

STRUCTURED BAYESIAN COMPRESSIVE SENSING EXPLOITING SPATIAL LOCATION DEPENDENCE

Qisong Wu[†], Yimin D. Zhang[†], Moeness G. Amin[†], and Braham Himed[‡]

[†] Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA

[‡] RF Technology Branch, Air Force Research Lab (AFRL/RVMD), WPAFB, OH 45433, USA

ABSTRACT

In this paper, we propose a novel structured compressive sensing algorithm based on non-parametric Bayesian framework for the reconstruction of sparse entries with a continuous structure. A paired spike-and-slab prior is first employed to impose signal sparsity. A logistic Gaussian kernel model, which involves the logistic model and location-dependent Gaussian kernel, is then proposed to encourage the underlying structure of a sparse signal. A closed-form and analytical posterior inference is carried out in a Gibbs sampling scheme. Simulation results demonstrate that the proposed algorithm outperforms existing state-of-the-art sparse Bayesian learning algorithms.

Index Terms— Compressive sensing, sparse Bayesian learning, Gaussian kernel, logistic model

1. INTRODUCTION

Sparse signal recovery and the associated compressive sensing (CS) approaches have attracted significant attention in recent years [1, 2]. CS techniques have the capability to recover signals from a small number of measurement samples with a high probability, provided that the signals are sparse or can be sparsely represented in some known domains.

A typical sparse reconstruction problem in the CS model is given by,

$$\mathbf{y} = \Phi \mathbf{w} + \varepsilon, \quad (1)$$

where $\mathbf{y} \in \mathcal{R}^N$ and $\varepsilon \in \mathcal{R}^N$ is an unknown zero-mean Gaussian noise vector. $\Phi \in \mathcal{R}^{N \times M}$ is a known and wide dictionary matrix due to $N \ll M$, and hence loses information in general. However, it can be shown to preserve the information in a sparse or compressible signal if the so-called restricted isometry property (RIP) is satisfied [2]. Define δ_A as the constant of RIP for a sensing matrix $\Phi \in \mathcal{R}^{N \times M}$. If

$$N > \frac{2}{c\delta_A} \left(\ln(2m_K) + K \ln \frac{12}{\delta_A} + t \right), \quad (2)$$

The work of Q. Wu, Y. D. Zhang, and M. G. Amin was supported in part by a subcontract with Defense Engineering Corporation for research sponsored by the Air Force Research Laboratory under Contract FA8650-12-D-1376.

where $m_K = \binom{M}{K}$ for a K -sparse signal, and $c > 0$ is a constant, then its RIP is held for all elements in the subspace with probability $1 - e^{-t}$ [3].

A number of algorithms have been proposed to recover sparse signals. Commonly used algorithms include greedy algorithms, such as block orthogonal matching pursuit (BOMP) [4], and dynamic programming algorithms, such as basis pursuit (BP) [5], its extended version for denoising [6], and Lasso algorithm [7]. While all these algorithms are able to reconstruct sparse signals, they require information about the number of non-zero element block either explicitly or through setting of the regularization parameter, which, in practice, may not be easily obtained. Bayesian approaches form a different class of sparse signal reconstruction algorithms, which generally yield improved performance [8–10]. Sparse Bayesian learning algorithms [8–15] provide effective solutions to a large class of problems based on a non-parametric Bayesian framework, and thus have the capability of inferring the sparsity parameter and avoiding the nuisance parameters. However, these algorithms do not exploit the underlying features described as the spatial structure of the pixel locations. In fact, various spatial characteristics can be exploited in practice. For example, in through-the-wall imaging and structural health monitoring, targets and flaws of interest often have extended occupancies that are clustered in the image domain [16–19]. In the time-frequency analysis, frequency modulated (FM) signals have a sparse and continuous signatures in the time-frequency domain [20].

In this paper, inspired by the algorithms in [21–24] with the exploitation of sparse signal structures, we propose a novel logistic Gaussian kernel based algorithm to improve sparse signal reconstruction by exploiting the locations features. A spike-and-slab prior is first employed to impose the signal sparsity. Unlike other approaches, such as [22–24], which assign specific pattern priors for the exploitation of clustered structures, a novel logistic Gaussian kernel model, which involves the logistic model and location-dependent Gaussian kernel, is proposed to automatically utilize the underlying contiguous structure. Finally, the posterior distribution of the model is analytically derived in the Gibbs sampling scheme, although it is difficult to acquire a closed-form posterior inference from either the logistic model or Gaussian kernel.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). $(\cdot)^T$ denotes the transpose

a matrix or vector. $\text{diag}(\mathbf{x})$ represents a diagonal matrix that uses the elements of \mathbf{x} as its diagonal elements. “ \circ ” denotes an element-wise multiplication. $p(\cdot)$ denotes the probability density function (pdf), and $\mathcal{N}(x|a, b)$ denotes that random variable x follows a Gaussian distribution with mean a and variance b . $\text{Gamma}(x|a, b)$ denotes that random variable x follows a Gamma distribution with parameters a and b , and $\text{Bern}(x|\pi)$ implies that random variable x follows a Bernoulli distribution with weight π .

2. GENERATIVE MODEL

2.1. Spike-and-slab prior for signal sparsity

To encourage signal sparsity, we place a spike-and-slab prior on \mathbf{w} , similar to [22, 25, 26], i.e.,

$$p(\mathbf{w}|\boldsymbol{\pi}, \boldsymbol{\beta}) = \prod_{i=1}^M [(1 - \pi_i)\delta(w_i) + \pi_i\mathcal{N}(w_i|0, \beta_i^{-1})], \quad (3)$$

where π_i is the prior probability of a non-zero element, i.e., a large weight π_i corresponds to a high probability that the entry takes a non-zero value, whereas a small π_i tends to generate a zero entry. In addition, β_i is the precision (reciprocal of the variance) of Gaussian distribution and $\delta(x)$ is the Dirac delta function.

The presence of the delta function in (3) makes the inference troublesome. However, there exists a simple reparameterization of the spike-and-slab prior [23, 25, 26]. Specially, assume a Gaussian random vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]^T$ with $p(\boldsymbol{\theta}) = \prod_{i=1}^M \mathcal{N}(\theta_i|0, \beta_i^{-1})$, and a Bernoulli random vector $\mathbf{z} = [z_1, \dots, z_M]^T$ with $p(\mathbf{z}) = \prod_{i=1}^M \text{Bern}(z_i|\pi_i)$, where $z_i = 1$ corresponds to a non-zero entry in the i th position. The product of these latent vectors $\mathbf{w} = \boldsymbol{\theta} \circ \mathbf{z}$ follows the pdf in (3), i.e.,

$$p(\boldsymbol{\theta}, \mathbf{z}) = \prod_{i=1}^M [\mathcal{N}(\theta_i|0, \beta_i^{-1})]^{z_i} \pi_i^{z_i} (1 - \pi_i)^{1-z_i}. \quad (4)$$

To acquire the trackable posterior of β_i , we place a Gamma prior, which is the conjugate to the Gaussian distribution, on β_i , i.e., $\text{Gamma}(\beta_i|a, b)$, $i \in [1, \dots, M]$, where a and b are hyper-parameters. A Gaussian prior is placed on the additive noise as $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \alpha_0^{-1}\mathbf{I}_K)$. In a similar way, we place a Gamma prior on α_0 to acquire an analytical posterior distribution, i.e., $\text{Gamma}(\alpha_0|c, d)$, where c and d are hyper-parameters.

2.2. Logistic Gaussian kernel model for signal structure

Inspired by the CluSS-MCMC algorithm [22], which imposes local patterned priors for the clustered structure, a logistic Gaussian kernel model is proposed to encourage the underlying continuous structure of pixels within the image. Unlike CluSS-MCMC BCS, which manually categorizes a few specific patterns and assigns corresponding hyper-parameters for each pattern to encourage the clustered structure, the proposed algorithm employs a kernel method to achieve this function by exploiting the physical location of pixels.

Therefore, the proposed algorithm has the capability to model the relationship of the entire pixels within the image.

In Section 2.1, the variable z_i follows a Bernoulli distribution with weight π_i . We employ a logistic function to express π_i in term of x_i as

$$\pi_i = \frac{1}{1 + e^{-x_i}}. \quad (5)$$

To model the continuous structure of the pixels’ support in the image, a Gaussian kernel is introduced and expressed as,

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad (6)$$

$$\Sigma_{ij} = \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\sigma_0}\right), \quad (7)$$

for $i, j \in (1, \dots, M)$, where \mathbf{s}_i and \mathbf{s}_j are physical locations of the i th and j th pixels within the image, and $\sigma_0 > 0$ is a scale parameter.

Notice in Eq. (6) that the kernel matrix is a real, symmetric, and Toeplitz, and all entries take values within $[0, 1]$. The diagonal entries take the value of unity because $\|\mathbf{s}_i - \mathbf{s}_j\|^2 = 0$ for $i = j$, and the values decrease for off-diagonal elements, depending on the distance from the main diagonal. When all the other pixels are involved in deciding the prior probability of a pixel under consideration, those which are closer to the underlying pixel have stronger influence than others. As such, the logistic model allows a soft-thresholding weights rather than the hard-thresholding ones used in the existing methods [17, 20, 22, 23].

It should be pointed out that σ_0 is an important parameter that determines the shape of a Gaussian function and, thereby, the locality scale of the signal structure. When $\sigma_0 \rightarrow \infty$, all entries in the matrix $\boldsymbol{\Sigma}$ would approach 1. It represents that the supports of all the pixels in the image are highly correlated. When $\sigma_0 \rightarrow 0$, on the other hand, the model will reduce to a typical sparse regression without consideration of the location information. In the conventional way, the value of σ_0 is assigned to be a factor of 2.

3. POSTERIOR INFERENCE

In this section, we adopt a Gibbs sampler to perform Bayesian inference. For convenience, we first define the collection of hyper-parameters as $\boldsymbol{\Xi} \triangleq \{a, b, c, d, \sigma_0\}$ and the collection of random variables as $\boldsymbol{\Theta} \triangleq \{\boldsymbol{\theta}, \mathbf{z}, \mathbf{x}, \alpha_0, \boldsymbol{\beta}\}$.

The explicit form of the joint pdf is

$$\begin{aligned} p(\mathbf{y}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\Xi}) &= \mathcal{N}(\mathbf{y}|\boldsymbol{\Phi}(\boldsymbol{\theta} \circ \mathbf{z}), \alpha_0^{-1}\mathbf{I}_N) \text{Gamma}(\alpha_0|c, d) \\ &\times \prod_{i=1}^M \mathcal{N}(\theta_i|0, \beta_i^{-1}) \text{Gamma}(\beta_i|a, b) \\ &\times \prod_{i=1}^M \text{Bern}\left(x_i \middle| \frac{1}{1 + e^{-x_i}}\right) \mathcal{N}(\mathbf{x}|\mathbf{0}, \boldsymbol{\Sigma}). \end{aligned} \quad (8)$$

3.1. Bayesian Inference Based on Gibbs Sampler

3.1.1. Updating paired variables $\{\theta_i, z_i\}$

Similar to the inference of $\{\theta_i, z_i\}$ in [22], the paired Gibbs sampler iteratively samples from the following conditional pdf

$$p(\theta_i, z_i | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) = p(\theta_i | z_i, \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) p(z_i | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}), \quad (9)$$

where $\boldsymbol{\theta}_{\setminus i}$ denotes vector $\boldsymbol{\theta}$ with variable θ_i excluded, and $\mathbf{z}_{\setminus i}$ denotes vector \mathbf{z} except variable z_i .

By utilizing the marginal distribution principle, we have

$$p(z_i | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) = \int p(z_i, \theta_i | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) d\theta_i, \quad (10)$$

where

$$\begin{aligned} p(z_i, \theta_i | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}, \mathbf{x}) &\propto \mathcal{N}(\mathbf{y}_{\setminus i} | \boldsymbol{\phi}_i z_i \theta_i, \alpha_0^{-1} \mathbf{I}_N) \\ &\times \mathcal{N}(\theta_i | 0, \beta_i^{-1}) [1 / (1 + e^{-x_i})]^{z_i} \\ &\times [e^{-x_i} / (1 + e^{-x_i})]^{1-z_i}, \end{aligned} \quad (11)$$

$\mathbf{y}_{\setminus i} = \mathbf{y} - \sum_{k \neq i} \phi_k z_k \theta_k$, and $\boldsymbol{\phi}_i$ represents the i th column of the measurement matrix $\boldsymbol{\Phi}$. The probability of $p(z_i = 1 | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y})$ is analytically acquired to be

$$p(z_i = 1 | \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) = \frac{1}{1 + e^{-u}}, \quad (12)$$

$$u = \frac{1}{2} \left(\log \beta_i - \log \sigma_i + \sigma_i^{-1} \alpha_0^2 \mathbf{y}_{\setminus i}^H \boldsymbol{\phi}_i \boldsymbol{\phi}_i^H \mathbf{y}_{\setminus i} + 2x_i \right), \quad (13)$$

$$\sigma_i = (\alpha_0 \boldsymbol{\phi}_i^H \boldsymbol{\phi}_i + \beta_i)^{-1}. \quad (14)$$

The conditional distribution of $p(\theta_i | z_i = 1, \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y})$ can be expressed as

$$p(\theta_i | z_i = 1, \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) = \mathcal{N}(\theta_i | \sigma_i^{-1} \alpha_0 \boldsymbol{\phi}_i^H \mathbf{y}_{\setminus i}, \sigma_i). \quad (15)$$

When $z_i = 0$, the variable θ_i is drawn from the prior. As such, we obtain

$$p(\theta_i | z_i = 0, \boldsymbol{\theta}_{\setminus i}, \mathbf{z}_{\setminus i}, \mathbf{y}) = 0. \quad (16)$$

3.1.2. Updating variables \mathbf{x}

It is generally difficult to perform inference for variable \mathbf{x} , because it is not conjugate between the logistical distribution and the Gaussian distribution [27]. Some strategies, such as analytic approximations, numerical integration, and Metropolis-Hastings, can be used to perform the approximation of complex pdf [28]. In the proposed model, however, we can acquire an analytical posterior inference for \mathbf{x} by following the auto-augmentation technique for the logistic model, which introduces an Polya-Gamma random variable vector \mathbf{q} [29],

$$\begin{aligned} p(\mathbf{x} | \mathbf{q}, \boldsymbol{\kappa}, \boldsymbol{\Sigma}) &\propto \mathcal{N}(\mathbf{x} | \boldsymbol{\nu}, \boldsymbol{\Lambda}) \mathcal{N}(\mathbf{x}, \mathbf{0}, \boldsymbol{\Sigma}) \\ &= \mathcal{N}(\mathbf{x} | \boldsymbol{\lambda}, \boldsymbol{\Gamma}), \end{aligned} \quad (17)$$

$$\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}^{-1} \boldsymbol{\nu}, \quad (18)$$

$$\boldsymbol{\Gamma} = [\boldsymbol{\Lambda}^{-1} + \boldsymbol{\Sigma}^{-1}]^{-1}, \quad (19)$$

where $\boldsymbol{\Lambda} = \text{diag}(q_1^{-1}, \dots, q_M^{-1})$ and $\boldsymbol{\nu} = [\kappa_1/q_1, \dots, \kappa_M/q_M]^T$ with $\kappa_i = z_i/2$. In addition, the augmented variable q_i is updated as,

$$p(q_i | x_i) = \text{PG}(q_i | 1, x_i). \quad (20)$$

where PG is the Polya-Gamma distribution.

3.1.3. Updating signal precision β

By utilizing the conjugate property of the Gaussian and Gamma distribution, we analytically acquire the posterior distribution of the precision variable β_i as

$$p(\beta_i | a, b, \boldsymbol{\theta}) = \text{Gamma}(\beta_i | a + 1/2, b + \theta_i^2/2), \quad (21)$$

where $i \in \{1, \dots, M\}$.

3.1.4. Updating noise precision α_0

Similar to β_i , we obtain the posterior distribution of noise precision α_0

$$p(\alpha_0 | c, d, \mathbf{y}, \boldsymbol{\Phi}, \boldsymbol{\theta}, \mathbf{z}) = \text{Gamma}(\alpha_0 | c + N/2, d + \eta), \quad (22)$$

where $\eta = \|\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\theta} \circ \mathbf{z})\|^2/2$.

In general, the posterior distributions of random variables $\{\boldsymbol{\theta}, \mathbf{z}, \alpha_0, \beta\}$ will easily be obtained due to the conjugate property. For the variables \mathbf{x} in the logistic model, an effective data-augmentation strategy is employed to acquire an analytical and closed-form posterior distribution. In conclusion, the proposed model can be analytically evaluated.

The computational cost of the proposed method is mainly due to the matrix inversion in Eq. (19), which involves $M \times M$ matrix inversion and generally requires $\mathcal{O}(M^3)$ operations. However, by exploiting the property of Toeplitz matrix inversion, the complexity can be reduced to $\mathcal{O}(M^2)$ [30], which is comparable to that of the other BCS methods [9, 10, 22, 24].

4. EXPERIMENTS

The experiments use the same setting where the hyper-parameters for the prior are set as follows: $a = b = c = d = 10^{-6}$. To evaluate the reconstruction performance of different CS methods, the normalized mean square error (NMSE) $\|\hat{\mathbf{w}} - \mathbf{w}_{\text{gen}}\|_2^2 / \|\mathbf{w}_{\text{gen}}\|_2^2$ is used as the performance index, where $\hat{\mathbf{w}}$ is the estimate of the true signal \mathbf{w}_{gen} .

The following state-of-art algorithms are included for performance comparison: basis pursuit denoising (BPDN)[6], BCS algorithm [9], BCS with Laplace prior (BCS-L) [10], and clustered BCS method (CluSS-MCMC) [22].

4.1. Binary image reconstruction

In the first experiment, we consider a binary image for the letter ‘C’ with the size of 16×16 and the sparsity number $K = 32$, as shown in Fig. 1(a). The measurement matrix $\boldsymbol{\Phi} \in R^{N \times M}$ is generated as a zero mean random Gaussian matrix with columns normalized to unit l_2 norm. Gaussian noise is added to the observations so that the signal-to-noise ratio (SNR) is 20 dB. The length of the observation K is 90, and the scale parameter is 2^4 .

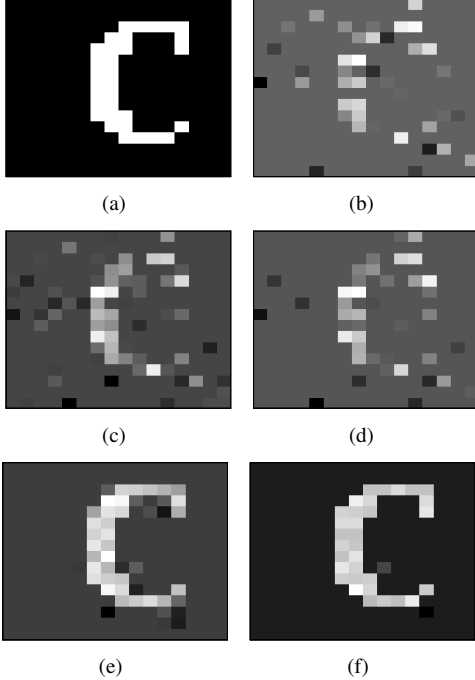


Fig. 1. Reconstructed results from different methods. (a) original binary image; (b) BPDN; (c) BCS based on RVM; (d) BCS with Laplace prior; (e) CluSS-MCMC; and (f) proposed algorithm.

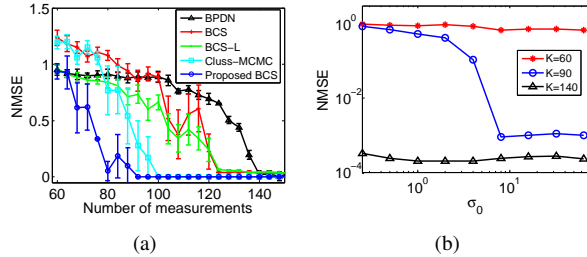


Fig. 2. NMSE versus (a) number of measurements and (b) scale parameter σ_0 .

The reconstructed images based on the aforementioned state-of-art algorithms are respectively given in Figs. 1(b)–1(f). It is evident that the proposed algorithm has the capability of recovering the binary image by exploiting the location information of pixels, as shown in Fig. 1(f). Fig. 1(e) shows the result obtained from the CluSS-MCMC algorithm, where only the four 1st-order neighboring pixels are considered for the sparsity pattern. However, it results in more spurious pixels adjacent to the true pixels than that in Fig. 1(f), because the imposed pattern prior favors the appearance of non-zero valued neighboring pixels. Due to the limited number of measurements N , CS methods that do not exploit the underlying continuous structure, i.e., BPDN, BCS based on RVM and BCS based on Laplace prior, yield highly noisy results in Figs. 1(b)–1(d), respectively. In general, the reconstructed results in Figs. 1(e) and 1(f), which exploit the structure features, are much better than those from Figs. 1(b)–1(d), and the proposed algorithm shows less distortion as compared to the CluSS-MCMC algorithm.

4.2. NMSE versus number of measurements

Now we examine the required number of measurements to achieve successful reconstruction. In this experiment, the same binary image is adopted and the number of measurements K varies between 60 and 150. No noise is assumed. Fig. 2(a) shows the errorbar plot, which includes the means and standard deviations, of the NMSE obtained from 100 repeated trials with respect to the number of measurements. It is evident that the proposed algorithm exhibits significant improvement. The sparse signals can be reconstructed by only $K = 90$ measurements in the proposed algorithm, whereas the required number for the CluSS-MCMC is $K = 98$, and a much higher number is required for reliable reconstruction using the BPDN, BCS and BCS-L algorithms.

4.3. NMSE versus scale parameter

Considering the effect of the scale parameter σ_0 in the Gaussian kernel on the covariance matrix Σ , we analyze the NMSE performance versus σ_0 and provide a guidance for the selection of σ_0 . Following the common practice, we take the value of σ_0 to be a factor of 2 and, in the underlying case, between 2^{-2} and 2^6 .

Fig. 2(b) shows the reconstructed NMSE, averaged over 100 repeated trials, with respect to σ_0 . The NMSE is robust to the σ_0 when the number of measurement is sufficiently high ($K \geq 140$). As mentioned in Section 2.2, when σ_0 takes a small value, the proposed algorithm reduces to the conventional BCS algorithm, which can recover sparse signals when a high number of measurements is available. When the number of measurements is small (< 60), on the other hand, the NMSE regrades rapidly, regardless of the value of σ_0 . It is interesting to note that, when K takes a moderate value of 90, the NMSE is poor in the case a small value of $\sigma_0 < 2^3$ is used. It implies that a weak dependency between pixels leads to a poor reconstruction result. On the other hand, the sparse binary image can be successfully recovered with a small NMSE (< 0.005) when a strong dependency is imposed with $\sigma_0 \geq 2^3$. As such, this result demonstrates the importance of exploiting the image structure and the flexibility of the proposed algorithm.

5. CONCLUSION

Compressive sensing techniques are capable of reconstructing sparse signals from a small number of measurements. In addition to the sparse property, the coefficients of sparse signals often exhibit special structures. In this paper, a structured Bayesian compressive sensing algorithm is proposed to improve the reconstruction performance by exploiting the location information of pixels in a non-parametric Bayesian framework. Specifically, an adaptive logistic Gaussian kernel model, which combines the logistic model and location-dependent Gaussian kernel, was developed to encourage the underlying continuous structure of signals, and an analytical and closed-form posterior inference was derived. Experimental results demonstrated the superiority of the proposed algorithm over the state-of-the-art sparse reconstruction algorithms.

6. REFERENCES

- [1] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Info. Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.
- [4] L. Jacob, G. Obozinski, and J. Vert, "Group Lasso with overlap and graph Lasso," in *Proc. Int. Conf. Machine Learning*, Montreal, Canada, June 2009.
- [5] E. V. D. Berg and M. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM J. Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.
- [6] P. R. Gill, A. Wang, and A. Molnar, "The in-crowd algorithm for fast basis pursuit denoising," *IEEE Trans. Signal Proc.*, vol. 59, no. 10, pp. 4595–4605, 2011.
- [7] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Stat. Soc. Series B*, vol. 68, no. 1, pp. 49–67, 2006.
- [8] M. E. Tipping, "Sparse Bayesian shrinkage and selection learning and the relevance vector machine," *J. Machine Learning Research*, vol. 1, no. 9, pp. 211–244, 2001.
- [9] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Proc.*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [10] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *IEEE Trans. Image Proc.*, vol. 19, no. 1, pp. 53–63, 2010.
- [11] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vector using sparse Bayesian learning," *IEEE J. Sel. Topics in Signal Proc.*, vol. 5, no. 5, pp. 1–15, 2011.
- [12] Z. Zhang and B. D. Rao, "Extension of SBL algorithm for the recovery of block sparse signals with inter-block correlation," *IEEE Trans. Signal Proc.*, vol. 92, no. 7, pp. 1580–1590, 2012.
- [13] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Proc.*, vol. 57, no. 1, pp. 92–106, 2009.
- [14] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE Trans. Signal Proc.*, vol. 62, no. 11, pp. 2906–2921, 2014.
- [15] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Complex multitask Bayesian compressive sensing," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014.
- [16] M. Leigsnering, F. Ahmad, M. G. Amin, and A. M. Zoubir, "Compressive sensing based specular multipath exploitation for through-the-wall radar imaging," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013.
- [17] Q. Wu, Y. D. Zhang, M. G. Amin, and F. Ahmad, "Through-the-wall radar imaging based on modified bayesian compressive sensing," in *Proc. IEEE China Summit and Int. Conf. Signal and Inform. Proc.*, Xi'an, China, May 2014.
- [18] Q. Wu, Y. D. Zhang, M. G. Amin, A. Golato, F. Ahmad, and S. Santhanam, "Structural health monitoring exploiting MIMO ultrasonic sensing and group sparse Bayesian learning," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2014.
- [19] Q. Wu, Y. D. Zhang, F. Ahmad, and M. G. Amin, "Compressive sensing based high-resolution polarimetric through-the-wall radar imaging exploiting target characteristics," *IEEE Antennas and Wireless Propagat. Lett.*, in press.
- [20] Q. Wu, Y. D. Zhang, and M. G. Amin, "Continuous structure based Bayesian compressive sensing for sparse reconstruction of time-frequency distributions," in *Proc. Int. Conf. Digital Signal Proc.*, Hong Kong, China, Aug. 2014.
- [21] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. Signal Proc.*, vol. 57, no. 9, pp. 3488–3497, 2009.
- [22] L. Yu, H. Sun, J. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Proc.*, vol. 92, no. 1, pp. 259–269, 2012.
- [23] L. Yu, J. P. Barbot, G. Zheng, and H. Sun, "Compressive sensing for cluster structured sparse signals: Variational Bayes approach," Technical Report, 2011. Available at http://hal.archives-ouvertes.fr/docs/00/57/39/53/PDF/class_vb.pdf.
- [24] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Mutli-task Bayesian compressive sensing exploiting intra-task dependency," *IEEE Signal Proc. Lett.*, vol. 22, no. 4, pp. 430–434, 2014.
- [25] E. I. George and R. E. McCulloch, "Variable selection via Gibbs sampling," *J. American Stat. Assoc.*, vol. 88, no. 423, pp. 881–889, 1993.
- [26] T. J. Mitchell and J. J. Beauchamp, "Bayesian variable selection in linear regression," *J. American Stat. Assoc.*, vol. 83, no. 404, pp. 1023–1032, 1988.
- [27] J. H. Albert and S. Chib, "Bayesian analysis of binary and polychotomous response data," *J. American Stat. Assoc.*, vol. 88, no. 422, pp. 669–679, 1993.
- [28] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [29] N. G. Polson, J. G. Scott, and J. Windle, "Bayesian inference for logistic models using Polya-Gamma latent variables," *J. American Stat. Assoc.*, vol. 108, no. 504, pp. 1339–1349, 2013.
- [30] W. R. Trench, "An algorithm for the inversion of finite Toeplitz matrices," *J. SAIM*, vol. 12, no. 3, pp. 515–522, 264.