COMPRESSIVE IMAGING WITH COMPLEX WAVELET TRANSFORM AND TURBO AMP RECONSTRUCTION

Chunli Guo, James D. B. Nelson

Department of Statistical Science, University College London

ABSTRACT

We extend the "turbo" belief propagation framework for compressive imaging to the dual-tree complex wavelet transform (DT-CWT) to exploit both sparsity and dependency across scales. Due to the near shift-invariance property and the improved angular resolution of DT-CWT, better reconstruction can be expected when incorporating with the compressed sensing (CS) algorithms. Two types priors to form the hidden Markov tree structure for the DT-CWT coefficients are considered. One models the real and imaginary components of DT-CWT separately while the other assumes the shared hidden states between the two. Simulations with natural images confirm an improved performance when iterating between the CS reconstruction and the DT-CWT HMT.

Index Terms— compressed sensing, dual-tree complex wavelet transform, HMT, turbo decoding, approximate message passing

1. INTRODUCTION

The typical compressive imaging problem is to estimate an image or its representation in the transformed domain (i.e. Fourier, wavelet) $\mathbf{x} \in \mathbb{R}^n$ by solving an under-determined system

$$\mathbf{y} = \mathbf{\Phi x} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the compressed sensing (CS) observation, and $\mathbf{\Phi} \in \mathbb{R}^{m \times n}$, $m < n$ is known as the measurement matrix. The ill-posed problem can be solved by exploiting the sparse property of $\mathbf{x}$ or $\boldsymbol{\theta}$ via ℓ_1-minimization [1], greedy methods [2] or approximate message passing (AMP) [3]. Nature images, when expressed in the wavelet domain, have an additional quad-tree structure and some statistical dependency along the branches of the wavelet trees [4]. To better aid the CS reconstruction, several authors have looked beyond the signal sparsity and incorporated wavelet dependencies [5–9], among which [9] has demonstrated the state-of-the-art performance.

The authors in [9] deployed the hidden Markov tree (HMT) [4] to model the interscale dependencies of the discrete wavelet transform (DWT) coefficients. The CS observation system in (1) and the HMT structure can be graphically represented by the corresponding factor graph [10]. The reconstruction is then based on loopy belief propagation [11] to propagate beliefs for the DWT coefficients on the factor graph. The turbo message passing schedule which alternates between the exploitation of CS observation and the HMT structure is proposed. The introduction of wavelet interscale dependencies provides a richer local belief model which significantly benefits the reconstruction.

The novelty of [9] is the introduction of the turbo scheme to incorporate additional signal properties into the standard CS reconstruction. Inspired by its success it is natural to consider the utility of richer, overcomplete wavelet representations in this framework. In this paper, we adopt and extend the turbo reconstruction approach to the dual-tree complex wavelet transform (DT-CWT) [12, 13]. The advantages of DT-CWT over the standard DWT have been shown in many image processing applications, such as denoising, detection, segmentation and classification [14, 15]. Compared to the DWT, the DT-CWT is near shift-invariant and more directional sensitive, making it better at characterizing geometric images features like edges. When imposed with the HMT structure, large and small DT-CWT coefficients cascade more consistently along branches of the wavelet trees, which leads to less edge blurring and artifacts.

Two statistical priors to form the HMT have been proposed for the DT-CWT coefficients [14]. The first one simply models the real and imaginary components of the complex coefficients as independent and separate two-state 1-D HMTs. The other is more realistic, which assigns a shared hidden state to the real and imaginary part to form the tree. In this paper, both HMT structures are exploited for the CS imaging. For the former, we directly deploy the factor graph and turbo approach in [9] for reconstruction. For the latter, a new factor graph and turbo scheme are proposed to incorporate the shared states. Numerical simulations demonstrate a significant improvement for the use of DT-CWT, both visually and quantitatively.

We begin with the introduction of the DT-CWT and two HMT models in Section 2. The corresponding factor graphs and turbo reconstruction are explained in Section 3. Image reconstruction examples are shown in Section 4. We finish with conclusion in Section 5.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

2. COMPLEX WAVELET TRANSFORM AND STATISTICAL HIDDEN MARKOV TREE MODEL

2.1. Dual-Tree Complex Wavelet Transform

The DT-CWT is inspired by the shift invariant property of the Fourier transform, namely that the energy in each frequency bin is invariant to any shifts in time or space. As with the Fourier transform, the DT-CWT encodes the signal information in the magnitude and phase of the complex coefficients. For example, a large magnitude indicates the presence of singularity while the phase encodes its position. The DT-CWT decomposition of a 1-D signal is implemented via two parallel DWTs. Each of the DWTs produces a set of real value coefficients $\{\mu_i\}$ and $\{\upsilon_i\}$. Together they form the complex wavelet coefficients $c_i = \mu_i + j\upsilon_i, j = \sqrt{-1}$. With the two DWTs jointly designed, the real and imaginary components are statistically uncorrelated and the DT-CWT is insensitive to small shifts [16].

The extension of DT-CWT to 2D signals is achieved by separately filtering along rows then columns by two parallel DWTs. The resulting coefficients are thus four times redundant. As a bonus, the additional information provides us good directional selectivity with six subbands at each scale, which capture image features strongly oriented at angles of $\{\pm 15°, \pm 30°, \pm 45°, \pm 75°\}$ [15]. With the more explicit orientation information of singularities, we are able to distinguish signals in a more subtle way.

Like the DWT for 2D signals, DT-CWT coefficients for images across scales form a quad-tree structure and exhibit the persistence across scale property [14]. To be specific, if the parent wavelet coefficient is large then the children wavelet coefficients are very likely to be large. Similarly, if the parent coefficient is small, its children are likely to be small. The HMT structure to model the persistence property has also been extended to the DT-CWT. Due to its near shift-invariant, the progressions of the magnitude are better preserved in DT-CWT coefficients.

2.2. Prior for independent HMT

The simplest prior is to model the real and imaginary part at each wavelet scale separately with the independent and identically distributed (i.i.d) two-state Gaussian mixture distribution.

$$\begin{aligned} p(\mu_{k,i}) &= p(\mu_{k,i}|s_{\mu,k,i}=1)p(s_{\mu,k,i}=1) \\ &\quad + p(\mu_{k,i}|s_{\mu,k,i}=0)p(s_{\mu,k,i}=0) \\ &= \lambda_{\mu,k}\mathcal{N}(\mu_{k,i};0,\sigma^2_{k,L}) + (1-\lambda_{\mu,k})\mathcal{N}(\mu_{k,i};0,\sigma^2_{k,S}) \end{aligned} \quad (2)$$

where k is the wavelet scale index and i is the coefficient index. In the sequel, we sometimes drop the scale index to imply the generality for all wavelet scales. Each $\mu_{k,i}$ can be seen as generated from either the large variance Gaussian distribution with variance $\sigma^2_{k,L}$ or from small variance Gaussian distribution with variance $\sigma^2_{k,S}$, depending on the associated hidden state $s_{\mu,k,i}$ being 1 or 0.

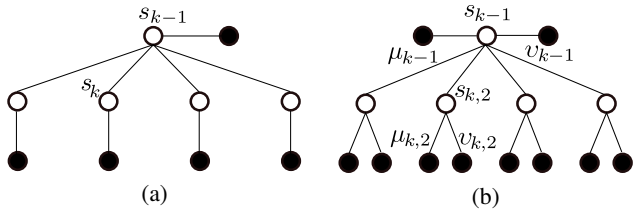

Fig. 1. HMT models for dual-tree complex wavelet coefficients: (a) standard HMT structure for real or imaginary part; (b) assign a shared hidden state for the associated real and imaginary components.

Similarly, the pdf of the imaginary part at scale k is

$$\begin{aligned} p(\upsilon_{k,i}) &= p(\upsilon_{k,i}|s_{\upsilon,k,i}=1)p(s_{\upsilon,k,i}=1) \\ &\quad + p(\upsilon_{k,i}|s_{\upsilon,k,i}=0)p(s_{\upsilon,k,i}=0) \\ &= \lambda_{\upsilon,k}\mathcal{N}(\upsilon_{k,i};0,\sigma^2_{k,L}) + (1-\lambda_{\upsilon,k})\mathcal{N}(\upsilon_{k,i};0,\sigma^2_{k,S}) \end{aligned} \quad (3)$$

Connecting the hidden states across scales for $\boldsymbol{\mu}$ and $\boldsymbol{\upsilon}$ results in a HMT structure that is the same as the one for the standard DWT, which is illustrated in Fig. 1(a).

2.3. Prior for HMT with shared hidden states

A more reasonable model to capture the magnitude persistence would be assigning a two-state hidden variable $s_{c,i}$ for the complex wavelet coefficient c_i, taking value 1 when $|c_i|$ being large and 0 when $|c_i|$ being small. One statistical model is to consider c_i as a 2-d random vector (μ_i, υ_i) and approximate $p(c_i)$ as a two-state 2-d Gaussian mixture, as proposed in [14].

$$\begin{aligned} p(c_{k,i}) &= p(c_{k,i}|s_{c,k,i}=1)p(s_{c,k,i}=1) \\ &\quad + p(c_{k,i}|s_{c,k,i}=0)p(s_{c,k,i}=0) \\ &= \frac{\lambda_{c,k}}{\sqrt{2\pi}\varsigma_{k,L}} e^{-\frac{\mu^2_{k,i}+\upsilon^2_{k,i}}{2\varsigma^2_{k,L}}} + \frac{1-\lambda_{c,k}}{\sqrt{2\pi}\varsigma_{k,S}} e^{-\frac{\mu^2_{k,i}+\upsilon^2_{k,i}}{2\varsigma^2_{k,S}}} \end{aligned} \quad (4)$$

Since μ_i and υ_i are uncorrelated, this is equivalent to model the real and imaginary parts as the two-state Gaussian mixture with a shared hidden state variable $s_{c,i}$ [14]. The quad-tree structure with shared hidden states is depicted in Fig. 1(b).

Compared to the independent HMT structure, this model is more realistic: small $|c_i|$ can only appears when $|\mu_i|$ and $|\upsilon_i|$ are simultaneously small, which is characterized by $s_{c,i} = 0$. While for the independent HMT model, small coefficients generated from a Gaussian mixture with the hidden variable being 1 is also possible.

2.4. Complete statistical model for HMT

So far we have presented two priors for DT-CWT coefficients. To complete the statistical model for the HMT structure, we

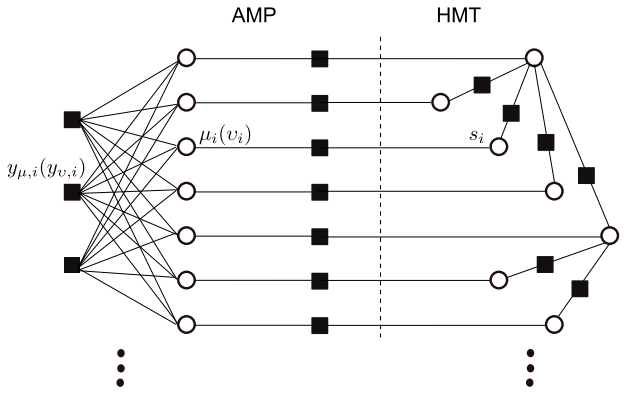

Fig. 2. Factor graph for reconstructing real $\boldsymbol{\mu}$ and imaginary $\boldsymbol{\upsilon}$ separately.

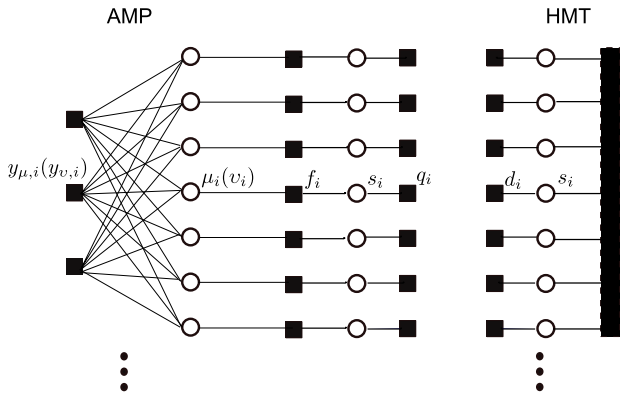

Fig. 3. The decoupled two subgraphs for the original factor graph in Fig. 2.

also need the prior for the hidden states and the statistical model for the transition probabilities ξ_k^{00}, ξ_k^{11} across scales, where $\xi_k^{00} = p(s_{k+1} = 0|s_k = 0)$ and $\xi_k^{11} = p(s_{k+1} = 1|s_k = 1)$. We follow the hierarchical Bayesian approach in [9] and model the prior for the hidden states and the transition probabilities as Beta distributed. The hyper-parameters of the Beta distribution are updated at each turbo iteration as described in [9].

3. TURBO RECONSTRUCTION

With the statistical model chosen, we now discuss the implementation of the turbo reconstruction method. Throughout, we assume the CS measurements are taken in the DT-CWT domain, which is

$$\mathbf{y} = \mathbf{\Phi c} = \mathbf{\Phi}\boldsymbol{\mu} + j\mathbf{\Phi}\boldsymbol{\upsilon} \tag{5}$$

where $\mathbf{c} = \{c_{k,i}\}$, $\boldsymbol{\mu} = \{\mu_{k,i}\}$, and $\boldsymbol{\upsilon} = \{\upsilon_{k,i}\}$ for all k, i.

To utilize the AMP algorithm for real quantities, we can separate the real and imaginary parts of the DT-CWT coefficients as

$$\mathbf{y}_\mu = \mathbf{\Phi}\boldsymbol{\mu}, \qquad \mathbf{y}_\upsilon = \mathbf{\Phi}\boldsymbol{\upsilon} \tag{6}$$

Ideally, we would like to calculate the posterior pdf $p(\mathbf{c}|\mathbf{y})$ to infer $\mathbf{c}$. However, exact computation of $p(\mathbf{c}|\mathbf{y})$ is intractable due to the dense measurement matrix. Alternatively, we can resort to the marginal posterior $p(c_i|\mathbf{y})$ which can be effectively approximated using loopy belief propagation on the factor graph.

3.1. Reconstruction for Independent HMT

When $\boldsymbol{\mu}$ and $\boldsymbol{\upsilon}$ are modeled separately with two-state 1-D HMT structure, the reconstruction can be performed independently in parallel by calculating $p(\mu_i|\boldsymbol{y}_\mu)$ and $p(\upsilon_i|\boldsymbol{y}_\upsilon)$. When dealing with complicated functions involving many variables, efficient algorithms may be derived by exploiting the factorization of the global function: expressing it as a product of several simpler "local" functions, each of which depends on only a subset of the variables. A factor graph is graphical representation of such factorization, which connects local functions with their related argument variables [10]. The factor graph for $p(\boldsymbol{\mu}|\boldsymbol{y}_\mu)$ or $p(\boldsymbol{\upsilon}|\boldsymbol{y}_\upsilon)$ is illustrated in Fig. 2.

We use round nodes to denote variables (i.e. μ_i, υ_i, s_i), square nodes for operations (i.e. transition function among hidden states, assigning priors) and lines to connect operations with the associated augments in the factor graph. "Messages" flowing to/from a variable node along edges (as illustrated by arrows in Fig. 5) can be calculated according to the sum-product algorithm [10]. The loopy belief propagation is conducted by exchanging messages passing along nodes and edges until they converge. In our context, messages take the form of pdfs and represent local beliefs for variables.

Unsurprisingly, the factor graph in Fig. 2 has the same structure as the one in [9] since we essentially treat the reconstruction of $\mathbf{c}$ as two DWT CS problems. Thus we can directly deploy the same turbo approach in [9] to obtain $p(\mu_i|\boldsymbol{y}_\mu)$ and $p(\upsilon_i|\boldsymbol{y}_\upsilon)$. The essence is to decouple the factor graph into two subgraphs (reconstruction parts) along the dash line in Fig. 2 and exchange the local belief on the hidden states $\{s_i\}$ between two subgraphs alternately, by treating the likelihood from one part as the prior for the other. More specifically, at the first turbo iteration, by assuming some initial prior $p(s_i = 1)$ for the hidden states, we start with the reconstruction on the left subgraph. Performing the loopy belief propagation on the left subgraph of Fig. 2 is equivalent to the AMP algorithm [3] and produces the "local likelihood" on hidden states. These local beliefs are then treated as priors for $\{s_i\}$ in the right subgraph for the HMT reconstruction. The belief propagation for factor graphs without loops has exact solution and yields the commonly known forward-backward algorithm [10]. The local likelihoods for $\{s_i\}$ from the right subgraph are fed back to the left subgraph as an updated prior for the AMP algorithm before invoking the new turbo iteration.

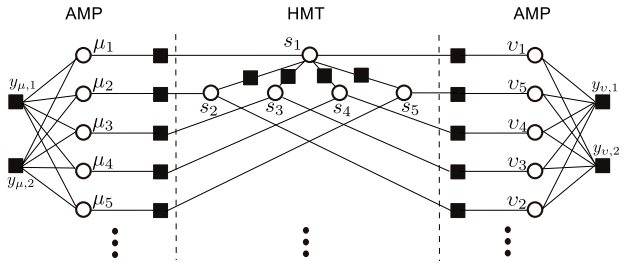

Fig. 4. Factor graph for modeling $\boldsymbol{\mu}$ and $\boldsymbol{\upsilon}$ with shared hidden states.

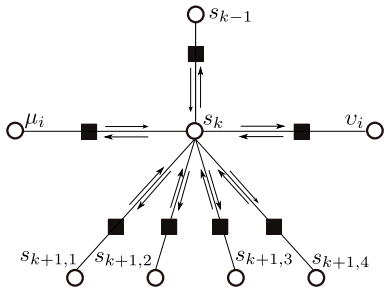

Fig. 5. One typical hidden variable node for the factor graph in Fig. 4.

The decoupled factor graph is shown in Fig. 3, where the quad-tree structure are abstracted as the supernode H in the right subgraph. Let t denote the turbo iteration index, $\{q_i(\cdot)\}$ are hidden states priors acting on the left subgraph, $\{d_i(\cdot)\}$ are hidden states priors acting on the right subgraph. Let $m^{(t)}_{A\to B}$ represents the message sent from node A to node B in the tth turbo iteration. At $t=1$, initialize $q_i(s_i)=p(s_i=1)$. Then the turbo message passing can be summarized as

$$d_i^{(t)}(s_i) = m^{(t)}_{f_i\to s_i}(s_i) \tag{7}$$

$$q_i^{(t+1)}(s_i) = m^{(t)}_{H\to s_i}(s_i) \tag{8}$$

for $t=2,3,\cdots$. We refer to this reconstruction as the Turbo-CHMT1 algorithm.

3.2. Reconstruction for HMT with shared hidden states

When connecting $\boldsymbol{\mu}$ and $\boldsymbol{\upsilon}$ with the shared hidden states, the reconstruction is no longer separable. The new factor graph for the system is illustrated in Fig. 4.

To perform the belief propagation, we also deploy the turbo method and split the loopy factor graph along the two dash lines into two AMP reconstruction parts and one HMT reconstruction part. The turbo iteration starts with the AMP reconstruction on the left and right subgraph simultaneously and produces two partial information for the hidden state $m_{\mu_i\to s_i}(s_i)$ and $m_{\upsilon_i\to s_i}(s_i)$, respectively. The forward-backward algorithm is then performed for the middle subgraph on the quad-tree structure, which yields two output messages $m_{s_i\to\mu_i}(s_i)$ and $m_{s_i\to\upsilon_i}(s_i)$. In the next turbo iteration these are used as the updated prior for s_i for the separate AMP reconstructions.

Fig. 5 features a hidden state node s_k with one parent node s_{k-1}, four children nodes $\{s_{k+1,i}\}_{i=1}^4$ and the associated real and imaginary coefficients μ_i, υ_i. Compared to the HMT structure in Fig. 3, the key modification is the extra input message for the hidden states, which needs to be taken care of when performing the forward-backward to calculate the messages exchanging between s_k and its parent and children. According to the sum-product algorithm, the output message from s_k to μ_i and υ_i are

$$\begin{aligned} m_{s_k\to\mu_i}(s_k) =& m_{\upsilon_i\to s_k}(s_k) m_{s_{k-1}\to s_k}(s_k) \\ &\times \prod_{i=1}^{4} m_{s_{k+1,i}\to s_k}(s_k) \end{aligned} \tag{9}$$

$$\begin{aligned} m_{s_k\to\upsilon_i}(s_k) =& m_{s_k\to\mu_i}(s_k) m_{s_{k-1}\to s_k}(s_k) \\ &\times \prod_{i=1}^{4} m_{s_{k+1,i}\to s_k}(s_k) \end{aligned} \tag{10}$$

The benefit of sending messages from both real and imaginary parts to the hidden states is twofold. For one thing, it reinforces the confidence of estimation for s_i when the messages agree with each other (both imply s_i is more likely to be 1 or 0). For another, it softens the decision for s_i when they indicate different state. We denote this turbo approach as the Turbo-CHMT2 algorithm. Numerical simulation with natural images indicates an accelerated convergence for the reconstruction with Turbo-CHMT2.

4. NATURAL IMAGE EXAMPLE

In this section, the proposed Turbo-CHMT1 and Turbo-CHMT2 algorithm are tested on three 128×128 grayscale images using 4 level Q-shift DT-CWT [1], and compared with the original Turbo algorithm with db2 wavelet decomposition [2] and the EM-GM-GAMP algorithm [17], which does not exploit the wavelet dependency. In all cases, the entries of the measurement matrix are i.i.d Gaussian elements. The maximum AMP iteration is set to be 200. The CS measurements are taken in the wavelet domain. We used the normalized mean squared error (NMSE) $\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2/\|\mathbf{x}\|_2^2$ for performance comparison.

Table 1 reports the quantitative comparison of the reconstruction algorithms under different sampling ratios for three standard images. The reconstruction of the lena images in Fig. 6 gives us a visual comparison. We can see that both Turbo-CHMT1 and Turbo-CHMT2 provide us a significant

[1] http://www-sigproc.eng.cam.ac.uk/Main/NGK

[2] http://www2.ece.ohio-state.edu/ schniter/turboAMPimaging/

(a) Original (b) EM-GAMP (c) Turbo (c) Turbo-CHMT1 (c) Turbo-GHMT2

Fig. 6. Reconstruction with 30% CS observations of the 128×128 lena image using various algorithms.

reconstruction improvement (about 1.5 dB) over the original Turbo algorithm. We attribute the improvement to the near shift-invariant property and good directional resolution of the DT-CWT. The best performance belongs to the Turbo-CHMT2, which validates the HMT model with shared hidden states and the proposed turbo reconstruction approach.

Table 1. NMSE (dB) for image reconstruction

Image	*Cameraman*		*Lena*		*Boat*	
m/n	0.2	0.3	0.2	0.3	0.2	0.3
EM-GAMP	-15.10	-19.03	-14.11	-17.43	-16.74	-19.29
Turbo	-17.96	-20.85	-15.60	-18.47	-18.36	-21.08
Turbo-CHMT1	-18.63	-23.32	-17.64	-20.36	-19.38	-21.61
Turbo-CHMT2	-19.12	-23.98	-18.06	-20.88	-19.76	-22.09

5. CONCLUSION

Given the advantageous near shift-invariant property and directional selectivity of the DT-CWT, we are motivated to exploit wavelet dependencies for the DT-CWT coefficients to aid the CS imaging. Two types of Gaussian mixture priors are considered for the DT-CWT coefficients with two different graphical representations for reconstruction. Inspired by the turbo decoding mechanism for the DWT coefficients, we propose two turbo algorithms for the loopy factor graphs. As the reconstruction examples have demonstrated, our algorithms have successfully leveraged the interscale dependencies among DT-CWT coefficients for CS reconstruction. Further work might involve the exploitation of the phase information encoded in the DT-CWT coefficients.

REFERENCES

[1] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci Comp.*, vol. 20, pp. 33–61, 1999.

[2] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[3] D. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. of the Nat. Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[4] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans on Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[6] L. He and L. Carin, "Exploiting structure in wavelet-based bayesian compressive sensing," *IEEE Trans. on Signal Process.*, vol. 57, no. 9, pp. 3488–3497, Sept. 2009.

[7] M. Duarte, M. Wakin, and G. Baraniuk, "Wavelet-domain compressive signal reconstruction using a hidden markov tree model," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Las Vegas, NV, 2008, pp. 5137–5140.

[8] P. Schniter, "Turbo reconstruction of structured sparse signals," in *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, March 2010, pp. 1–6.

[9] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," *IEEE Trans. on Signal Process.*, vol. 60, no. 7, pp. 3439–3448, July 2012.

[10] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. on Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[11] B. J. Frey and D. J. MacKay, "A revolution: Belief propagation in graphs with cycles," *Advances in neural information processing systems*, pp. 479–485, 1998.

[12] N. Kingsbury, "The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement," in *Proc. EUSIPCO*, vol. 98, 1998, pp. 319–322.

[13] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.

[14] H. Choi, J. Romberg, R. Baraniuk, and N. Kingsbury, "Hidden markov tree modeling of complex wavelet transforms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 2000, pp. 133–136.

[15] N. Kingsbury, "Image processing with complex wavelets," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 357, no. 1760, pp. 2543–2560, 1999.

[16] ——, "Shift invariant properties of the dual-tree complex wavelet transform," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3. IEEE, 1999, pp. 1221–1224.

[17] J. Vila and P. Schniter, "Expectation-maximization gaussian-mixture approximate message passing," *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4658–4672, Oct 2013.