



Dependent Hierarchical Beta Process for Image Interpolation and Denoising

¹Mingyuan Zhou, ²Hongxia Yang, ³Guillermo Sapiro,
²David Dunson and ¹Lawrence Carin

¹Department of Electrical & Computer Engineering, Duke University

²Department of Statistical Science, Duke University

³Department of Electrical & Computer Engineering, University of Minnesota

AI & Statistics, Ft. Lauderdale, April 12 2011

Outline

- Introduction
- Dependent Hierarchical Beta Process (dHBP)
- Dictionary Learning with dHBP
- Image Interpolation and Denoising
- Conclusions

Introduction: *Background*

- Dictionary learning and sparse coding

$$\min_{\mathbf{D}, \mathbf{W}} \|\mathbf{X} - \mathbf{D}\mathbf{W}\|_F + \lambda \sum_{i=1}^N |\mathbf{w}_i|_1$$

- Sparse factor analysis model
(Factor/feature/dish/dictionary atom)
- Indian Buffet process and beta process

Introduction: *Background*

- Beta process and Bernoulli process (Thibaux & Jordan AISTATS2007)

$$B \sim \text{BP}(c, B_0)$$

$$X_i \sim \text{BeP}(B)$$

$$B = \sum_{k=1}^{\infty} \pi_k \delta_{d_k}$$
$$X_i = \sum_{k=1}^{\infty} z_{ik} \delta_{d_k}$$

$$B|\{X_i\}_{i=1,n} \sim \text{BP}\left(c+n, \frac{c}{c+n}B_0 + \frac{1}{c+n} \sum_{i=1}^n X_i\right)$$

- Indian Buffet process (Griffiths & Ghahramani 2005)

$$X_{n+1}|\{X_i\}_{i=1,n} \sim \text{BeP}\left(\frac{c}{c+n}B_0 + \frac{1}{c+n} \sum_{i=1}^n X_i\right)$$

Introduction: *Motivation*

- Exchangeability assumption is not true

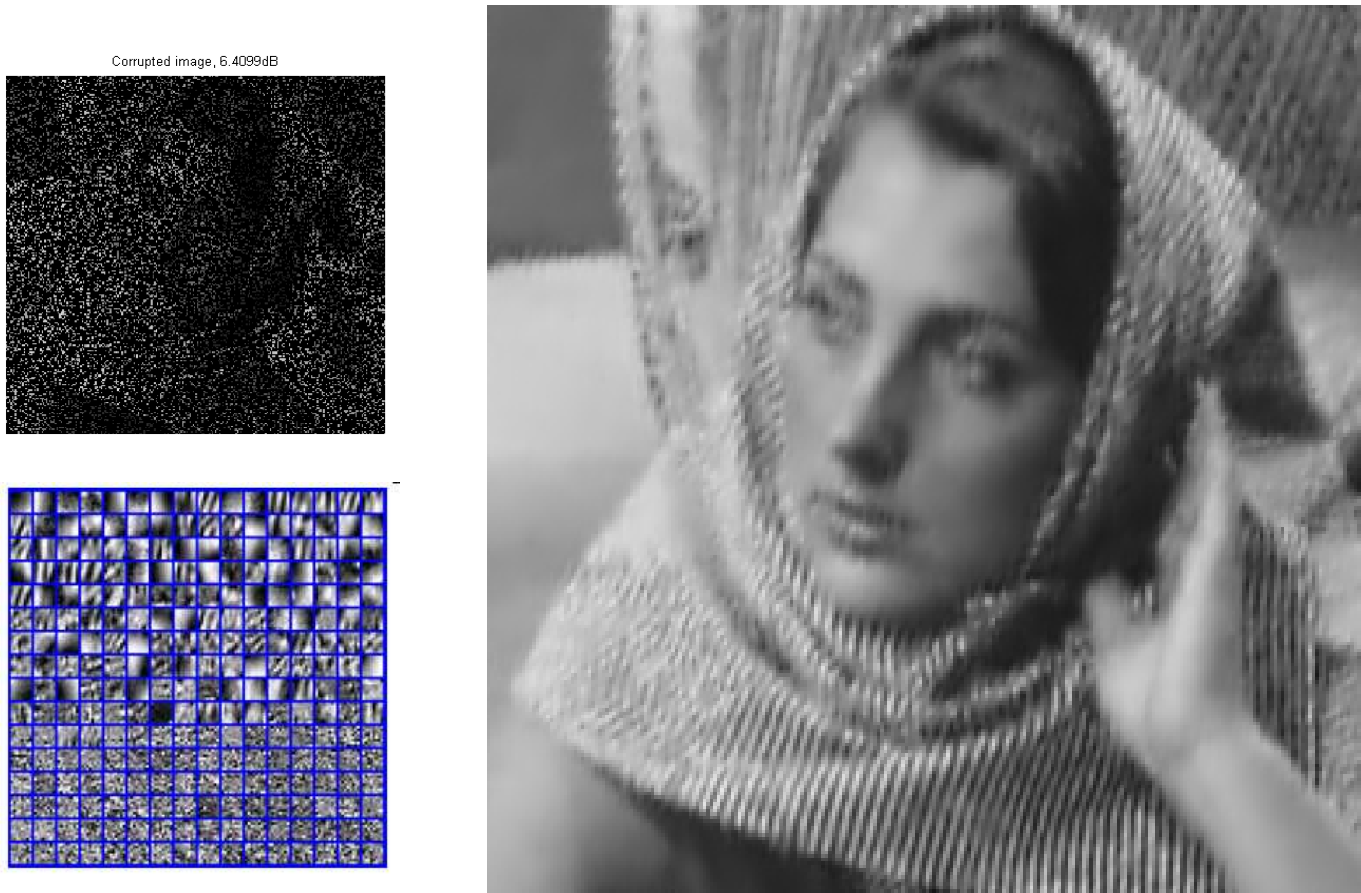


Image interpolation with BPFA (Zhou et al. 2009)
80% pixels are missing at random

Introduction: *Covariate Dependence*

- Dependent Dirichlet process
 - MacEachern (1999), Duan et al. (2007), Griffin & Steel (2006)
- Non-exchangeable IBP
 - Phylogenetic IBP (Miller, Griffiths & Jordan 2008)
 - Dependent IBP (Williamson, Orbanz & Ghahramani 2010)
- Bayesian density regression (Dunson & Pillai 2007)

Review of Beta Process (Thibaus & Jordan 2007)

- A beta process is a positive Levy process, whose Levy measure lives on $\Omega \otimes [0,1]$ and can be expressed as

$$\nu(d\omega, dp) = cp^{-1}(1-p)^{c-1}dpB_0(d\omega)$$

- If the base measure B_0 is continuous, then $B = \sum_{k=1}^{\infty} p_k \delta_{w_k}$, and p_k is drawn from a degenerate beta distribution parameterized by c

- If $B_0 = \sum_{k=1}^{\infty} q_k \delta_{w_k}$, then

$$B = \sum_{k=1}^{\infty} p_k \delta_{w_k}, \text{ and } p_k \sim \text{Beta}(cq_k, c(1-q_k))$$

Dependent Hierarchical Beta Process

- Random walk matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$

$$a_{ij} = \mathcal{K}(\ell_i, \ell_j) / \sum_{j'=1}^N \mathcal{K}(\ell_i, \ell_{j'})$$

- dHBP

$$X_i \sim \text{BeP}(B_i)$$

$$B_i = \sum_{j=1}^N a_{ij} B_j^* , \quad B_j^* \sim \text{BP}(c_1, B) , \quad B \sim \text{BP}(c_0, B_0)$$

- Covariate-dependent correlations

$$\text{corr}\{B_i(S), B_{i'}(S)\} = \frac{\langle \mathbf{a}_i, \mathbf{a}_{i'} \rangle}{\|\mathbf{a}_i\| \cdot \|\mathbf{a}_{i'}\|}$$

Dictionary Learning with dHBP

$$\begin{aligned} \mathbf{x}_i &\sim \mathcal{N}(\mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i), \gamma_\epsilon^{-1} \mathbf{I}_P) \\ \mathbf{d}_k &\sim \mathcal{N}(0, P^{-1} \mathbf{I}_P), \quad \mathbf{s}_i \sim \mathcal{N}(0, \gamma_s^{-1} \mathbf{I}_K) \end{aligned}$$

$$\begin{aligned} z_{ik} &\sim \text{Bernoulli}(\pi_k) \\ \pi_k &\sim \text{Beta}(c\eta, c(1 - \eta)) \end{aligned}$$

BP

$$\begin{aligned} z_{ik} &\sim \text{Bernoulli}(\pi_{ik}), \quad \pi_{ik} = \sum_{j \in \mathcal{Q}_i} a_{ij} \pi_{jk}^* \\ \pi_{jk}^* &\sim \text{Beta}(c_1 \eta_k, c_1 (1 - \eta_k)) \\ \eta_k &\sim \text{Beta}(c_0 \eta_0, c_0 (1 - \eta_0)). \\ a_{ij} &= \mathcal{K}(\ell_i, \ell_j) / \sum_{j'=1}^N \mathcal{K}(\ell_i, \ell_{j'}) \\ \mathcal{K}(\ell_i, \ell_j) &= \delta(j \in \mathcal{Q}_i) \exp(-\|\ell_i - \ell_j\|_2 / \sigma) \end{aligned}$$

dHBP

Missing Data and Outliers

- Missing data

- Full data: \mathbf{x}_i

- Observed: $\mathbf{y}_i = \Sigma_i \mathbf{x}_i$, Missing: $\bar{\Sigma}_i \mathbf{x}_i$

$$\begin{aligned} \mathcal{N}(\mathbf{x}_i; \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i), \gamma_\epsilon^{-1} \mathbf{I}_P) = \\ \mathcal{N}(\Sigma_i^T \mathbf{y}_i; \Sigma_i^T \Sigma_i \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i), \Sigma_i^T \Sigma_i \gamma_\epsilon^{-1} \mathbf{I}_P) \\ \mathcal{N}(\bar{\Sigma}_i^T \bar{\Sigma}_i \mathbf{x}_i; \bar{\Sigma}_i^T \bar{\Sigma}_i \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i), \bar{\Sigma}_i^T \bar{\Sigma}_i \gamma_\epsilon^{-1} \mathbf{I}_P) \end{aligned}$$

- Sparse spiky noise

$$\mathbf{x}_i = \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i) + \epsilon_i + \mathbf{v}_i \odot \mathbf{m}_i$$

$$\mathbf{v}_i \sim \mathcal{N}(0, \gamma_v^{-1} \mathbf{I}_P), \quad m_{ip} \sim \text{Bernoulli}(\pi'_{ip}), \quad \pi'_{ip} \sim \text{Beta}(a_0, b_0)$$

- Recoverd data:

$$\hat{\mathbf{x}}_i = \mathbf{D}(\mathbf{s}_i \odot \mathbf{z}_i)$$

MCMC Inference

- Independence chain Metropolis-Hastings

$$\pi_{jk}^* \sim \text{Beta}\left(c_1\eta_k + \sum_{i:\{j \in \mathcal{Q}_i\}} z_{ik}, c_1(1 - \eta_k) + \sum_{i:\{j \in \mathcal{Q}_i\}} (1 - z_{ik})\right).$$

- Slice sampling

$$p(\eta_k | -) \propto \eta_k^{c_0\eta_0 - 1} (1 - \eta_k)^{c_0(1 - \eta_0) - 1} \sin^N(\pi\eta_k) \exp\left(c_1\eta_k \sum_{j=1}^N \log\left(\frac{\pi_{jk}^*}{1 - \pi_{jk}^*}\right)\right)$$

- Gibbs sampling

Experiments

- Image interpolation
 - Missing pixels
 - Locations of missing pixels are known
- Image denoising
 - WGN + sparse spiky noise
 - Amplitudes unknown
 - Locations of spiky noise are unknown
- Covariates: patch spatial locations

Image Interpolation: BP vs. dHBP

BP: 26.9 dB



dHBP: 29.92 dB

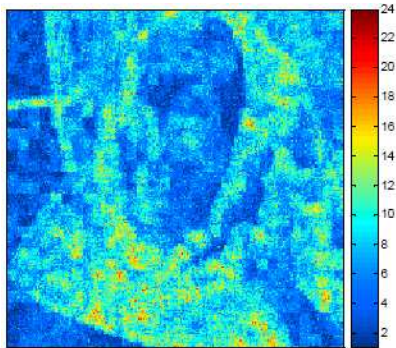


80% pixels missing at random

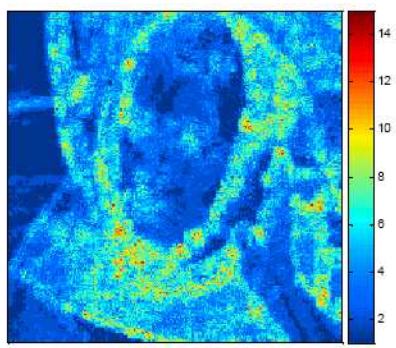
Image Interpolation: BP vs. dHBP

Observed	BP recovery
dHBP recovery	Original

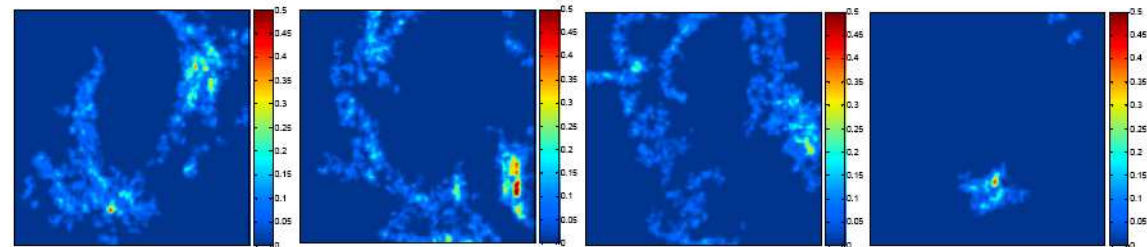
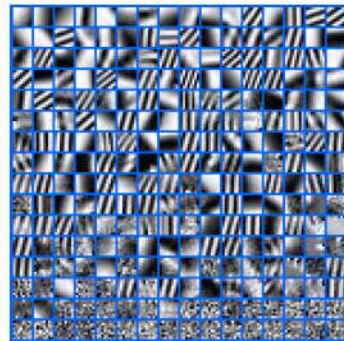
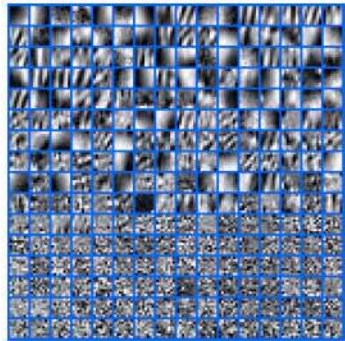
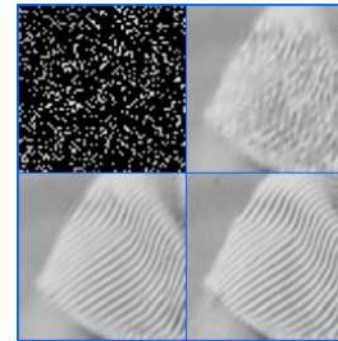
BP atom usage



dHBP atom usage



dHBP recovery



BP dictionary

dHBP dictionary

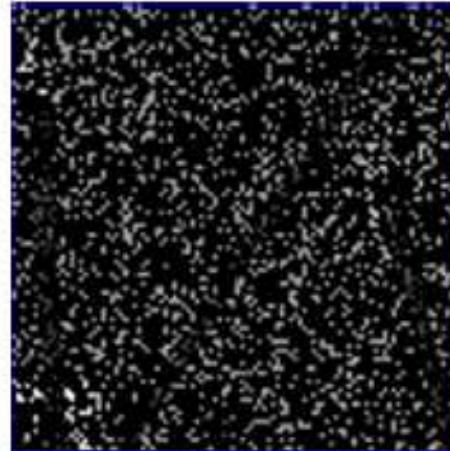
Dictionary atom activation probability map

Image Interpolation: BP vs. dHBP

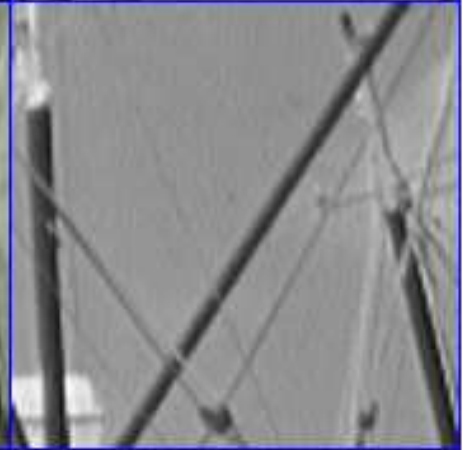
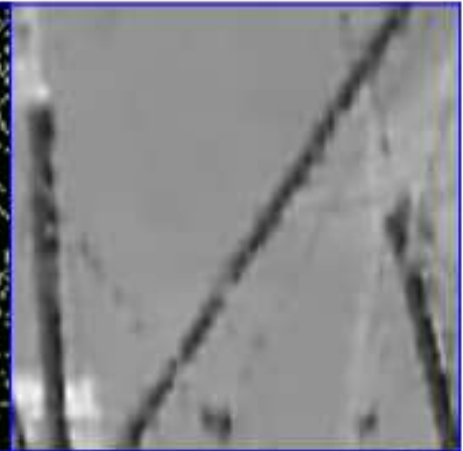
dHBP recovery



Observed (20%)



BP recovery



dHBP recovery

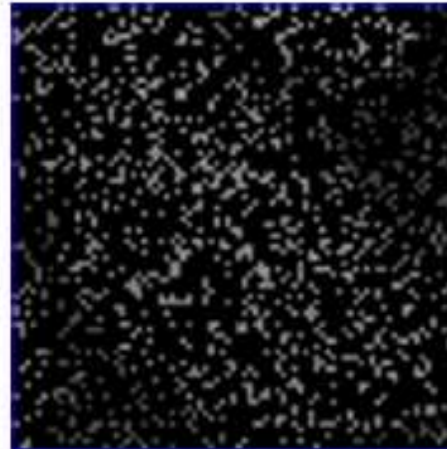
Original

Image Interpolation: BP vs. dHBP

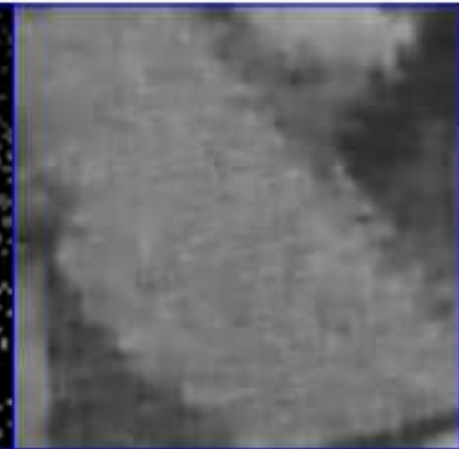
dHBP recovery



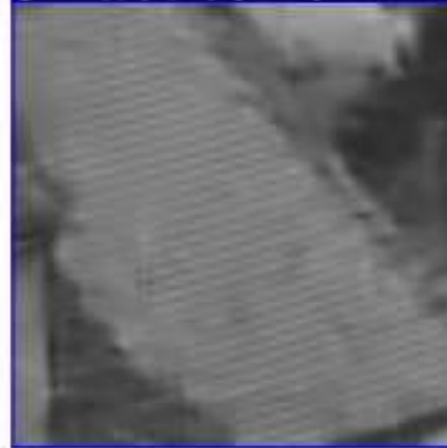
Observed (20%)



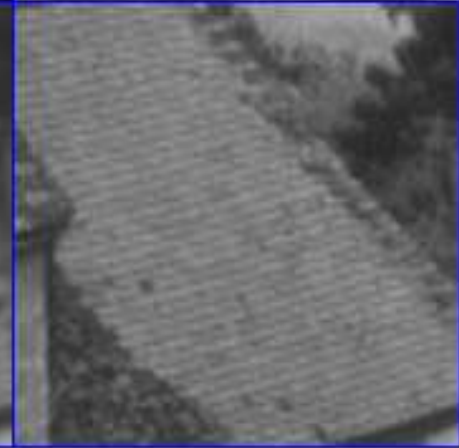
BP recovery



dHBP recovery



Original

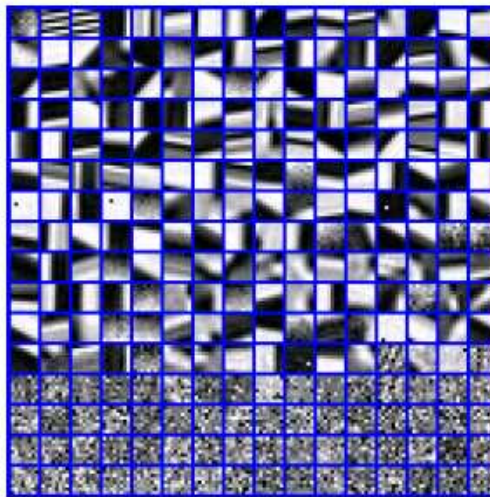


Spiky Noise Removal: BP vs. dHBP

Original image



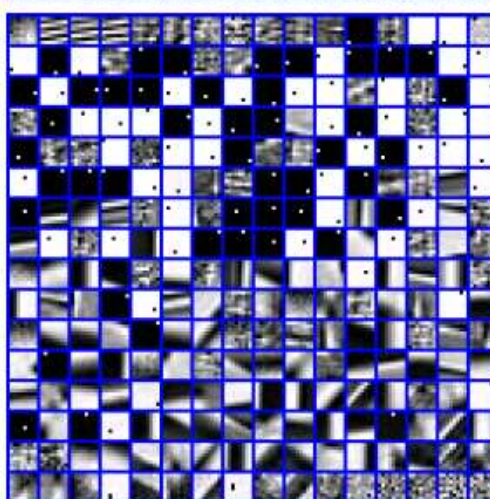
dHBP dictionary



dHBP denoised image



Noisy image (WGN + Sparse Spiky noise)



BP dictionary



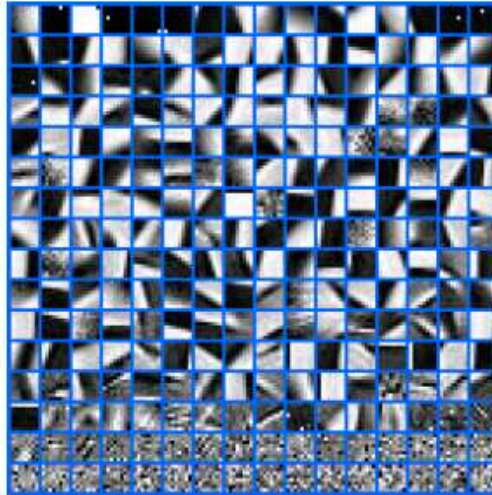
BP denoised image

Spiky Noise Removal: BP vs. dHBP

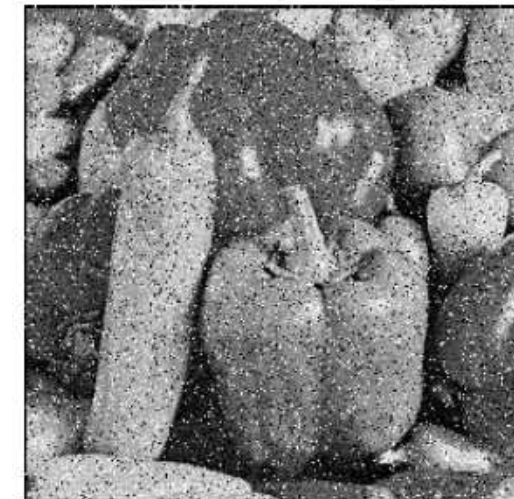
Original image



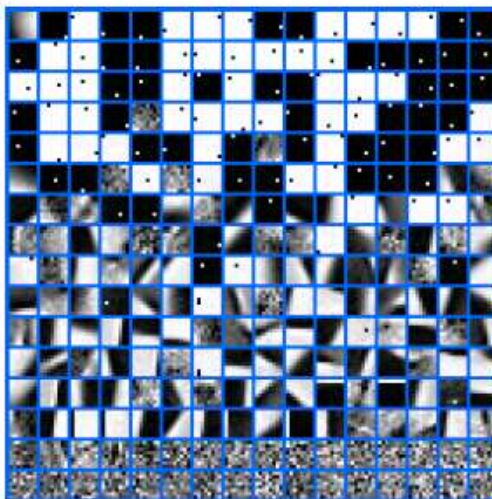
dHBP dictionary



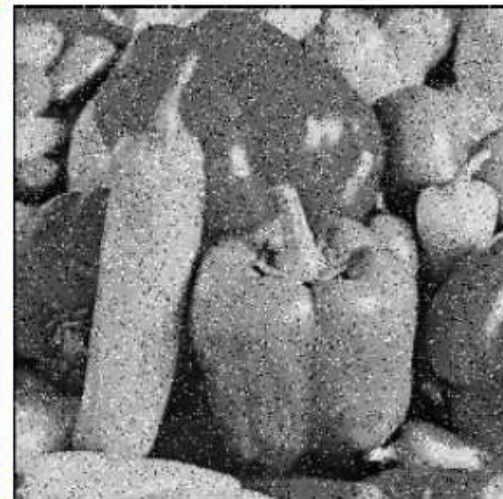
dHBP denoised image



Noisy image (WGN + Sparse Spiky noise)



BP dictionary



BP denoised image

Future Work

- Landmark-dHBP
 - J landmarks
 - $J \ll N$
- Locality constraint for manifold learning
 - Covariates: cosine distance between samples
 - Dictionary atoms look like the data
- Variational inference, online learning
- Other applications
 - Super-resolution
 - Deblurring
 - Video background & foreground modeling

Conclusions

- A dependent hierarchical beta process is proposed
- Efficient hybrid MCMC inference is presented
- Encouraging performance is demonstrated on image-processing applications