Dependent Hierarchical Beta Process for Image Interpolation and Denoising

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

[1]Department of Electrical & Computer Engineering, Duke University
[2]Department of Statistical Science, Duke University
[3]Department 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

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

- 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)

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

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

$$\boxed{\begin{aligned} B &= \sum_{k=1}^{\infty} \pi_k \delta_{d_k} \\ X_i &= \sum_{k=1}^{\infty} z_{ik} \delta_{d_k} \end{aligned}}$$

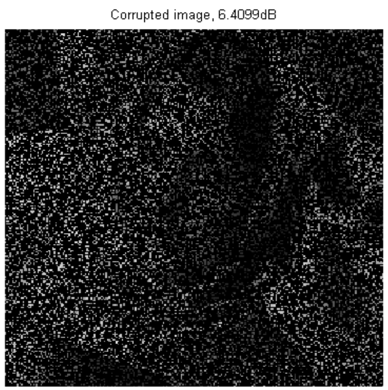
$$B | \{X_i\}_{i=1,n} \sim \mathrm{BP}\left(c+n, \frac{c}{c+n} B_0 + \frac{1}{c+n} \sum_{i=1}^{n} X_i\right)$$

- Indian Buffet process (Griffiths & Ghahramani 2005)

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

Introduction: *Motivation*

- Exchangeability assumption is not true



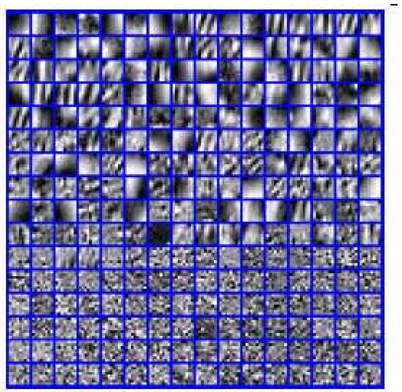


Image interpolation with BPFA (Zhou et al. 2009)
80% pixels are missing at random

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

$$\nu(d\omega, dp) = cp^{-1}(1-p)^{c-1} dp B_0(d\omega)$$

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

$$a_{ij} = \mathcal{K}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j) \Big/ \sum_{j'=1}^{N} \mathcal{K}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_{j'})$$

- dHBP

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

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

- Covariate-dependent correlations

$$\mathrm{corr}\{B_i(S), B_{i'}(S)\} = \frac{< \boldsymbol{a}_i, \boldsymbol{a}_{i'} >}{\|\boldsymbol{a}_i\| \cdot \|\boldsymbol{a}_{i'}\|}$$

Dictionary Learning with dHBP

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

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

BP

$$z_{ik} \sim \text{Bernoulli}(\pi_{ik}), \qquad \pi_{ik} = \sum_{j \in \mathcal{Q}_i} a_{ij}\pi_{jk}^*$$
$$\pi_{jk}^* \sim \text{Beta}(c_1\eta_k, c_1(1-\eta_k))$$
$$\eta_k \sim \text{Beta}(c_0\eta_0, c_0(1-\eta_0)).$$
$$a_{ij} = \mathcal{K}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j) \Big/ \sum_{j'=1}^{N} \mathcal{K}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_{j'})$$
$$\mathcal{K}(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j) = \delta(j \in \mathcal{Q}_i)\exp(-\|\boldsymbol{\ell}_i - \boldsymbol{\ell}_j\|_2/\sigma)$$

dHBP

Missing Data and Outliers

- Missing data
  - Full data: $\boldsymbol{x}_i$
  - Observed: $\boldsymbol{y}_i = \boldsymbol{\Sigma}_i \boldsymbol{x}_i$, Missing: $\bar{\boldsymbol{\Sigma}}_i \boldsymbol{x}_i$

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

- Sparse spiky noise

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

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

- Recoverd data:

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

MCMC Inference

- Independence chain Metropolis-Hastings

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

- Slice sampling

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

- 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

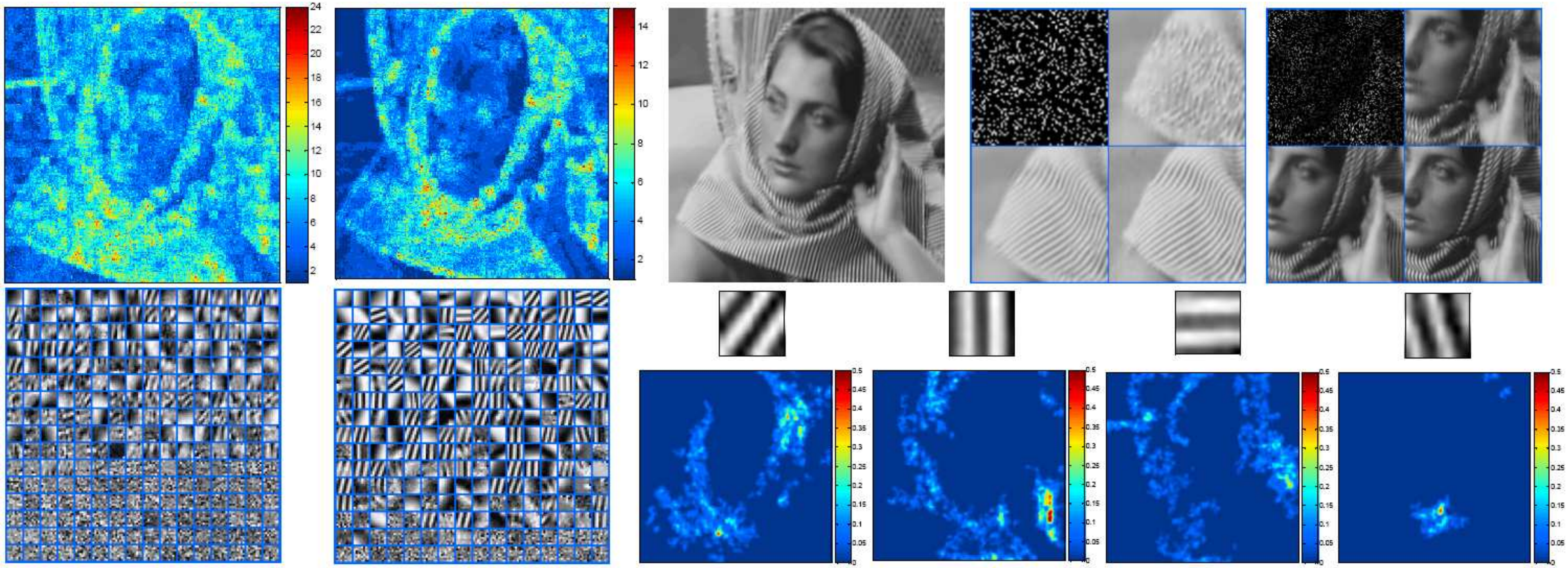
dHBP: 29.92 dB

80% pixels missing at random

Image Interpolation: BP vs. dHBP

Observed	BP recovery
dHBP recovery	Original

BP atom usage | dHBP atom usage | dHBP recovery



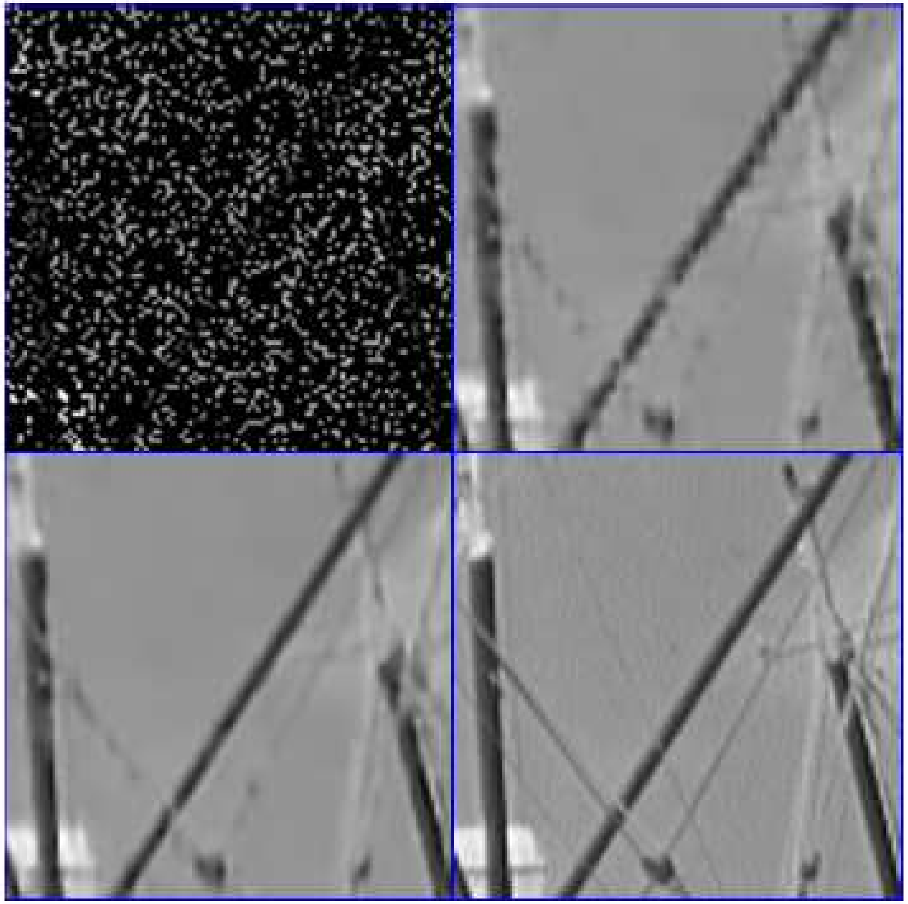
BP dictionary | dHBP dictionary | Dictionary atom activation probability map

Image Interpolation: BP vs. dHBP

dHBP recovery

Observed (20%)

BP recovery



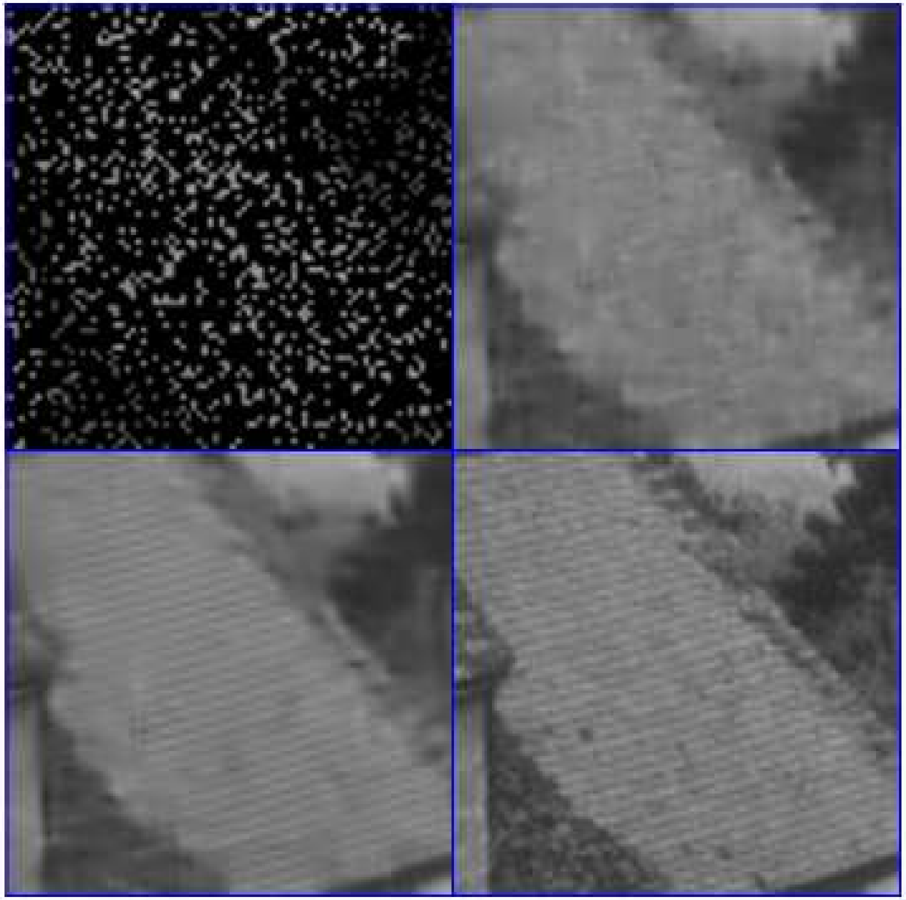
dHBP recovery

Original

Image Interpolation: BP vs. dHBP

dHBP recovery

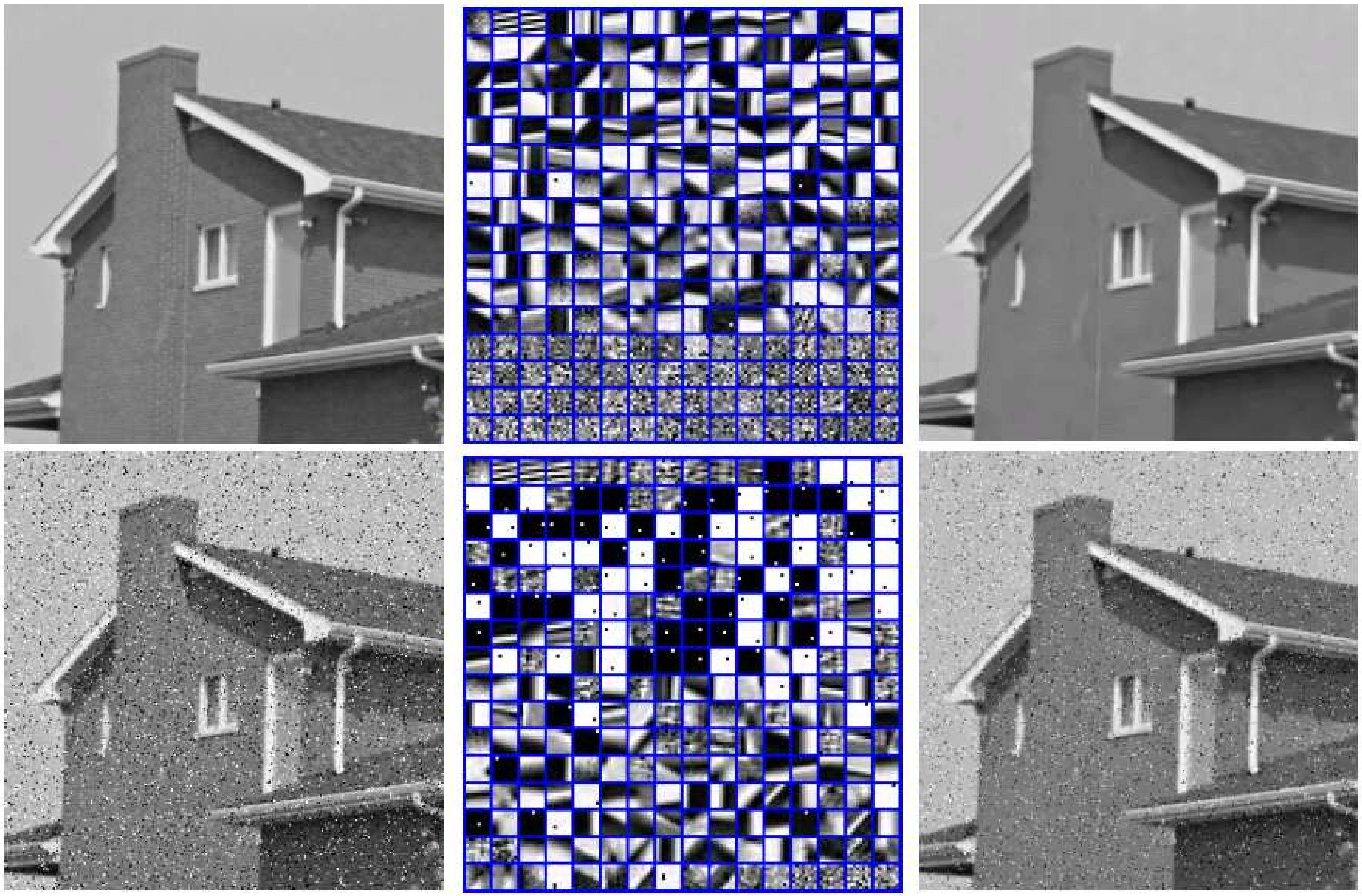
Observed (20%) BP recovery



dHBP recovery Original

Spiky Noise Removal: BP vs. dHBP

Original image | dHBP dictionary | dHBP denoised image

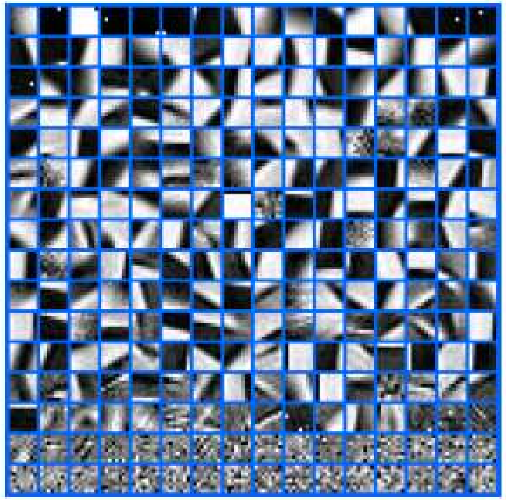


Noisy image (WGN + Sparse Spiky noise) | BP dictionary | BP denoised image

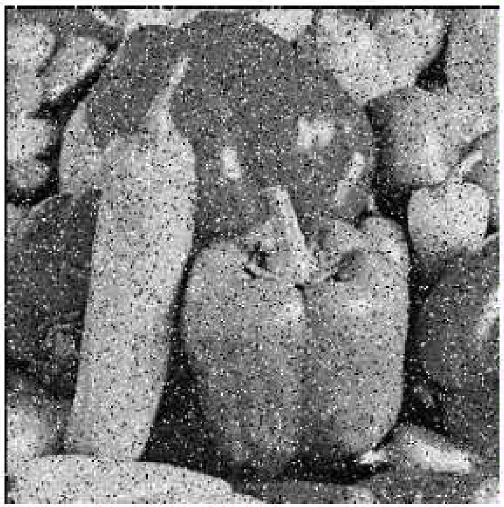
Spiky Noise Removal: BP vs. dHBP

Original image

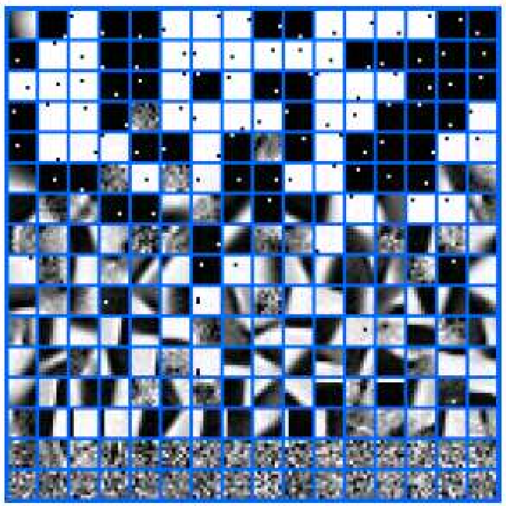
dHBP dictionary



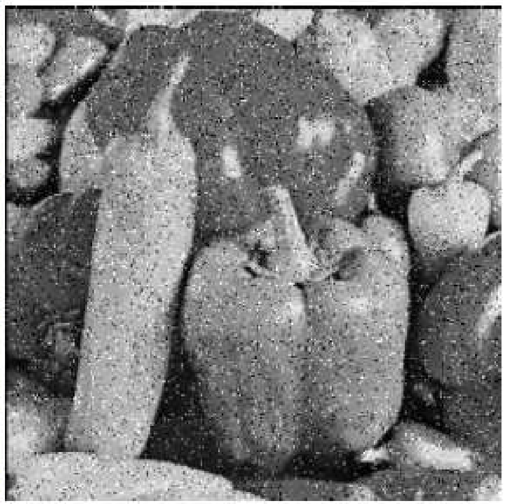
dHBP denoised image



Noisy image (WGN + Sparse Spiky noise)



BP dictionary



BP denoised image

Future Work

- Landmark-dHBP
  - *J* landmarks
  - $J << N$
- 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