SPIE's 45th Annual Meeting, Internat. Symp. on Optical Science and Technology, August 2000, San Diego, CA

Multiplicative Multiscale Image Decompositions: Analysis and Modeling

Justin K. Romberg, Rolf Riedi, Hyeokho Choi, and Richard G. Baraniuk

Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA

ABSTRACT

Multiscale processing, in particular using the wavelet transform, has emerged as an incredibly effective paradigm for signal processing and analysis. In this paper, we discuss a close relative of the Haar wavelet transform, the *multiscale multiplicative decomposition*. While the Haar transform captures the differences between signal approximations at different scales, the multiplicative decomposition captures their ratio. The multiplicative decomposition has many of the properties that have made wavelets so successful. Most notably, the multipliers are a sparse representation for smooth signals, they have a dependency structure similar to wavelet coefficients, and they can be calculated efficiently.

The multiplicative decomposition is also a more natural signal representation than the wavelet transform for some problems. For example, it is extremely easy to incorporate positivity constraints into multiplier domain processing. In addition, there is a close relationship between the multiplicative decomposition and the Poisson process; a fact that has been exploited in the field of photon-limited imaging. In this paper, we will show that the multiplicative decomposition is also closely tied with the Kullback-Leibler distance between two signals. This allows us to derive an n-term KL approximation scheme using the multiplicative decomposition.

1. INTRODUCTION

Multiscale signal and image processing, in particular the wavelet transform, has enjoyed tremendous success in recent years. In particular, wavelets are responsible for state-of-the-art algorithms for compression, segmentation, denoising, and many other applications in statistical signal processing.[1,2] In this paper, we will consider another multiscale transformation closely related to the Haar wavelet transform: the multiscale multiplicative decomposition. The multiplicative decomposition is similar to the Haar transform in that it decomposes the image at a nested set of scales, but for many problems it provides a more natural representation.

Most of the discussion in this paper with focus around multiplicative decompositions for 1-D signals. Although extending the multiplicative decomposition to two dimensions is not quite as straightforward as with, say, the Haar transform (see Section 2.3), most of the results are the same.

To illustrate the multiplicative decomposition, we compare it to its close (and perhaps more familiar) relative the Haar transform. Let $x = (x_0, x_1, \ldots, x_{N-1}), N = 2^L$ be a 1-D signal. We form the (un-normalized) Haar scaling coefficients $u_{j,k}$ by taking

$$\begin{aligned} u_{L,k} &= x_k & & k = 0, \ldots, N-1 \\ u_{j,k} &= u_{j+1,2k} + u_{j+1,2k+1} & & j = L-1, \ldots, 0 \quad k = 0, \ldots, 2^j - 1 \end{aligned} \tag{1}$$

The Haar transform divides the signal into dyadic intervals at each scale and calculates the (renormalized) average of the signal over these intervals. The scaling coefficients can easily be arranged into a binary tree (see Figure 1). Each node represents a scaling coefficient $u_{j,k}$ with *children* $u_{j+1,2k}, u_{j+1,2k+1}$ that further refine the signal in the same spatial location (we will refer to $u_{j,k}$ as the *parent* of $u_{j+1,2k}$ and $u_{j+1,2k+1}$).

The Haar wavelet coefficients $w_{j,k}$ are calculated by taking the difference (again, renormalized) between each pair of scaling coefficients that share the same parent:

$$w_{j,k} = 2^{(L-j)/2}(u_{j+1,2k} - u_{j+1,2k+1}) \quad j = L-1, \ldots, 0 \quad k = 0, \ldots, 2^j - 1 \tag{2}$$

Email: jrom, riedi, choi, richb@rice.edu. Web: www.dsp.rice.edu. This work was supported by the National Science Foundation, grant CCR-9973188, DARPA/AFOSR, grant F49620-97-1-0513, and Texas Instruments.

Since the Haar coefficients tell us how to refine the local averages from scale to the next, each one can be associated with an edge in the binary tree in Figure 1. The Haar coefficients represent *additive* refinements; we have $u_{j+1,2k} = u_{j,k} + 2^{(L-j)/2} w_{j,k}$ and $u_{j+1,2k+1} = u_{j,k} - 2^{(L-j)/2} w_{j,k}$.

In contrast, the multipliers $m_{j,k}$ are calculated by taking the ratio of each left (by convention) child over its parent

$$m_{j,k} \quad = \quad \frac{u_{j+1,2k}}{u_{j,k}} \quad j = L-1, \dots, 0 \quad k = 0, \dots, 2^j - 1 \tag{3}$$

(we only need to store one of these ratios, the ratio of a right child over its parent is given by $1 - m_{j,k}$). Each multiplier can also be associated with an edge in the binary tree. The multipliers (as the name suggests) represent multiplicative refinements; $u_{j+1,2k} = m_{j,k} u_{j,k}$ and $u_{j+1,2k+1} = (1 - m_{j,k}) u_{j,k}$. Since mass is preserved from scale to scale (up to a normalization constant), we can interpret $m_{j,k}$ as the ratio of the total mass that is "pushed" to the left as we move to a finer scale.

Although closely related to the Haar transform, the multiplicative decomposition is a different representation of the signal. A moments reflection shows that there is not an easy relationship between the $w_{j,k}$ and $m_{j,k}$. In particular, $\log m_{j,k} \neq w_{j,k}$ for a given signal.

The multiplicative decomposition has several properties that make it a more natural representation than the Haar transform in certain situations. For example, when the signals of interest are positive, e.g. images, we would like to constrain any processing to produce a result with all positive values. This constraint is perfectly natural in the multiplicative domain: if all the multipliers are between 0 and 1, the resulting signal will be positive. This is not true of the Haar transform; thresholding the Haar coefficients of a positive valued signal can give a result with values below zero.

The Haar wavelet transform (and wavelet transforms in general) owes much of its success in image processing to two basic facts. First, it provides a sparse decomposition for real-world images. Most of the image is represented by a few large Haar coefficients, with the rest being close to zero. Second, it is very easy to capture the statistical structure of images in the Haar domain. Simple statistical models with local dependencies have yielded state-of-the-art algorithms in denoising,[3,4] compression,[5] segmentation,[6] and many other problems in statistical image processing. Because of its close relationship to the Haar transform, the multipliers also enjoy these properties, as can be seen in Figure 3. Representation of real-world images is as natural in the multiplicative domain as it is in the Haar domain.

The multiplicative decomposition can also be used to model Poisson processes. This is useful when analyzing images that have been formed by photon-limited imaging; images with Poisson noise. As discussed in Section 3, the multiplicative decomposition retains the decoupling between the signal and the Poisson noise. Since images themselves can be modeled well in the multiplicative domain, we can solve denoising/estimation problems using powerful Bayesian techniques.[7,8]

Another useful property of the multiplicative decomposition is that the KL distance between two signals, given by

$$d_{KL}(x,y) = \sum_{i=0}^{N-1} x_i \log \frac{x_i}{y_i} + \sum_{i=0}^{N-1} (y_i - x_i). \tag{4}$$

and described in Section 5, can be easily expressed in terms of their multipliers. Using this fact, we can derive an n-term approximation algorithm using the multipliers; it is easy to choose which multipliers represent the most information about the image. Since the decomposition is sparse for real-world images, an image is well approximated using a small percentage of its multipliers (see Figure 4).

We begin in Section 2 by describing the multiplicative decomposition and its relationship to the Haar transform. Section 3 discusses the statistics of multipliers, and shows that the multiplicative decomposition is well suited for Poisson processes. In Section 4, we introduce the KL signal distance as a measure of the distinctness of two images. In Section 5, we show that the KL signal distance is easily expressed using the multiplicative decomposition and use this fact to derive an n-term approximation algorithm.

2. MULTIPLICATIVE DECOMPOSITIONS

The multiplicative decomposition is a multiscale signal representation. Like the Haar transform, the multiplicative decomposition describes how piecewise constant approximations of the signal evolve from scale to scale. While the Haar coefficients describe the *additive* changes in the approximation from scale to scale, the multiplicative decomposition describes the change by a *multiplicative cascade.*

2.1. The multiplicative and Haar decompositions

We start by recalling the unnormalized Haar scaling coefficients of a discrete, 1-D, finite length signal $x = (x_0, x_1, \ldots, x_{N-1})$, $N = 2^L$:

$$\begin{array}{rcll} u_{L,k} & = & x_k & k = 0, \ldots, N-1 \\ u_{j,k} & = & u_{j+1,2k} + u_{j+1,2k+1} & j = L-1, \ldots, 0 \\ & & & k = 0, \ldots, 2^j - 1. \end{array} \tag{5}$$

The scaling coefficients are found by subdividing the signal into dyadic sets at each scale, and finding the total mass over each set. As such, it is convenient to represent the $u_{j,k}$ as nodes in a binary tree, as shown in Figure 1. By summing the scaling coefficients of the two child nodes, we calculate the scaling coefficient of the parent. The leaves of the tree, the nodes at scale L, represent the signal itself.

To generate the Haar coefficients from the scaling coefficients, we set

$$\begin{array}{rcll} w_{j,k} & = & C_j(u_{j+1,2k} - u_{j+1,2k+1}) & j = L-1, \ldots, 0 \\ & & & k = 0, \ldots, 2^j - 1 \end{array} \tag{6}$$

where $C_j = 2^{(L-j)/2}$. The Haar coefficients are associated with the edges between the nodes on the binary tree (Figure 1). The result of adding $w_{j,k}/C_j$ to the parent scaling coefficient $(u_{j,k})$ is the left child scaling coefficients $(u_{j+1,2k})$. The right child coefficient is given by subtracting $w_{j,k}/C_j$ from the parent coefficient. Thus, the Haar coefficients can be thought of as an additive cascade. To calculate a leaf x_i, we can take the scaling coefficient uj, k of any ancestor of x_i, and add to it all the Haar coefficients (or negative Haar coefficients) associated with edges on the path from $u_{j,k}$ to x_i.

As the structure of equations (5) and (6) suggests, calculating the Haar coefficients is extremely efficient. The $O(N)$ computational complexity is a significant factor of the popularity of the Haar transform (and of wavelet transforms in general).

To generate the multiplicative decomposition of x, we simply take the ratio

$$m_{j,k} \quad = \quad \frac{u_{j+1,2k}}{u_{j,k}} \quad j = L-1, \ldots, 0 \quad k = 0, \ldots, 2^j - 1 \tag{7}$$

where we use the (admittedly odd) convention $0/0 = 1/2$. To reconstruct x from its multipliers, we also need knowledge of the top scaling coefficient $u_{0,0}$. The multiplicative decomposition consists of $u_{0,0}$ and the $m_{j,k}$ for $j = L-1, \ldots, 0$ and $k = 0, \ldots, 2^j - 1$.

The multipliers $m_{j,k}$ can also be associated with edges on the tree of scaling coefficients (Figure 1). To calculate the left child scaling coefficient $u_{j+1,2k}$, we simply multiply the parent scaling coefficient $u_{j,k}$ by the multiplier $m_{j,k}$ on the edge between the two. To calculate the right child, we multiply the parent by $1 - m_{j,k}$. The multiplicative decomposition represents a multiplicative cascade. To calculate a leaf x_i, we can take the scaling coefficient $u_{j,k}$ of any ancestor of x_i, and multiply it by all the $m_{j',k'}$ (or $1 - m_{j',k'}$) associated with the edges on the path from $u_{j,k}$ to x_i. The multiplicative decomposition also has $O(N)$ computational complexity, so no efficiency has been lost.

The scaling coefficients at a scale J can be used to construct a piecewise constant approximation $\hat{x}^J$ of x

$$\hat{x}_i^J \quad = \quad 2^{-j} u_{J,k}, \qquad k2^{L-j} \leq i < (k+1)2^{L-j}. \tag{8}$$

The approximation $\hat{x}^J$ is the average of x over a dyadic index set. As we saw above, increasing the scale j splits the scaling coefficient into two children coefficients. This mean that each region in which $\hat{x}_i^J$ is constant gets broken up into two smaller piecewise constant regions. Whereas on the tree we add a Haar coefficient to the parent scaling coefficient, a series of pulses with heights proportional to the Haar coefficient are added to $\hat{x}^J$ resulting in $\hat{x}^{J+1}$ (this is illustrated in Figure 2). Using the multiplicative decomposition, this refinement from $\hat{x}^J$ to $\hat{x}^{J+1}$ is equivalent to point-wise multiplying $\hat{x}^J$ by a series of pulses with values between 0 and 1 (also shown in Figure 2).

2.2. Basic properties of multipliers

When the signal is flat in a dyadic region, there is no refinement from one scale to the next, and the corresponding "no-information" multipliers are $1/2$ (note that the "no-information" Haar coefficients are 0). Jumps in the signal correspond to multipliers away from $1/2$. However, unlike the Haar transform, the multiplicative decomposition is nonlinear. As a consequence, jumps of equal height are not represented by the same multipliers. For example, if $u_{j,k}=3$ with $u_{k+1,2k}=1$ and $u_{k+1,2k+1}=2$, then $w_{j,k}=-C_j$ and $m_{j,k}=1/3$. If, however, $u_{j,k}=5$ with $u_{k+1,2k}=2$ and $u_{k+1,2k+1}=3$, we still have $w_{j,k}=-C_j$ but $m_{j,k}=2/5$. A jump of a certain size appears more drastic to the multiplicative decomposition if it happens at small signal values. We will exploit this property later in Section 5.2.

When the signal to be analyzed is predominantly smooth (much of the signal is well approximated by piecewise constants), most of the multipliers will be close to $1/2$. As discussed in the next section, we see this behavior in images (the corresponding behavior for the Haar wavelet coefficients is well documented[1]). This suggests that a few multipliers carry most of the image information, an idea we will develop further in Section 5.1.

2.3. Multiplicative decompositions in 2-D

Extending the multiplicative decomposition to higher dimensions is not as straightforward as for the Haar transform. In this section, we will extend the multiplicative decomposition in two different way, and discuss the properties of each. We will refer to 2-D signals as images, since this is the primary application for higher dimensional signal processing.

In the analog of (5), the 2-D scaling coefficients are given by sums of the image over dyadic squares. The coefficients are best represented on a quad tree, with each $u_{j,k}$ breaking up into four smaller coefficients $u_{j+1,4k}, u_{j+1,4k+1}, u_{j+1,4k+2}, u_{j+1,4k+3}$ as we move to finer and finer scale.

In the first representation, which we'll call the non-separable 2-D multiplicative decomposition, the multipliers are calculated by taking the ratio of each child with the parent

$$m^{\text{non}}_{j,k,0} = \frac{u_{j+1,4k}}{u_{j,k}} \tag{9}$$

$$m^{\text{non}}_{j,k,1} = \frac{u_{j+1,4k+1}}{u_{j,k}} \tag{10}$$

$$m^{\text{non}}_{j,k,2} = \frac{u_{j+1,4k+2}}{u_{j,k}} \tag{11}$$

$$m^{\text{non}}_{j,k,3} = \frac{u_{j+1,4k+3}}{u_{j,k}} = 1 - m^{\text{non}}_{j,k,0} - m^{\text{non}}_{j,k,1} - m^{\text{non}}_{j,k,2} \tag{12}$$

for $j=1\dots,L$, $k=0,\dots,4^j-1$. We have used the notation $m^{\text{non}}_{j,k,\ell}$ for the multiplier associated with the edge on the quad tree that connects $u_{j,k}$ and $u_{j+1,4k+\ell}$. The transform saves the first three coefficients (by convention), since the last is redundant.

Again, for positive valued images, we have $m^{\text{non}}_{j,k,\ell} \in [0,1]$ for all j,k,ℓ. Since we must also have $m^{\text{non}}_{j,k,0} + m^{\text{non}}_{j,k,1} + m^{\text{non}}_{j,k,2} \leq 1$, we have an implicit dependence between the three multipliers we store. When we process the multipliers, it is not enough to simply keep each ones value between $[0,1]$ independently, we must also make sure that the three multipliers that share the same parent sum to less than 1. Hence, do any kind of processing with the $m^{\text{non}}_{j,k,\ell}$ is tricky, since we must account for this dependence.

To address this problem, we introduce the "separable" 2-D multiplicative decomposition *. Here, we first break the dyadic square into right and left halves, and then each of these into top and bottom halves:

$$m_{2j,k} = \frac{u_{j+1,4k} + u_{j+1,4k+1}}{u_{j,k}} \tag{13}$$

$$m_{2j+1,2k} = \frac{u_{j+1,4k}}{u_{j+1,4k} + u_{j+1,4k+1}} \tag{14}$$

$$m_{2j+1,2k+1} = \frac{u_{j+1,4k+2}}{u_{j+1,4k+2} + u_{j+1,4k+3}}. \tag{15}$$

*The word separable is a bit of a misnomer here. Although this representation separates into two stages the right-left and up-down splits of the dyadic squares, these operations are not reversible; the decomposition that splits left-right first is slightly different than the decomposition that splits up-down first.

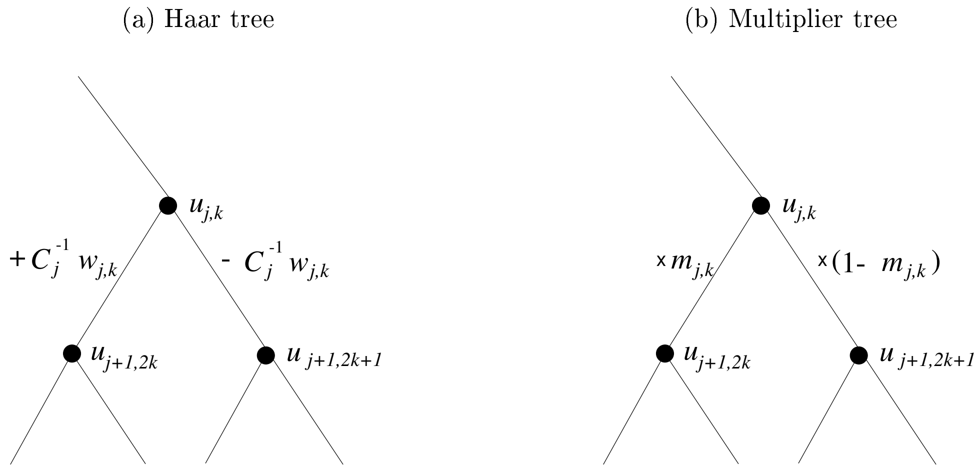

Figure 1. *(a) The Haar transform can be naturally represented with a binary tree with the scaling coefficients on the nodes and the Haar wavelet coefficients on the edges. (b) The multiplicative decomposition again has the scaling coefficients on the nodes, but now the edges are associated with multipliers.*

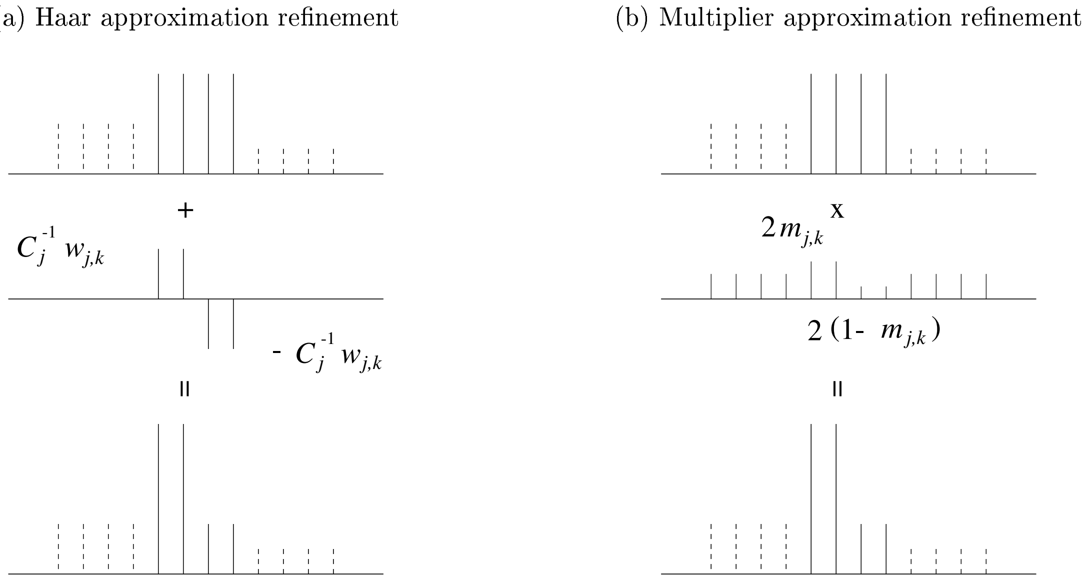

Figure 2. *Function refinement moving down one scale. (a) For the Haar transform, this is captured by adding in a pulse with height proportional to the size of the Haar wavelet coefficient. (b) For the multiplicative decomposition, we point-wise multiply the approximation at one scale by another piecewise constant function with values proportional to the multipliers.*

As the indexing in (13)–(15) suggests, this basically flattens the quad tree into a binary tree, with each level transition in the quad tree being represented by two level transitions (one for left-right, one for up-down) in the binary tree. The separable 2-D and the 1D multiplicative decompositions share the same properties.

In the remainder of the paper, whenever we refer to the multiplicative decomposition of an image, we will be referring to the separable decomposition in (13)–(15).

3. STATISTICAL MODELING AND MULTIPLIERS

One of the main advantages of a multiscale representation of an image is that it makes the image's salient statistical properties easier to characterize. For example, models that treat the wavelet coefficients of images as independent[9,5] or as having simple, local dependencies[3,4,6] have produced tremendously successful algorithms for many problems in statistical image processing.

3.1. Poisson noise and the multiplicative decomposition

Although multiscale decompositions provide a domain in which the signal is easy to describe, they don't necessarily do the same for noise processes. In situations where the algorithm is dealing with corrupted data (such as denoising or any kind of processing with photon-limited images), having a tractable noise model is critical. For example, wavelet based denoising results, in both theory and in practice, rely heavily on the noise being Gaussian. Independent Gaussian noise has a very simple wavelet-domain description; if an image is corrupted by additive white Gaussian noise (AWGN), the wavelet coefficients are also corrupted by AWGN. Other types of noise, e.g. Poisson, Gamma, or Cauchy, do not have nice representations in the wavelet domain.

Just as the Haar transform is a natural representation for Gaussian noise, the multiplicative decomposition, as described in Section 2, is natural for Poisson noise.[7,10] To make these notions more precise, consider a 1-D signal $x = (x_0, x_1, \ldots, x_{N-1})$ with scaling coefficients $u_{j,k}$, Haar coefficients $w_{j,k}$ and multipliers $m_{j,k}$. We'll consider two "noisy" versions of x. Let y be an observation of x in additive white Gaussian noise;

$$y_i \sim \mathcal{N}(x_i, \sigma^2) \tag{16}$$

with the y_i independent given the x_i. By Parseval's theorem, the Haar coefficients of y, denoted $(Hy)_{j,k}$, are distributed

$$(Hy)_{j,k} \sim \mathcal{N}(w_{j,k}, \sigma^2). \tag{17}$$

The independence of the noise in the time domain has been preserved in the transform domain. Let z be an observation of x in Poisson noise,

$$z_i \sim \text{Poisson}(x_i) \tag{18}$$

with the z_i independent given the x_i. It is clear that the distribution of the Haar transform Hz can not be expressed nicely as in (17). However, we can write distribution of Mz, the multiplicative decomposition of z, by

$$u_{0,0} \sim \text{Poisson}\left(\sum x_i\right) \tag{19}$$

$$(Mz)_{j+1,2k} | u_{j,k} \sim \text{Binomial}(u_{j,k}, m_{j+1,2k}) \tag{20}$$

with the $(Mz)_{j,k}$ independent.[10]

The factorization given by (19)–(20) shows that the multipliers of a Poisson process are independent. This makes the multiscale multiplicative decomposition natural for analyzing Poisson processes, just as the Haar transform is natural for analyzing Gaussian processes.

3.2. Image modeling with the multiplicative decomposition

In order to take a Bayesian approach to image processing problems, we need a statistical model of the image itself as well as the corruption process. The statistical structure of real-world images can be as readily characterized in the multiplier domain as in the wavelet domain. Indeed, because the multiplicative decomposition and the Haar transform are so intimately related, the multipliers and Haar coefficients of images share the same statistical "grammar" (dependency structure). Since this structure is much easier to characterize than dependencies in the spatial domain, and since this statistical grammar (in the context of the wavelet domain) has been exploited in the past to create powerful transform-domain algorithms,[3–6] the multiplicative decomposition is a promising domain in which to build an image model.

The two key properties of the multipliers of real-world images, illustrated in Figure 3, are

1. The decomposition is sparse (most of the multipliers are clustered around 1/2).
2. The multipliers exhibit persistence across scale (large values— values that are much smaller or much greater than 1/2— tend to exist on connected paths in the tree in Figure 1).

(a) image (b) multiplicative decomposition (c) histogram

Figure 3. *(a) Cameraman image and (b) its multiplicative decomposition. A histogram of the multipliers at one scale is given in (c). Notice that most of the multipliers are close to* $1/2$*. Also note the persistence of "large" multipliers (those not close to* $1/2$*) near the edges in the image.*

These properties suggest a relatively simple statistical model for the multiplicative decomposition of an image following the hidden Markov modeling paradigm in Crouse et al.[3] Since most of the multipliers will be close to $1/2$, and all of them lie between 0 and 1, the marginal distributions can be given by the symmetric beta ($B(p,p)$, with a large value of p for sparsity) distribution as in Timmermann and Nowak.[7] The dependencies between multipliers at different scales can be specified by a conditional distribution on the parameter p; $Q(p_1, p_2) =$ the probability that a multiplier is distributed $B(p_1, p_1)$ given that its parent is distributed $B(p_2, p_2)$. We propose this model only as an illustration of how to capture the structure of image multipliers, and we don't pursue it further in the sequel.

4. KULLBACK-LEIBLER DISTANCE AND METRICS FOR SIGNAL SPACES

Section 3 and Kolaczyk[10] show that different kinds multiscale decompositions are appropriate for processing signals with different kinds of degradation (noise processes). In this section, we will show that the appropriate metric to use on the image space (a measure of the distinctness of two images) also depends on the type of degradation. We derive this metric by looking at the Kullback-Leibler distance between the distributions of the noisy observations of two signals.

We start by recalling the notion of Kullback-Leibler (KL) distance, a metric that quantifies the distinctness of two discrete probability distributions $p(x)$ and $q(x)$ both with support S:

$$D(p||q) = \sum_{x \in S} p(x) \log \frac{p(x)}{q(x)}. \tag{21}$$

The KL distance $D(p||q)$ has a number of interpretations.[11] Perhaps the most notable comes from detection theory via Stein's Lemma. Consider the following classic problem: given a set of independent and identically distributed samples $\{x_1, x_2, \ldots, x_n\}$ determine which of the distributions (p or q) is more likely to have produced the data. We do this simply by finding the larger of $\prod_i^n p(x_i)$ and $\prod_i^n q(x_i)$. The Stein Lemma tells us that the probability that we misclassify samples from $q(x)$ as being from $p(x)$ (termed the "miss probability" P_M) decays as $n \to \infty$ with an exponent proportional to the KL distance between p and q:

$$P_M \sim e^{-C \cdot n D(p||q)}. \tag{22}$$

Equation (22) tells us that the KL distance is fundamentally related to our ability to distinguish between data sets drawn from different distributions. If the KL distance between two distributions is small, then we can think of them as being similar.

We can leverage this idea to assign a metric on a space of signals (1-D in this discussion, but the analysis can be extended to images) that have been corrupted with noise. Given a space of signals X, we want to assign a number $d(s,t)$ to every pair of signals $s, t \in X$ that quantifies the similarity of s and t. For example, if X is a set of images, we want $d(s,t)$ to be small if s and t look alike, and large if they look very different. Since we are observing noisy versions of the signals, $d(s,t)$ should be small if it is hard to distinguish statistically the observations of s and observations of t.

To make this more precise, consider two signals $s = (s_0, s_1, \ldots, s_{N-1})$ and $t = (t_0, t_1, \ldots, t_{N-1})$ observed under less than ideal conditions. Since we are making observations of these signals in the presence of noise, a measurement of s is a random variable distributed $p_s(x)$ and a measurement of t is distributed $p_t(x)$. The KL distance between $p_s(x)$ and $p_t(x)$ dictates (roughly) how easy it is to confuse observations of s and t, so a reasonable metric is $d(s,t) = D(p_s(x)||p_t(x))$. Of course, since $D(p||q) \neq D(q||p)$ in general, $d(s,t)$ will not be a metric in the strict mathematical sense.

For example, if we make measurements of s and t in the presence of additive white Gaussian noise with variance σ^2, we have

$$d(s,t) = D(\mathcal{N}(s, \sigma^2 I)||\mathcal{N}(t, \sigma^2 I)) = \frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (s_i - t_i)^2. \tag{23}$$

Equation (23) show that for signals observed in Gaussian noise, the KL distance between the observations induces the normal Euclidean (L_2) metric on the space X of signals.

Now assume that s and t are corrupted by Poisson noise (we'll assume that $s_i, t_i > 0$). Each component of an observation of s (and likewise t) is a Poisson random variable with parameter s_i (and t_i). For the induced metric on the signal space we have

$$d(s,t) = D(\text{Poisson}(s)||\text{Poisson}(t)) = \sum_{i=0}^{N-1} s_i \log \frac{s_i}{t_i} + \sum_{i=0}^{N-1} (t_i - s_i). \tag{24}$$

The KL distance between the observations becomes a sort of deterministic KL on the underlying signals. We will call (24) the *KL signal distance* (to distinguish it from the classical KL distance for probability distributions) between s and t. Note that in general, $d_{KL}(x,y) \neq d_{KL}(y,x)$, so (24) does not define a "distance" in the strict mathematical sense. However, it can still be used as a measure of the similarity of two signals.

Since Stein's lemma gives a fundamental relationship between the KL distance of two processes and the ability of an optimal detector to distinguish between them, we expect that the metric ($d(s,t)$) the KL distance imposes on the parameter space (signal space X in our case) will do likewise. Equation (24) tells us that if we are observing signals created by a Poisson processes, the KL signal distance determines our ability to distinguish between them. In Section 5, we will explore the properties of (24) as a metric for images.

5. MULTIPLICATIVE DECOMPOSITIONS AND THE KL SIGNAL DISTANCE

In Section 4 we saw that the KL signal distance (24) is appropriate for images that will be observed as Poisson processes. In this section, we show that the multiplicative decomposition provides a convenient way to express the KL signal distance between two signals; we can write down the contribution of each multiplier individually. Using this fact, we can form n-term approximations to signals by finding the n multipliers of a signal that contribute most to the KL signal distance.

Much of the success of the wavelet transform in image processing is based on the fact that the wavelet coefficients are a sparse representation for images. Since a small number of large coefficients represent most of the information about an image, we can approximate the image by setting the small coefficients to zero and inverting the wavelet transform. A distance metric measures the effectiveness of this procedure; the smaller the distance between the original image and the approximate (for a given order of approximation), the more the approximation resembles the original. For an image x and a given n, the n-term approximation is the image that is closest (under the given metric) to x using n of its coefficients.

For these types of approximations to be feasible, there must be an easy way to choose the n coefficients that will give the best approximation. To do this, the distance metric must be easily expressible in terms of the transform coefficients. If this is the case, then we can easily deduce the effect of including each coefficient, and the selection procedure simply becomes choosing the coefficient at each step that decreases the distance the most.

The most commonly used distance metric for n-term approximation is L_2; the n-term approximate finds the n coefficients that minimize the sum of the square of the differences between the original and approximate images. The L_2 norm is easily expressed in terms of the Haar coefficients. Indeed, by Parseval's theorem (and the linearity of the Haar transform), the L_2 error is the same in the Haar domain as it is in the spatial domain. This means that choosing the n-terms for the approximation is easy: we just choose the n largest Haar coefficients.

5.1. n-term KL approximation using multipliers

The L_2 metric can not be written easily in terms of the multiplicative decomposition of an image. However, another useful metric between two images, namely the KL signal distance introduced in Section 4, can be simply expressed using the multipliers. Recalling (24), the KL signal distance between two discrete signals is given in the spatial domain by

$$d_{KL}(x,y) = \sum_i x_i \log \frac{x_i}{y_i} + \sum_i y_i - x_i. \tag{25}$$

A quick example will illustrate the ease of calculating the KL metric between two 1-D signals in terms of their multiplicative distributions. Let signals $x = \{x_0, x_1, x_2, x_3\}$ and $x' = \{x'_0, x'_1, x'_2, x'_3\}$ have multiplicative decompositions $\{u_{0,0}, m_{0,0}, m_{1,0}, m_{1,1}\}$ and $\{u'_{0,0}, m'_{0,0}, m'_{1,0}, m'_{1,1}\}$ respectively, following the notation from Section 2. For ease of exposition, we'll assume that $u_{0,0} = u'_{0,0}$. By simply plugging in the reconstruction equations for x_i and x'_i into equation (25), we have

$$\begin{aligned} d_{KL}(x,x') \quad = \quad & m_{0,0}m_{1,0}(\log m_{0,0}m_{1,0} - \log m'_{0,0}m'_{1,0}) + \\ & m_{0,0}(1-m_{1,0})(\log m_{0,0}(1-m_{1,0}) - \log m'_{0,0}(1-m'_{1,0})) + \\ & (1-m_{0,0})m_{1,1}(\log(1-m_{0,0})m_{1,1} - \log(1-m'_{0,0})m'_{1,1}) + \\ & (1-m_{0,0})(1-m_{1,1})(\log(1-m_{0,0})(1-m_{1,1}) - \log(1-m'_{0,0})(1-m'_{1,1})). \end{aligned} \tag{26}$$

Expanding (26) and collecting like log terms we write

$$d_{KL}(x,x') = k(m_{0,0}||m'_{0,0}) + m_{0,0}k(m_{1,0}||m'_{1,0}) + (1-m_{0,0})k(m_{1,1}||m'_{1,1}) \tag{27}$$

where

$$k(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{(1-p)}{(1-q)} \tag{28}$$

is the KL distance between two Binomial distributions with different parameters. In (27) we have written the KL signal distance as a sum of positive terms, each of which depends only on one of the multipliers of x'. For length N signals, we extend (27) to

$$d_{kl}(x,x') = \sum_{j,k} \frac{u_{j,k}}{u_{0,0}} k(m_{j,k}||m'_{j,k}) \tag{29}$$

with $\frac{c_i}{c_1}$ being the product of all the multipliers on the path down the tree to branch i.

Let Λ_n be the set of indices of multipliers included in an n-term approximate x' of x. The distance between a signal x and its x' is given by (29) with

$$m'_{j,} = \begin{cases} m_{j,k} & (j,k) \in \Lambda_n \\ \frac{1}{2} & (j,k) \notin \Lambda_n \end{cases} \tag{30}$$

To pick the n multipliers that minimize this distance, we want to choose the multipliers corresponding to the largest terms in (29). Since each of the terms depends on only one of the $m'_{j,k}$, the selection of a $m'_{j,k}$ (changing it from 1/2 to $m_{j,k}$) does not affect the other terms in the sum. As a result, we have that $\Lambda_n \subset \Lambda_{n+1}$ and we can easily measure each multiplier's contribution to the distance. By noticing that

$$m'_{j,k} = m_{j,k} \Rightarrow k(m_{j,k}||m'_{j,k}) = 0 \tag{31}$$

Figure 4. *n-term approximation of the 256×256 cameraman image ($n = 2103$). The original cameraman image is shown in (a). (b) L_2 2103-term approximation using Haar coefficients, $||\hat{x} - x||_2 = 581.8$. (c) KL 2103-term approximation using multipliers, $d_{KL}(x, \hat{x}) = 57940.1$.*

and

$$m'_{j,k} = 1/2 \Rightarrow k(m_{j,k}||m'_{j,k}) = h(m_{j,k}) + 1 \tag{32}$$

where $h(p) = p \log p + (1-p)\log(1-p)$ is the negative entropy of a binomial random variable with parameter p, we can summarize n-term approximation with multipliers of a signal x with the following steps

1. Calculate the multipliers $m = \mathcal{M}(x)$ using the method outlined in Section 2
2. Calculate $T_{j,k} = \frac{u_{j,k}}{u_{0,0}}(h(m_{j,k}) + 1)$ for each multiplier $m_{j,k}$
3. Include in the index set Λ_n the values of (j,k) corresponding to the n largest $T_{j,k}$
4. Construct d' with the rule in (30)
5. Reconstruct the approximate signal $x' = \mathcal{M}^{-1}m'$ from m'.

Because the multipliers are (empirically) sparse, we expect most of the $T_{j,k}$ to be small. Therefor, most of the distance in (29) can be made up with a few multipliers (much like most of the L_2 energy of a signal is represented by few Haar coefficients). n-term approximation results for an image are shown in Figure 4.

5.2. KL as a metric for images

We have shown that the multiplicative decomposition is a natural setting for n-term KL approximation. For this to have application in image processing, the KL signal distance must be compatible with visual perception; if the KL distance is small between a pair of images, the images should appear to be similar. In this section, we'll discuss which features in images KL discriminates between and how it compares to using L_2.

The most important aspect of an image in terms of perception is its edge structure. Indeed, the wavelet transform owes its success across the whole of image processing to this fact. Wavelet coefficients tend to be large near the edges of an image and small in the smooth regions. Since the ideal domain in which to analyze signals has a few large coefficients representing important features, wavelets are well suited to image processing.

Haar wavelet coefficients capture local differences. Therefor, the size of a Haar wavelet coefficient on an edge is proportional to the difference in grayscale values on either side of the edge. An edge jumping from 3 to 7 is represented the same way as an edge from 8 to 13. However, the human visual system picks up edges at low grayscale (white areas) better than at high grayscale (black areas). The metric we use to distinguish images should be more sensitive to the low grayscale areas.

This is precisely how the KL distance behaves. Figure 5.2(a) shows a simple image with squares in white and black areas. The square in the white region lies 10 grayscales above its white background, while the square in the

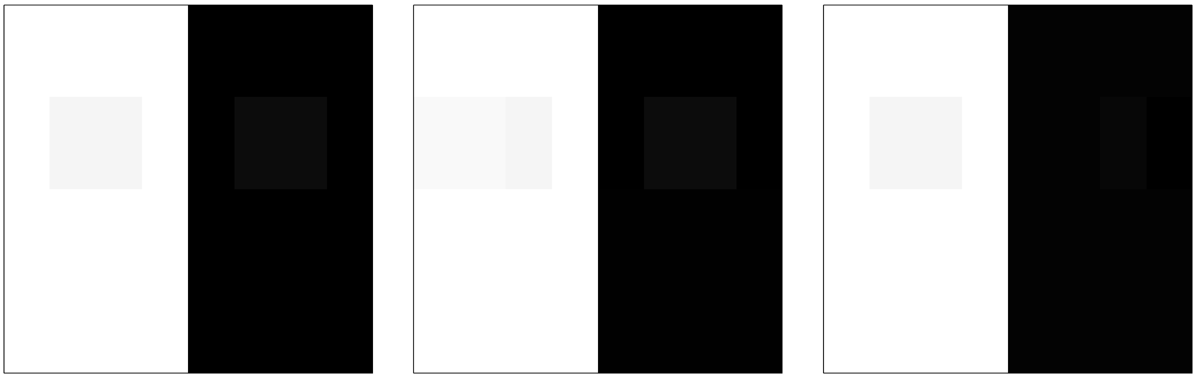

Figure 5. *(a) Squares image, (b) low order Haar L_2 approximation, (c) low order multiplicative KL approximation. Notice that the Haar approximation captures the square in the black region, as opposed to the multiplicative approximation which captures the higher contrast square in the white region.*

black region lies 20 grayscales below its black background. Notice that the light gray square in the white region is easier to see than the dark gray square in the black region. Figure 5.2(b) shows a low order Haar L_2 approximation of the image. The visually more important square in the white region is not well represented, the Haar transform treats square in the black region (with the greater size jump) with greater care. In Figure 5.2(c) we see a multiplicative KL approximation of the same order. The square in the white region is well represented; the multipliers have done a better job of representing the visually important features in the image.

6. CONCLUSIONS AND FUTURE WORK

The multiplicative decomposition, closely related to but very distinct from the Haar wavelet transform, is a multiscale representation well-suited to certain types of problems. In particular, signals represented by Poisson processes are naturally represented (and hence easily processed) in the multiplicative domain. Since the multipliers also provide a sparse decomposition of images, Bayesian modeling in the multiplicative domain offers many of the same advantages as modeling in the wavelet domain.

We have also showed that the KL distance between two positive valued signals is easily expressed by their multiplicative decompositions. Using this fact, we introduced an n-term approximation scheme using the multipliers which compared favorably to n-term approximation using Haar wavelets.

We have, however, avoided mentioning until now one real shortcoming of the multiplicative decomposition. Many wavelet based processing algorithms show greatly improved results when using smoother wavelet than Haar. While the extension in the wavelet domain is easy, It is not clear how to extend the multiplicative decomposition to include more complicated scale-to-scale transitions. This is a subject of the author's current research.

In this paper (although it is not the first to do such), we have presented the multiplicative decomposition as a paradigm for multiscale signal and image analysis. However, some of the techniques could be useful elsewhere. Specifically, given the similarity of equations (21) and (24), we can use the n-term approximation scheme described in Section 5.1 to approximate probability density functions in general. In this context, the closest KL approximation has a solid statistical meaning, so the algorithm may be of more general interest.

REFERENCES

1. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.
2. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, New York, 1992.
3. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.* **46**, pp. 886–902, Apr. 1998.
4. J. K. Romberg, H. Choi, and R. G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *Submitted to IEEE Trans. on Image Proc.* , 2000. Available at www.dsp.rice.edu/publications.

5. S. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Data Compression Conference '97*, pp. 221–230, (Snowbird, Utah), 1997.
6. H. Choi and R. G. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models," *Submitted to IEEE Trans. Image Proc. - Available at www.dsp.rice.edu/publications/* , 1999.
7. K. E. Timmermann and R. D. Nowak, "Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging," *IEEE Trans. on Info. Theory* , Apr. 1999.
8. E. D. Kolaczyk, "Bayesian multi-scale models for Poisson processes," *Journal of the American Statistical Association* , to appear 1999.
9. H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Amer. Stat. Assoc.* **92**, 1997.
10. E. D. Kolaczyk, "Some observations on the tractability of certain multi-scale models," in *Wavelets and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1999.
11. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.