Image Denoising Using Complex Wavelets and Markov Prior Models

Fu Jin, Paul Fieguth, and Lowell Winger

Dept. of Systems Design Engineering, Univ. of Waterloo,
Waterloo, Ontario, Canada N2L 3G1
{fjin, pfieguth, lwinger}@uwaterloo.ca

Abstract. We combine the techniques of the complex wavelet transform and Markov random fields (MRF) model to restore natural images in white Gaussian noise. The complex wavelet transform outperforms the standard real wavelet transform in the sense of shift-invariance, directionality and complexity. The prior MRF model is used to exploit the clustering property of the wavelet transform, which can effectively remove annoying pointlike artifacts associated with standard wavelet denoising methods. Our experimental results significantly outperform those using standard wavelet transforms and are comparable to those from overcomplete wavelet transforms and MRFs, but with much less complexity.

Keywords: image denoising, complex wavelet transform, Markov random field.

1 Introduction

Images and image sequences are frequently corrupted by noise in the acquisition and transmission phases. The goal of denoising is to remove the noise, both for aesthetic and compression reasons, while retaining as much as possible the important signal features. Traditionally, this is achieved by approaches such as Wiener filtering, which is the optimal estimator in the sense of mean squared error (MSE) for Gaussian processes. However, the Wiener filter requires stationarity and an accurate statistical model of the underlying process, these performing poorly for natural images failing these assumptions. In practice, adaptive methods [1,2] were mostly used. These methods are good in that they are fast and can effectively suppress noise for most *natural* images. More importantly their adaptivity allows them to work for non-stationary processes (it is well-known the natural images are non-stationary). The main problem with such methods is their assumption that the natural images are independent random processes, which usually is not true. For example, image textures are correlated and are successfully modelled as Gaussian MRF (GMRF).

The last decade has seen a good deal of effort in exploiting the wavelet transform to suppress noise in natural images [3,4,5,7], because of its effectiveness and simplicity. It is now well-known that wavelet transforms with some regularity have strong decorrelation ability, thus well-representing many natural images

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 73–80, 2005.

with relatively few large coefficients. So it is thus far more reasonable to assume that the wavelet coefficients are independent, than the original spatial domain pixels. This explains why good denoising results have been achieved by simply thresholding or shrinking each wavelet coefficient *independently* [3,4]. Indeed, this kind of approach has much better results than traditional methods [1,2], both subjectively and objectively. However, for natural images the wavelet transform is not quite equivalent to the ideal eigen value/Karhunen-Loeve decomposition, so some correlation (dependence) still exists among the wavelet coefficients. For example, large (in magnitude) wavelet coefficients tend to be clustered within a scale and across scales. If these characteristics could be exploited in some way for denoising, better performance might be expected. Indeed, MRF models have been used for this very reason [5,7] and significantly better results obtained, both subjectively and objectively. Specifically, pointlike artifacts associated with the independence model have been effectively suppressed.

Considering the shift-variability of standard wavelets, or the complexity of shift-invariant (undecimated) transforms, in this paper we propose to use complex wavelet transform together with a MRF model for image denoising. A different formulation of the problem is also proposed. Sec. 2 describes the problem formulation. Sec.3 introduces the basic ideas of the complex wavelets and its useful properties. Sec.4 is about the probability models we use in the paper. In Sec.5 we show some experimental results with discussions.

2 The Denoising Method

Standard wavelet-based denoising methods consist of three steps:

1. The wavelet decomposition of the image is computed:
   Given noise-free image $\underline{x}_o$ and wavelet transform H, then
   $$\begin{aligned} \underline{x} &= H\underline{x}_o \\ \underline{y} &= H\underline{y}_o = H(\underline{x}_o + \underline{w}_o) \end{aligned} \tag{1}$$
   where $\underline{y}$ are the noisy observations.
2. The obtained wavelet coefficients are modified:
   $$\underline{\hat{x}} = f(\underline{y}) \tag{2}$$
   where $f()$ denotes our proposed estimator.
3. The cleaned image is obtained from the modified wavelet coefficients by inverse wavelet transform:
   $$\underline{\hat{x}}_o = H^{-1}\underline{\hat{x}}(\underline{y}) \tag{3}$$

For the first step we need to choose the "best" wavelet transform for an application. Commonly used wavelets include Daubechies orthogonal wavelets with the lengths from 2 to 10, bi-orthogonal wavelets with symmetry and several regular (smooth) overcomplete wavelets implemented by the *a trous* algorithm . Generally speaking, overcomplete wavelets outperform the fully decimated wavelets for

signal and image denoising because they are shift-invariant. For images, wavelets with good orientation-selectivity (e.g. curvelet, ridgelet) are preferred. In this paper we use the dual-tree complex wavelet proposed by Kingsbury [8] because of its shift-invariance and orientation-selectivity properties. These properties will be shown in Section 3.

For Step two we propose to use a Bayesian decision and estimation method to modify the wavelet coefficients [5,7]. We classify the wavelet coefficients into two groups:H_0 and H_1, representing noise and signal, respectively. Then the *ith* coefficient is changed to minimize the mean square error:

$$\begin{aligned}\hat{x}_i &= E(x_i|\underline{y}) \\ &= E(x_i|\underline{y}, L_i = H_0)P(L_i = H_0|\underline{y}) \\ &\quad + E(x_i|\underline{y}, L_i = H_1)P(L_i = H_1|\underline{y})\end{aligned} \tag{4}$$

where $E()$ denotes expectation. $L_i \in \{H_0, H_1\}$ is the label of the *ith* coefficient.

In practice, we can assume $E(x_i|\underline{y}, H_0) = 0$. There are several methods to evaluate $E(x_i|\underline{y}, H_1)$. In this paper we simply set $E(x_i|\underline{y}, H_1) = y_i$. Thus we have

$$\hat{x}_i = P(L_i = H_1|\underline{y}) \cdot y_i \tag{5}$$

So to find $\hat{x}_i$ the only unknown quantity is $P(L_i = H_1|\underline{y})$. To get $P(L_i = H_1|\underline{y})$ one method is to find $P(\underline{L}|\underline{y})$ first [5,7], using the Bayesian rule

$$P(\underline{L}|\underline{y}) = \frac{P(\underline{y}|\underline{L})P(\underline{L})}{P(\underline{y})} \tag{6}$$

where $\underline{L}$ is the label field. Then based on models $P(\underline{y}|\underline{L})$ and $P(\underline{L})$ we can use stochastic sampling to find the joint probability $P(\underline{L}|\underline{y})$ and then the marginal probability $P(L_i = H_1|\underline{y})$. However, it is usually difficult to model $P(\underline{y}|\underline{L})$; in [5,7], $\underline{y}$ was heuristically assumed to be an independent process given $\underline{L}$.

In this paper we find $\hat{x}_i$ in a different way, separating detection and estimation. First we find the label field $\underline{L}$ by maximizing *a posteriori* probability (MAP)

$$\max_{\underline{L}} P(\underline{L}|\underline{m}) \propto \max_{\underline{L}} P(\underline{m}|\underline{L}) \cdot P(\underline{L}) \tag{7}$$

where $\underline{m}$ is the feature vector for the classification. In Sec. 4 we also empirically model $P(\underline{m}|\underline{L})$ as an independent process. However, it should be noted this model is just used for *detecting* labels, not directly for *estimating* x_i. This means we only require the model of $P(\underline{m}|\underline{L})$ be good enough to classify labels correctly. With $\underline{L}$ known we can then estimate x_i by

$$\begin{aligned}\hat{x}_i &= E(x_i|y_i, \underline{L}_{(-i)}) \\ &\approx P(L_i = H_1|y_i, \underline{L}_{(-i)}) \cdot y_i \\ &= \frac{P(y_i|L_i = H_1, \underline{L}_{(-i)}) \cdot P(L_i = H_1|\underline{L}_{(-i)})}{\sum_{k=0}^{1} P(y_i|L_i = H_k, \underline{L}_{(-i)})P(L_i = H_k|\underline{L}_{(-i)})} \cdot y_i\end{aligned} \tag{8}$$

where $\underline{L}_{(-i)}$ is the whole label field $\underline{L}$ excluding L_i.

The required probability models ($P(\underline{m}|\underline{L})$, $P(y_i|L_i = H_1, \underline{L}_{(-i)})$ and $P(L_i = H_1|\underline{L}_{(-i)})$) are discussed in Sec. 4.

3 The Complex Wavelet Transform

The present standard wavelet transforms are almost all *real-value* transforms, such as Daubechies and biorthogonal wavelets [6]. They have some interesting properties and are successfully used in many image processing applications (e.g., compression, denoising, feature extraction). However, under the constraint of being real-valued they suffer from a few disadvantages [8]:

1. The real-valued orthogonal wavelet can not be symmetric, which is expected for some applications;
2. Lack of shift invariance, which means small shifts in the input signal can cause major variations in the distribution of energy between coefficients at different scales;
3. Poor directional selectivity for diagonal features, when the wavelet filters are real and separable.

A well-known way of providing shift-invariance is to use the undecimated form of the dyadic filter tree, which is implemented most efficiently by the algorithm *a trous*. However, this suffers from substantially increased computation requirements compared to the fully decimated DWT. In addition, separable $2D$ overcomplete wavelet transforms still have poor directional selectivity. Designing non-separable direction-selective $2D$ wavelet bases is usually a complicated task.

If the wavelet filters are allowed to be complex-valued (this results in single-tree complex wavelet (ST-CWT)) all the above three problems can be overcome [8]. However, though (ST-CWT) can solve these problems it suffers from poor frequency selectivity. Thus, Kingsbury [8] proposed the dual-tree complex wavelet transform (DT-CWT) (Fig. 1). DT-CWT uses two real DWT trees to implement its real part and imaginary part, separately. In addition to the other attractive properties of the ST-CWT, DT-CWT has good frequency selectivity and easy to achieve perfect reconstruction. Indeed, Selesnick [9] found that the real and imaginary parts of the DT-CWT can be linked by the Hilbert transform. This observation further explains why DT-CWT has those useful characteristics.

The $2D$ DT-CWT can be easily implemented by the tensor products of $1D$ DT-CWT. Because the $1D$ filters are complex the $2D$ DT-CWT consists of six wavelets (Fig. 2). Note the good directionality of the wavelet bases.

4 The a Prior Models

1. $P(\underline{m}|\underline{L})$

As mentioned in Sec.2 $P(\underline{m}|\underline{L})$ is modelled as an independent process, i.e.

$$P(\underline{m}|\underline{L}) = \prod_{i=1}^{N} P(m_i|L_i) \tag{9}$$

where $\underline{m}$ is the feature vector and $\underline{L}$ is the label field. We use the magnitude of complex wavelet coefficient as the feature:

$$m(u,v) = \sqrt{{r_e}^2(u,v) + {i_m}^2(u,v)}, \quad u,v = 1,2,...,N \tag{10}$$

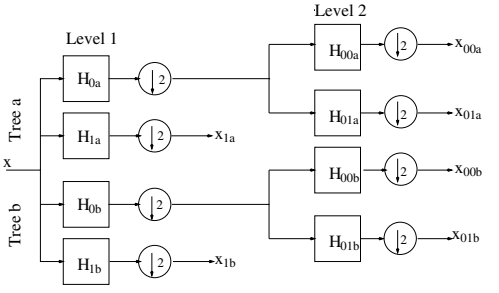

Fig. 1. *two*-level dual-tree complex wavelet transform

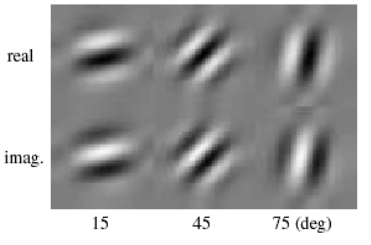

Fig. 2. $2D$ dual-tree complex wavelet bases (only three orientations are shown)

where $u, v = 1, 2, ..., N$ are coordinates of the image field. $r_e(u, v)$ and $i_m(u, v)$ are the real and imaginary parts, respectively

$$y(u, v) = r_e(u, v) + \sqrt{-1} \cdot i_m(u, v) \tag{11}$$

In this paper we assume $r_e(u, v)$ and $i_m(u, v)$ are Gaussian, as did in several other papers (e.g. [3]). Thus, a good approximate model for $m(u, v)$ is Rayleigh distribution. Fig. 3 shows the histogram (from a group of natural images) of $m(u, v)$ and estimated Rayleigh function.

2. $P(y_i|L_i = H_k, \underline{L}_{(-i)}), \quad k = 0, 1$
We assume a conditional independence

$$P(y_i|L_i = H_k, \underline{L}_{(-i)}) = P(y_i|L_i = H_k) \tag{12}$$

and then model $P(y_i|L_i = H_k)$ as complex Gaussian.

3. $P(L_i = H_1|\underline{L}_{(-i)})$
The label process $\underline{L}$ is modelled as a MRF. Specifically we use the auto-logistic model [5]. This kind of models have also been successfully applied for texture segmentation. It is described as

$$P(\underline{L}) = 1/Z \cdot exp(-V(\underline{L})) \tag{13}$$

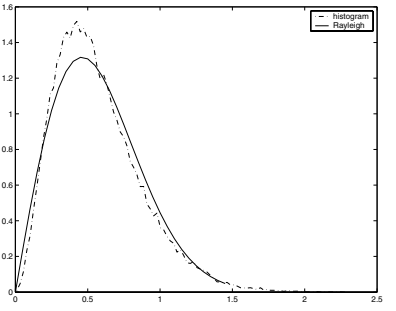

Fig. 3. Histogram of the magnitude and estimated Rayleigh function $P(m_i|L_i = H_1)$

(a) (b)

Fig. 4. Labels without (a) and with (b) *a priori* model for one wavelet orientation (15^o)

where the energy function $V(\underline{L}) = \sum_i V_{Ni}(\underline{L})$ and the clique potentials are defined as

$$V_{Ni}(\underline{L}) = \sum_{j \in Ni} V_{i,j}(L_i, L_j) \quad \text{with}$$
$$V_{i,j}(L_i, L_j) = \begin{cases} -\gamma \text{ if } L_j = L_i \\ \gamma \text{ if } L_j \neq L_i \end{cases} \tag{14}$$

where γ is a positive scalar. N_i is the first-order neighborhood system.

This *a priori* model for the label field tries to exploit the clustering property of the wavelet coefficients. It has been shown to be useful for suppressing separate noise artifact [5]. In Fig. 4 the influence of the *a priori* model is illustrated. We used iterated conditional mode (ICM) in the maximization process.

5 Experimental Results and Discussions

We applied the proposed technique to several natural images with artificial additive Gaussian noise. One result is shown in Fig. 5. For comparison we also show the denoising result from [3]. This method is widely used in references for comparison. It did not employ any *a priori* label model. Visually, Fig. 5(d) looks

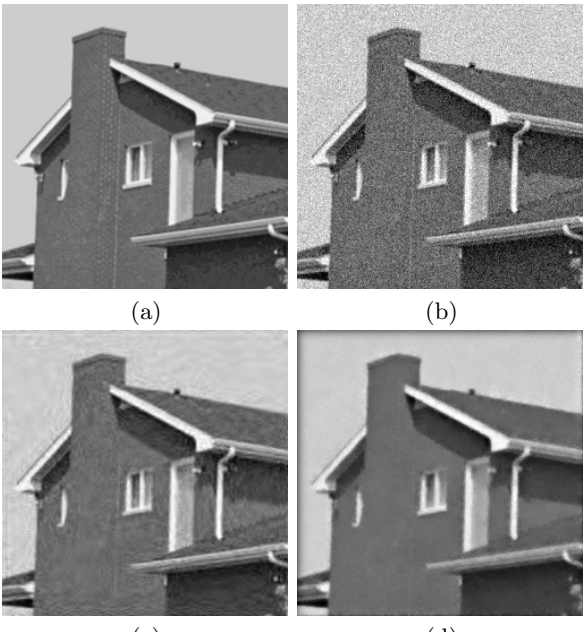

Fig. 5. (a) original (b) noisy (c) denoised image by adaptive thresholding [3] (d) proposed method

much cleaner and thus more pleasing than Fig. 5(c), though the latter looks a little bit sharper. Objectively the proposed approach is also about $1dB$ better in the sense of SNR.

We also compared with the methods in [5,7] because they also used MRF *a priori* model. We found the results looks similar and the differences in SNR are less than $0.5dB$. However, in [5,7] *undecimated* overcomplete wavelets were used. Thus their complexity is much higher than the *decimated* complex wavelet (For example for $2D$ decomposition the complex wavelet transform has a redundancy of 4, independent of number of levels. But for the overcomplete wavelet transform the redundancy is $4+3(NL-1)$, where NL denotes the number of levels). This is especially true in the MRF iteration process. Furthermore, $2D$ complex wavelets have better direction-selectivity which means potential for better denoising performance. It should also be noted our problem formulation is different. We first find edge masks according to *MAP* and then combine this knowledge with the measurement to evaluate the conditional probability $P(L_i = H_1|y_i, \underline{L}_{(-i)})$.

References

1. Kuan D., Sawchuk A., Strand T., Chavel P.: Adaptive noise smoothing filter for images with signal-dependent noise. IEEE Trans. PAMI **2**(7) (1985) 165–177
2. Lee J.: Digital image enhancement and noise filtering by use of local statistics. IEEE Trans. PAMI. **2**(2) (1980) 165–168

3. Chang S. and Bin Y. and Vetterli M.: Spatially adaptive wavelet thresholding with context modeling for image denoising. IEEE Trans. on Image Processing., **9**(9) (2000) 1522 - 1531
4. Mihcak M. and Kozintsev I. and Ramchandran K. and Moulin P.: Low-complexity image denoising based on statistical modeling of wavelet coefficients. IEEE Sig Proc.Lett., **12**(6) (1999) 300 - 303
5. Malfait M. and Roose D.: Wavelet-based image denoising using a Markov random field a prior model. IEEE Trans. Imag Proc.,**4**(6) (1997) 545 - 557
6. Daubechies I.: Orthonormal Bases of Compactly Supported Wavelets. Comm. Pure Appl. Math. **5** 41 (1988) 909-996
7. Pizurica A., Philips W., Lemahieu I. and Acheroy M.: A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans. Imag Proc.,**5**(11) (2002) 545 - 557
8. kingsbury N.: Image processing with complex wavelets. Phil. Trans. Royal Society London A, **9**(29), (1999) 2543 – 2560
9. Selesnick I.: The design of approximate Hilbert transform pairs of wavelet bases. IEEE Trans. Signal Processing, **5**(50), (2002) 1144 - 1152.