol{d}_{km}\boldsymbol{d}_{km}^\top|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}], \tag{5}$$

$$\boldsymbol{\Sigma}_{dm}^{[t+1]} = n_m^{-1}\sum_k E[q_{km}(\boldsymbol{d}_{km}\boldsymbol{d}_{km}^\top - \boldsymbol{d}_{km}\boldsymbol{\xi}_m^{[t+1]\top} - \boldsymbol{\xi}_m^{[t+1]}\boldsymbol{d}_{km}^\top + \boldsymbol{\xi}_m^{[t+1]}\boldsymbol{\xi}_m^{[t+1]\top})|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}], \tag{6}$$

where n_m is the number of wavelet coefficients in the m-th subband.

3.1. A Conditional E-step

To carry our the above calculation, we need to compute $\boldsymbol{S}^{[t+1]}(\boldsymbol{x}) = E\left[\boldsymbol{S}(\boldsymbol{x})\middle|\boldsymbol{y}_{\rm obs}; \boldsymbol{\theta}^{[t]}\right]$. This computation involves the t-distribution for $\boldsymbol{q}$, and therefore is complicated. However, when conditioned on $\boldsymbol{q}$, only multivariate normal distributions are involved. So our general strategy is to first take the expectation conditioned on $\boldsymbol{q}$ (and $\boldsymbol{y}_{\rm obs}$ with $\boldsymbol{\theta} = \boldsymbol{\theta}^{[t]}$), and then integrate out $\boldsymbol{q}$ numerically when we complete the M-step.

Specifically, let $\boldsymbol{\varphi} = E(\boldsymbol{y}|\boldsymbol{q}; \boldsymbol{\theta}) = \boldsymbol{W}^T E[w|\boldsymbol{q}]$ and $\boldsymbol{\Omega}_{|\boldsymbol{q}} = \mathrm{Cov}(\boldsymbol{y}|\boldsymbol{q}, \boldsymbol{\theta}) = \boldsymbol{W}^\top \mathrm{Cov}(\boldsymbol{w}|\boldsymbol{q}, \boldsymbol{\theta})\boldsymbol{W}$ be the conditional mean and variance of $\boldsymbol{y}$ given $\boldsymbol{q}$ under our model (1)-(3). Note that $\boldsymbol{\varphi}$ is free of $\boldsymbol{q}$ under our model, and given $\boldsymbol{q}$, $\boldsymbol{\Omega}_{|\boldsymbol{q}}$ and $\boldsymbol{\varphi}$ can be defined completely by the parameters in $\boldsymbol{\theta}$. Define $\boldsymbol{\varphi}_{\rm obs}$ the sub-vector of $\boldsymbol{\varphi}$ that corresponds to $\boldsymbol{y}_{\rm obs}$, and partition $\mathrm{Cov}(\boldsymbol{y}|\boldsymbol{q}, \boldsymbol{\theta})$ into four submatrices according to the partition $\boldsymbol{y} = (\boldsymbol{y}_{\rm obs}, \boldsymbol{y}_{\rm mis})$ as $\boldsymbol{\Omega}_{|\boldsymbol{q}} = [\boldsymbol{\Omega}_{o|\boldsymbol{q}}, \boldsymbol{\Omega}_{om|\boldsymbol{q}}; \boldsymbol{\Omega}_{mo|\boldsymbol{q}}, \boldsymbol{\Omega}_{m|\boldsymbol{q}}]$. Conditioned on $\boldsymbol{q}$, the noisy image $\boldsymbol{y}$ is imputed via the standard regression estimate:

$$\hat{\boldsymbol{y}}_{|\boldsymbol{q}} = E[\boldsymbol{y}|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}] = \boldsymbol{\varphi} + [\boldsymbol{\Omega}_{o|\boldsymbol{q}}; \boldsymbol{\Omega}_{mo|\boldsymbol{q}}]\boldsymbol{\Omega}_{o|\boldsymbol{q}}^{-1}(\boldsymbol{y}_{\rm obs} - \boldsymbol{\varphi}_{\rm obs}).$$

And $\mathrm{Cov}(\boldsymbol{y}_{\rm mis}|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta})$ is a Schur complement of $\boldsymbol{\Omega}_{o|\boldsymbol{q}}$ in $\boldsymbol{\Omega}_{|\boldsymbol{q}}$:

$$\mathrm{Cov}(\boldsymbol{y}_{\rm mis}|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}) = \boldsymbol{\Omega}_{m|\boldsymbol{q}} - \boldsymbol{\Omega}_{mo|\boldsymbol{q}}\boldsymbol{\Omega}_{o|\boldsymbol{q}}^{-1}\boldsymbol{\Omega}_{om|\boldsymbol{q}}.$$

Imputation of other data vectors follow easily:

$$\hat{\boldsymbol{w}}_{km|\boldsymbol{q}} = E[\boldsymbol{w}_{km}|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}] = \boldsymbol{W}_{km}E[\boldsymbol{y}|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}] = \boldsymbol{W}_{km}\hat{\boldsymbol{y}}_{|\boldsymbol{q}}$$
$$\hat{\boldsymbol{d}}_{km|\boldsymbol{q}} = E[\boldsymbol{d}_{km}|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}]$$
$$= E[E[\boldsymbol{d}_{km}|\boldsymbol{w}_{km}, \boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}]|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}]$$
$$= E[\boldsymbol{\xi}_m + (\boldsymbol{\Sigma}_{wm}^{-1} + q_{km}\boldsymbol{\Sigma}_{dm}^{-1})^{-1}\boldsymbol{\Sigma}_{wm}^{-1}(\boldsymbol{w}_{km} - \boldsymbol{\xi}_m)|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}]$$
$$= \boldsymbol{\xi}_m + (\boldsymbol{\Sigma}_{wm}^{-1} + q_{km}\boldsymbol{\Sigma}_{dm}^{-1})^{-1}\boldsymbol{\Sigma}_{wm}^{-1}(\hat{\boldsymbol{w}}_{km|\boldsymbol{q}} - \boldsymbol{\xi}_m).$$

Using the formulae above, all the covariance matrices (conditioned on $\boldsymbol{q}$) that are needed to carry out $E\left[\boldsymbol{S}(\boldsymbol{x})\middle|\boldsymbol{q}, \boldsymbol{y}_{\rm obs}; \boldsymbol{\theta}\right]$ can be calculated analytically. See (7), (8), and (9), where $\boldsymbol{I}$ is an identity matrix, and $\boldsymbol{C}_{dkm|\boldsymbol{q}} = (\boldsymbol{\Sigma}_{wm}^{-1} + q_{km}\boldsymbol{\Sigma}_{dm}^{-1})^{-1}$.

3.2. The M-step

By (4)-(6), the M-step updates $\boldsymbol{\theta}_1 = \{\boldsymbol{\Sigma}_{wm}, \boldsymbol{\Sigma}_{dm}, \boldsymbol{\xi}_m\}$ by

$$\boldsymbol{\xi}_m^{[t+1]} = \frac{\sum_k \int_0^\infty q_{km}\hat{\boldsymbol{d}}_{km|\boldsymbol{q}}p(\boldsymbol{q}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]})d\boldsymbol{q}}{\sum_k \int_0^\infty q_{km}p(\boldsymbol{q}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]})d\boldsymbol{q}},$$

$$\boldsymbol{\Sigma}_{wm}^{[t+1]} = n_m^{-1}\sum_k \int_0^\infty (\boldsymbol{\Lambda}_{wkm|\boldsymbol{q}} - \boldsymbol{\Lambda}_{wdkm|\boldsymbol{q}} - \boldsymbol{\Lambda}_{wdkm|\boldsymbol{q}}^\top + \boldsymbol{\Lambda}_{dkm|\boldsymbol{q}})p(\boldsymbol{q}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]})d\boldsymbol{q},$$

$$\boldsymbol{\Sigma}_{dm}^{[t+1]} = n_m^{-1}\sum_k \int_0^\infty q_{km}(\boldsymbol{\Lambda}_{dkm|\boldsymbol{q}} - \hat{\boldsymbol{d}}_{km|\boldsymbol{q}}\boldsymbol{\xi}_m^{[t+1]\top} - \boldsymbol{\xi}_m^{[t+1]}\hat{\boldsymbol{d}}_{km|\boldsymbol{q}}^\top + \boldsymbol{\xi}_m^{[t+1]}\boldsymbol{\xi}_m^{[t+1]\top})p(\boldsymbol{q}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]})d\boldsymbol{q}.$$

Likewise, the estimate for the noise-free wavelet coefficients are found by integration:

$$\hat{\boldsymbol{d}}_{km} = E[\boldsymbol{d}_{km}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}] = \int_0^\infty \hat{\boldsymbol{d}}_{km|\boldsymbol{q}}p(\boldsymbol{q}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta})d\boldsymbol{q}.$$

The integrals above can be simplified via the Bayes rule. Specifically, for any integrable $g(\boldsymbol{q})$, we have

$$E[g(\boldsymbol{q})|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}] = \frac{\int_0^\infty g(\boldsymbol{q})\phi_{\rm obs}(\boldsymbol{q})\prod_{k,m}\tilde{p}(q_{km}|\nu_m)dq_{km}}{\int_0^\infty \phi_{\rm obs}(\boldsymbol{q})\prod_{k,m}\tilde{p}(q_{km}|\nu_m)dq_{km}}$$

$$
\begin{aligned}
\boldsymbol{\Lambda}_{wkm|\boldsymbol{q}} =& E[\boldsymbol{w}_{km}\boldsymbol{w}_{km}^\top|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}] = \boldsymbol{W}_{km}E[\boldsymbol{y}\boldsymbol{y}^\top|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}]\boldsymbol{W}_{km}^\top \\
=& \boldsymbol{W}_{km}\left[\mathrm{Cov}(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}) + \hat{\boldsymbol{y}}_{|\boldsymbol{q}}\hat{\boldsymbol{y}}_{|\boldsymbol{q}}^\top\right]\boldsymbol{W}_{km}^\top = \boldsymbol{W}_{km}\mathrm{Cov}(\boldsymbol{y}|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta})\boldsymbol{W}_{km}^\top + \hat{\boldsymbol{w}}_{km|\boldsymbol{q}}\hat{\boldsymbol{w}}_{km|\boldsymbol{q}}^\top \qquad (7) \\
\boldsymbol{\Lambda}_{dkm|\boldsymbol{q}} =& E[\boldsymbol{d}_{km}\boldsymbol{d}_{km}^\top|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}] = E\left\{E[\boldsymbol{d}_{km}\boldsymbol{d}_{km}^\top|\boldsymbol{w},\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}]\Big|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}\right\} \\
=& E\left\{\mathrm{Cov}(\boldsymbol{d}_{km}|\boldsymbol{w},\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}) + E[\boldsymbol{d}_{km}|\boldsymbol{w},\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}]E[\boldsymbol{d}_{km}^\top|\boldsymbol{w},\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}]\Big|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}\right\} \\
=& \boldsymbol{C}_{dkm|\boldsymbol{q}} + E\left\{\left[\boldsymbol{\xi}_m + \boldsymbol{C}_{dkm|\boldsymbol{q}}\boldsymbol{\Sigma}_{wm}^{-1}(\boldsymbol{w}_{km}-\boldsymbol{\xi}_m)\right]\left[(\boldsymbol{w}_{km}-\boldsymbol{\xi}_m)^\top\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}} + \boldsymbol{\xi}_m^\top\right]\Big|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}\right\} \qquad (8) \\
=& \boldsymbol{C}_{dkm|\boldsymbol{q}} + (\boldsymbol{I}-\boldsymbol{C}_{dkm|\boldsymbol{q}}\boldsymbol{\Sigma}_{wm}^{-1})(\boldsymbol{\xi}_m\boldsymbol{\xi}_m^\top)(\boldsymbol{I}-\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}}) + (\boldsymbol{I}-\boldsymbol{C}_{dkm|\boldsymbol{q}}\boldsymbol{\Sigma}_{wm}^{-1})(\boldsymbol{\xi}_m\hat{\boldsymbol{w}}_{km|\boldsymbol{q}}^\top)(\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}}) \\
&+ (\boldsymbol{C}_{dkm|\boldsymbol{q}}\boldsymbol{\Sigma}_{wm}^{-1})(\hat{\boldsymbol{w}}_{km|\boldsymbol{q}}\boldsymbol{\xi}_m^\top)(\boldsymbol{I}-\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}}) + (\boldsymbol{C}_{dkm|\boldsymbol{q}}\boldsymbol{\Sigma}_{wm}^{-1})\Lambda_{wkm|\boldsymbol{q}}(\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}}) \\
\boldsymbol{\Lambda}_{wdkm|\boldsymbol{q}} =& E[\boldsymbol{w}_{km}\boldsymbol{d}_{km}^\top|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}] = E\left\{\boldsymbol{w}_{km}E[\boldsymbol{d}_{km}^\top|\boldsymbol{w},\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}]\Big|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}\right\} \qquad (9) \\
=& E\left\{\boldsymbol{w}_{km}\left[(\boldsymbol{w}_{km}-\boldsymbol{\xi}_m)^\top\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}} + \boldsymbol{\xi}_m^\top\right]\Big|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}\right\} = \boldsymbol{\Lambda}_{wkm|\boldsymbol{q}}\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}} + \hat{\boldsymbol{w}}_{km|\boldsymbol{q}}\boldsymbol{\xi}_m^\top(\boldsymbol{I}-\boldsymbol{\Sigma}_{wm}^{-1}\boldsymbol{C}_{dkm|\boldsymbol{q}})
\end{aligned}
$$

where, denoting $|\cdot|$ the determinant operator,

$$
\phi_{\rm obs}(\boldsymbol{q}) = |\boldsymbol{\Omega}_{o|\boldsymbol{q}}|^{-\frac{1}{2}}\exp\left\{-\frac{(\boldsymbol{y}_{\rm obs}-\boldsymbol{\varphi}_{\rm obs})^\top\boldsymbol{\Omega}_{o|\boldsymbol{q}}^{-1}(\boldsymbol{y}_{\rm obs}-\boldsymbol{\varphi}_{\rm obs})}{2}\right\}
$$

$$
\tilde{p}(q_{km}|\nu_m) = q_{km}^{(\nu_m/2)-1}\exp\left\{-\frac{\nu_m q_{km}}{2}\right\}. \qquad (10)
$$

3.3. When ν_m is unknown

Both the E-step and M-step becomes a bit more involved when ν_m also needs to be estimated. For such cases, we can employ the ECME (expectation-conditional maximization either) algorithm [6, 8, 7] —an extension of EM algorithm–to speed up the computation. Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$, where $\boldsymbol{\theta}_1 = \{\boldsymbol{\Sigma}_{wm}, \boldsymbol{\Sigma}_{dm}, \boldsymbol{\xi}_m\}$ and $\boldsymbol{\theta}_2 = \{\nu_m\}$. The ECME uses two conditional M-steps instead of one M-step:

$$
\begin{aligned}
\boldsymbol{\theta}_1^{[t+1]} &= \arg\max_{\boldsymbol{\theta}_1} Q(\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{[t]}\}; \boldsymbol{\theta}^{[t]}) \\
\boldsymbol{\theta}_2^{[t+1]} &= \arg\max_{\boldsymbol{\theta}_2} \log p(\boldsymbol{y}_{\rm obs}|\{\boldsymbol{\theta}_1^{[t+1]}, \boldsymbol{\theta}_2\}).
\end{aligned}
$$

Here the first CM-step is identical to the M-step described above. The second CM-step can be carried out via Newton-Raphson:

$$
\nu_{m\,\text{new}}^{[t+1]} = \nu_{m\,\text{old}}^{[t+1]} - \frac{\ell'(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{[t+1]})}{\ell''(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1^{[t+1]})}, \qquad (11)
$$

where $\ell(\boldsymbol{\theta}_2|\boldsymbol{\theta}_1) = \log p(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ and

$$
p(\boldsymbol{y}_{\rm obs}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \propto \int_0^\infty \phi_{\rm obs}(\boldsymbol{q}) \prod_{k,m} p(q_{km}|\nu_m) dq_{km}
$$

can be evaluated numerically as before. Note here $p(q_{km}|\nu_m)$ must be the proper density of $\chi^2_{\nu_m}/\nu_m$, not the unnormalized density as given in (10), because the normalizing constant is a function of the unknown parameter ν_m and thus it is needed in calculating the derivatives required by (11). Note also that ν_m is continuous as χ^2_ν distribution is a special case of the Gamma distribution with the shape parameter $\alpha = \nu/2$.

4. IMPLEMENTATION, APPLICATIONS, AND RESULTS

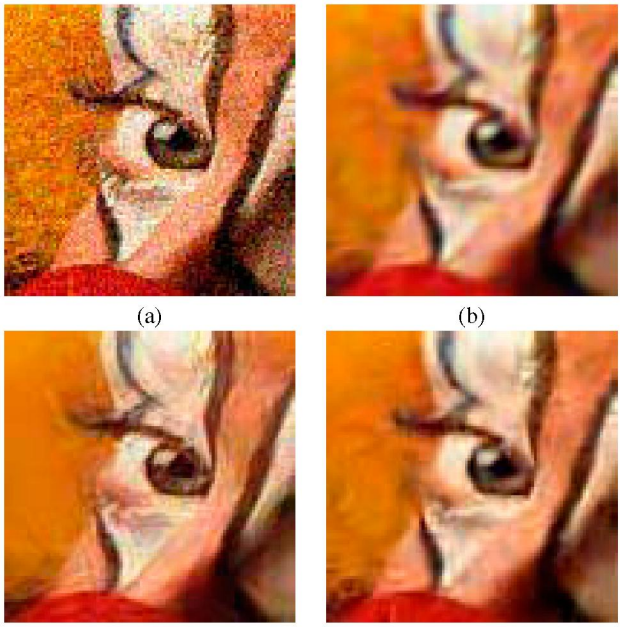
Once $\boldsymbol{\theta}^{[t]}$ is found by our algorithm, our estimate of the true image $\boldsymbol{f}$ is the posterior mean, $\boldsymbol{f}^{[t]} = \boldsymbol{W}^\top E[\boldsymbol{d}|\boldsymbol{y}_{\rm obs}, \boldsymbol{\theta}^{[t]}]$, where $\boldsymbol{W}$ now stands for wavelet transform applied to each color component independently. This is clearly useful to a number of applications.

The algorithm is implemented using the overcomplete Daubechies 4 separable wavelets. The overcomplete expansion is accomplished by convolution (i.e. ignoring the downsample operator), and the redundant representation of wavelet coefficients are projected back to the span of the image signal by averaging together the shifted versions of inverse wavelet transforms, amounting to the model averaging in the pixel domain. The elements in $\boldsymbol{w}_{km}$ are the wavelet coefficients taken from a 3×3 spatial window centered around w_{km}. Here the members of $\boldsymbol{w}_{km}$ are taken only from the m-th subband, although the inclusion of parent/child wavelets is a straightforward extension. For color image $\boldsymbol{f}$, $\boldsymbol{w}_{km}$ includes the wavelet coefficients from all color components belonging to their respective 3×3 spatial window centered around w_{km}. Note that there are redundant estimates of the wavelet coefficients in $\boldsymbol{d}_{km}$ because of the overlapping neighborhoods of 3×3 moving window. These redundant estimates are averaged together before the inverse wavelet transform.

For our experiment, the degree of freedom ν_m is set to 0.1 for wavelet coefficients and 1 for scaling coefficients (so the second CM-step is unneeded). All other parameters are estimated by our algorithm. The initial values $\boldsymbol{\Sigma}_{dm}$ are trained using a reference image, or a *typical* noise-free (color) image that with no missing data. The initial noise correlation matrices $\boldsymbol{\Sigma}_{wm}$ are set to an identity matrix multiplied by the initial guess of the noise variance. The prior mean vectors $\boldsymbol{\xi}_m$ are set to 0 for the wavelet coefficients and the average color values in $\boldsymbol{y}_{\rm obs}$ for the scaling coefficients.

Because the literal execution of this algorithm is computationally intensive, we made a few approximations in our implementation (due to page constraints, the impact of such approximations will be reported in future work). First, we assume q_{km} is spatially slowly varying and therefore we effectively treat $q_{k'm} = q_{km}$ when k and $k's$ belongs to the same 3×3 neighborhood (or we can view this as a model modification to (3) by assuming a common q within each neighborhood). This approximation greatly reduces the dimension of the numerical integration over q to one. Second, to greatly reduce the size of matrices in computing the regression imputation, we approximate $E[y_i|\boldsymbol{q},\boldsymbol{y}_{\rm obs},\boldsymbol{\theta}]$ by $E[y_i|\boldsymbol{q},\boldsymbol{y}_{{\rm obs}\,i},\boldsymbol{\theta}]$, where $\boldsymbol{y}_{{\rm obs}\,i}$ is a subset of $\boldsymbol{y}_{\rm obs}$ restricted to a $n_i \times n_i$ neighborhood around the pixel location i; we found a value $n_i \in (10, 20)$ typically adequate.

In the case that all noisy pixels are observed, the proposed algorithm presents a method for estimating the noise-free pixel values, $\boldsymbol{f}$, given noisy pixels $\boldsymbol{y}$. Conversely, when only a subset of the noise-free pixels are observed, the proposed algorithm offers a method to interpolate the missing pixels. Furthermore, by interpreting spectral channels as separate different *color* components of the image,



(a) (b) (c) (d)

Fig. 1. Zoomed portion of the output images ($\sigma_n = 25$). Method(s) in (a) [13], (b) [13]+[1], (c) [9], (d) proposed (3rd iteration).

the interpolation method generalizes to hyper-spectral images seamlessly. In the real-world image interpolation problems, however, the observed data most certainly contain noise, sometimes to a severe degree. This is especially apparent in digital cameras, as the noise in the image sensor in a poor lighting condition is often amplified in the interpolation step known as demosaic(k)ing. While the recently published methods yield impressive results in the absence of noise, to the best of the authors' knowledge, none of the methods address the noise problems *explicitly*, with the exception of [9, 10]. Note also that the proposed algorithm readily generalizes to irregular sampling patterns (e.g. [11]) or with a choice of color filters other than red, green, and blue.

Given the page constraints, we only present experimental results on demosaicing. We assume that the image sensor, arranged in RGB Bayer pattern [12], is corrupted by white additive Gaussian noise. As the test images, we use widely available 24-bit color images often referred to as the "clown" and "lena," and the MSE of the output images are shown in table 1 (treating all color components equally). In the presence of noise, the proposed algorithm clearly outperforms demosaicing [13], demosaicing followed by denoising [13, 1], and method in [9] that perform demosaicing and denoising simultaneously. Zoomed portion of the output images ($\sigma_n = 25$), shown in figure 1, demonstrate that the proposed algorithm reduces noise while preserving the sharpness of the image details, and introduces less noticeable artifacts when compared to the alternatives. Note that the MSE increases after several iterations partially due to overfitting by the MLE and approximate nature of our current implementation.

5. CONCLUSION

We presented a new framework for coupling the EM algorithm with the Bayesian hierarchical modeling of wavelet coefficients of image signals. The E-step in the algorithm yields an estimate of the noise-free pixel values, even if only a subset of noisy pixels are observed. The M-step finds the maximal likelihood estimates (MLE) of the model hyper-parameters. The proposed method unifies the statistical treatments of the missing pixels with that of the noisy observation data, thereby yielding results potentially substantially better than when interpolation and denoising are performed independently.

Table 1. The MSE of the demosaiced "clown" and "lena" images corrupted by white noise with standard deviation σ_n.

method	clown		lena	
	$\sigma_n = 0$	$\sigma_n = 25$	$\sigma_n = 0$	$\sigma_n = 25$
method in [13]	50.56	492.08	26.47	578.82
methods in [13]+[1]	50.56	203.92	26.47	92.13
method in [9]	105.95	194.00	36.17	95.00
proposed, $t = 1$	45.77	163.42	21.94	83.64
proposed, $t = 2$	47.30	151.51	19.81	76.21
proposed, $t = 3$	52.54	156.48	20.26	82.43
proposed, $t = 4$	61.48	160.97	21.30	86.17
proposed, $t = 5$	70.35	165.04	22.28	89.74

6. REFERENCES

[1] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixture of Gaussians in the wavelet domain," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[2] L. Sendur and I.W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Processing*, vol. 50, 2002.

[3] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 5, pp. 545–557, May 2002.

[4] J. Portilla, "Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain," *IEEE Int'l Conf on Image Processing*, pp. 1217–1220, October 2004.

[5] J. Portilla, "Blind non-white noise removal in images using Gaussian scale mixtures in the wavelet domain," *Proc. of the IEEE Benelux Signal Processing Symposium*, pp. 17–20, April 2004.

[6] G.J. McLachelan and T. Krishnan, *The EM Algorithm and Extentions*, John Wiley & Sons, Inc., 1997.

[7] X.L. Meng and D. van Dyk, "The EM algorithm–An old folk song sung to a fast new tune (with discussions)," *J.R. Statist. Soc., Ser B*, vol. 59, pp. 511–567, 1995.

[8] C. Liu and D.B. Rubin, "The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence," *Biometrika*, vol. 81, pp. 633–648, 1994.

[9] K. Hirakawa and T.W. Parks, "Joint demosaicing and denoising," *IEEE Int'l Conf on Image Processing*, September 2005.

[10] J. Portilla, D. Otaduy, and C. Dorronsoro, "Low-complexity linear demosaicing using joint spatial-chromatic image statistics," *IEEE Int'l Conf on Image Processing*, September 2005.

[11] M. Parmar and S.J. Reeves, "A perceptually based design methodology for color filter arrays," *IEEE Int'l Conf on Acoustics, Speech, and Signal Processing*, 2004.

[12] B.E. Bayer, "Color imaging array," US Patent 3 971 065, 1976.

[13] B.K. Gunturk, Y. Altunbasak, and R.M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 997–1013, September 2002.