MAP Estimation with Gaussian Mixture Markov Random Field Model for Inverse Problems

Ruoqiao Zhang[1], *Charles A. Bouman*[1],
Jean-Baptiste Thibault[2], *and Ken D. Sauer*[3]

1. *Department of Electrical and Computer Engineering, Purdue University*
2. *GE Healthcare Technologies*
3. *Department of Electrical Engineering, University of Notre Dame*

Better Prior Models

- Why do we need better prior models?
  - Better prior models will be needed as data becomes sparser
  - Models must be adaptive to different classes of images
  - Low, mid, and high level representations are needed
- What is needed?
  - More expressive models of images
  - Trained on real data (scientific/medical data)
  - Computationally efficient to implement
- Promising recent approaches:
  - Dictionary learning; kSVD; Non-local means; BM3D; Bilateral filters
  - Many of these are not really consistent prior models
  - Do not quantify multivariate distribution of image

Mission statement:

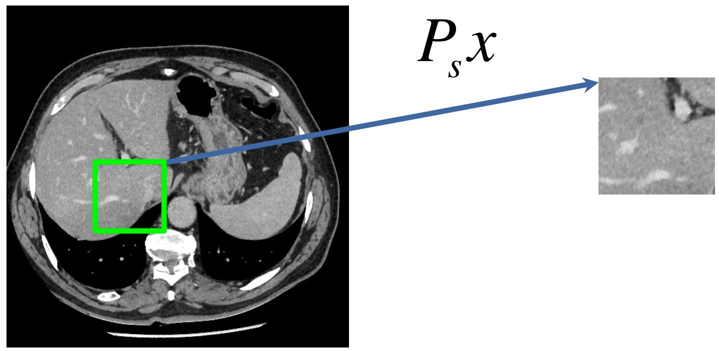
Formulate a single, consistent, robust, and expressive prior model for any image, *x*, that can be used in computationally efficient Bayesian estimation algorithms.

$$p_\theta(x) = \frac{1}{z} \exp\{-u(x)\}$$

θ - parameterizes model

So we need to construct *u(x)*

Modeling Patches with Gaussian Mixture

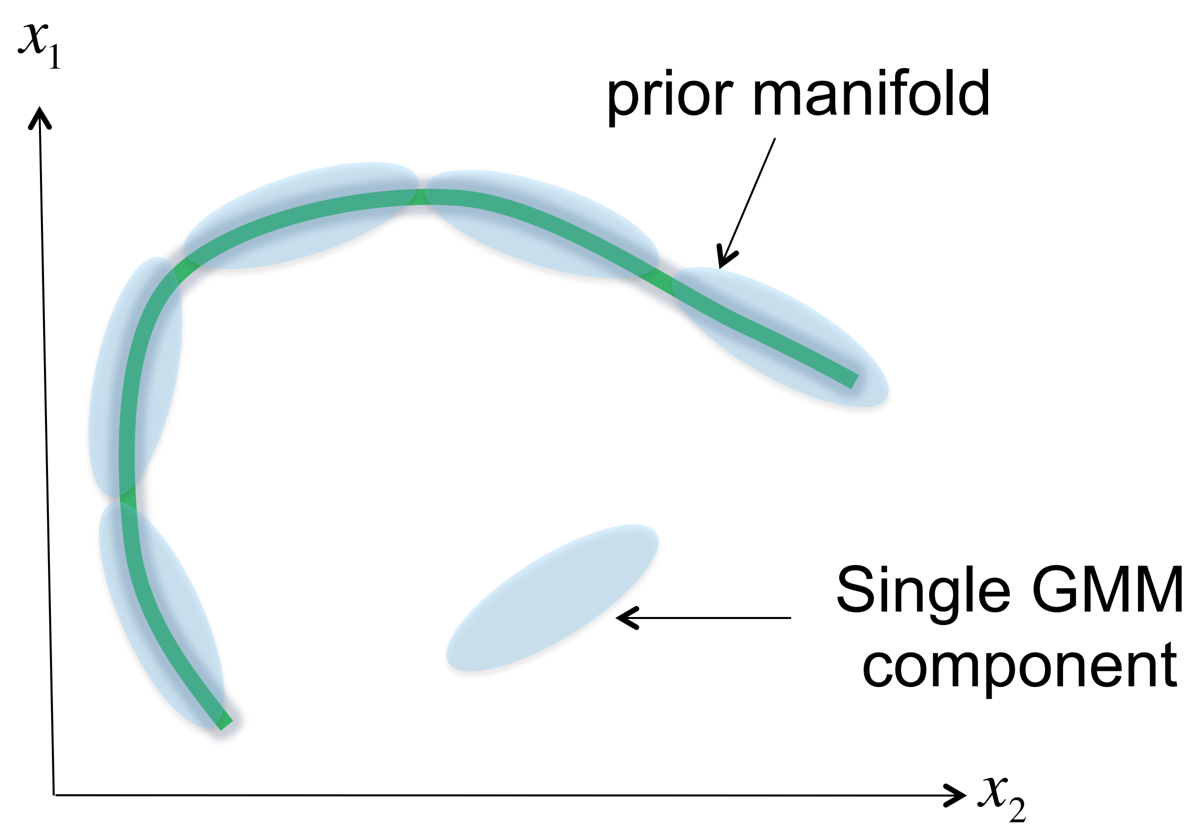


- Gaussian mixture model (GMM) for image patches

$$g\left(P_s x\right) = \sum_{k=0}^{K-1} \pi_k \frac{1}{(2\pi)^{p/2}} \left|B_k\right|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left\| P_s x - \mu_k \right\|_{B_k}^2 \right\}$$

Advantage: We can approximate any distribution with GMM

Advantage of GMM Patch Model

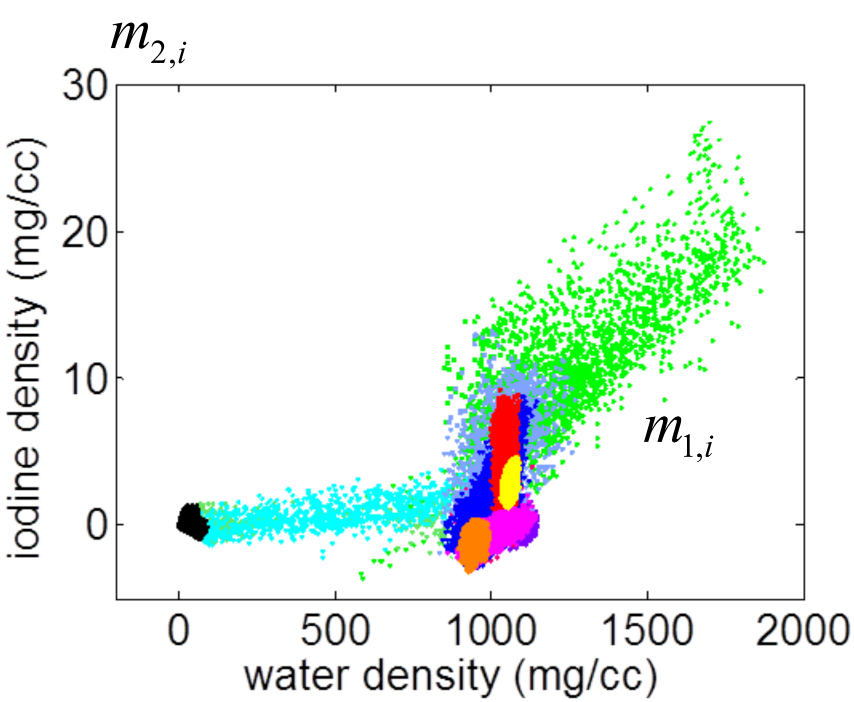


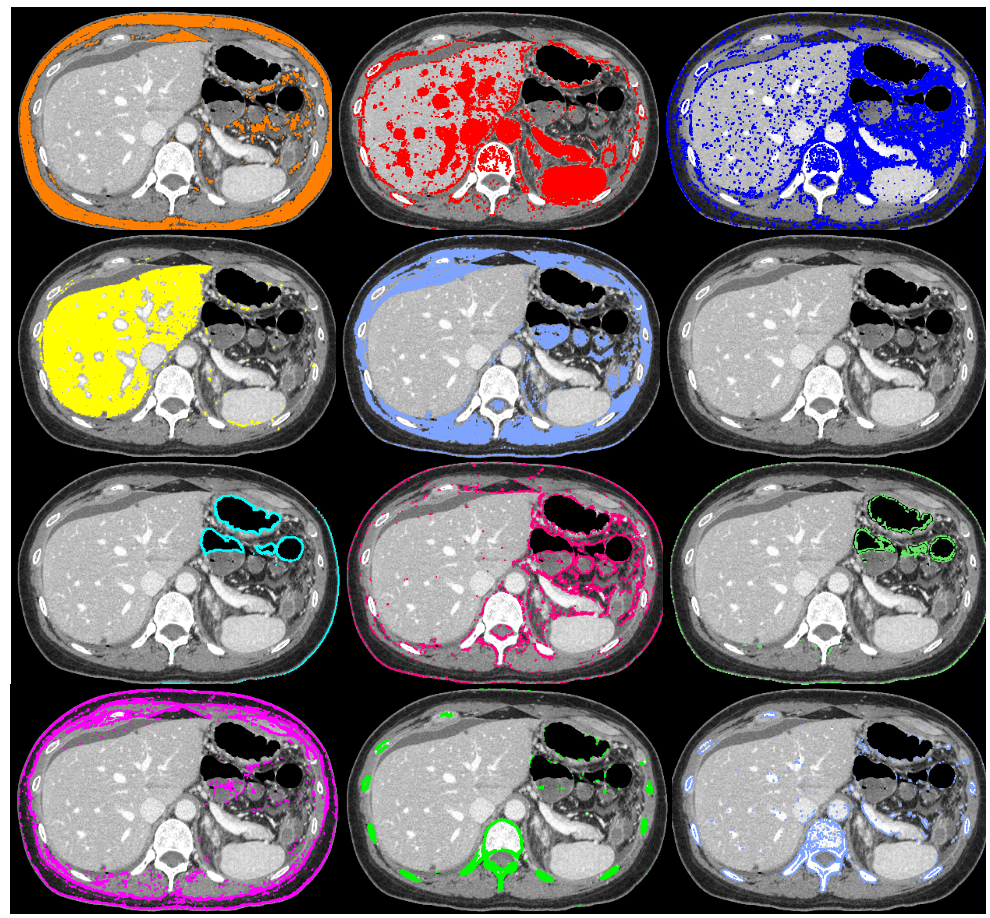
- Advantage of multivariate Gaussian mixture
  - Can model any distribution with enough GM components
  - Capture multivariate distribution of a patch
  - Model interaction between density and texture

GMM with 2x2 Image Patch

- Dual energy CT example
  - 12 clusters.
  - Display 2 dimensions out of 8
  - Water/iodine decomposition

    color-coded scatter plot



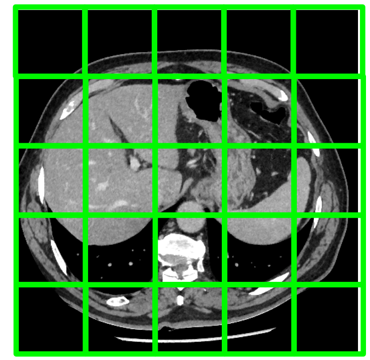


Question?

How to build a consistent image model out of GMM patches?

Model 1:
Non-Overlapping Tiling with GMM Patches

- Tile image with non-overlapping patches



- Image distribution

$$p_0(x) = \prod_{s \in S_0} g\left(P_s x\right)$$

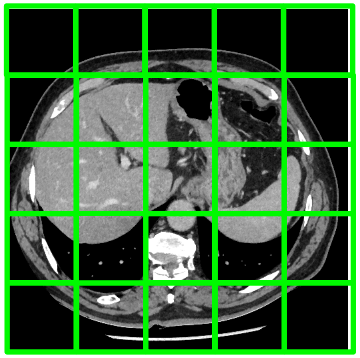
- Energy function

$$u(x) = \sum_{s \in S_0} V\left(P_s x\right)$$

$$V\left(P_s x\right) = \log g\left(P_s x\right)$$

Model 1:
Non-Overlapping Tiling with GMM Patches

- Tile image with non-overlapping patches



- Energy function

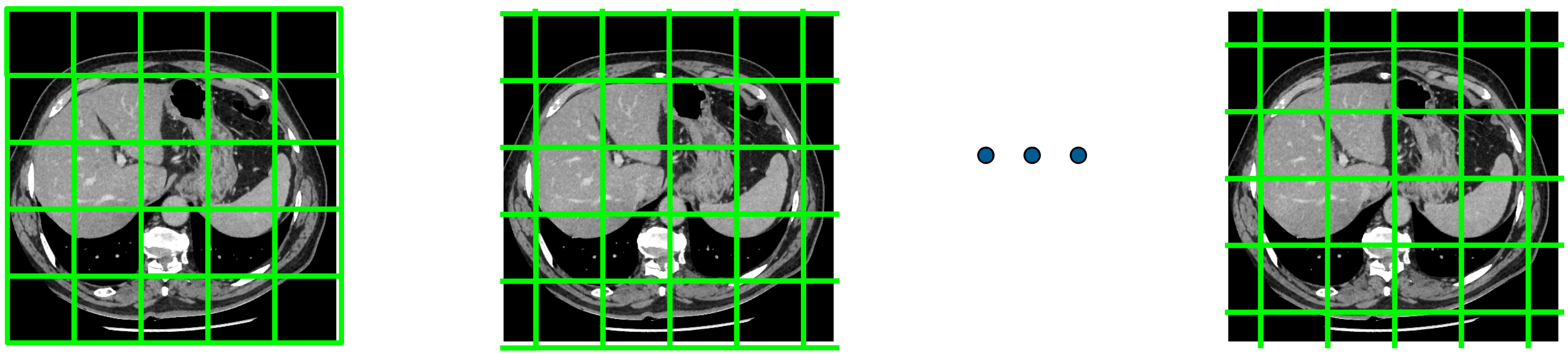
$$u(x) = \sum_{s \in S_0} V\left(P_s x\right)$$

sums over
non-overlapping patches

$$V\left(P_s x\right) = \log\left\{\sum_{k=0}^{K-1} \pi_k \frac{1}{(2\pi)^{p/2}} \left|B_k\right|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\|P_s x - \mu_k\right\|_{B_k}^2\right\}\right\}$$

Model 2:
Non-Overlapping Tiling with GM Patches

- M^2 different tilings with non-overlapping patches



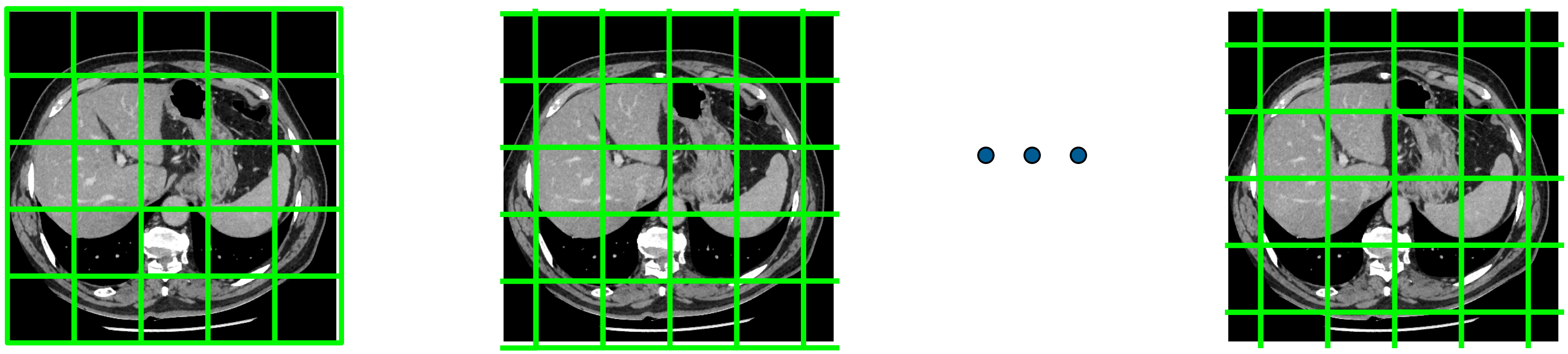
$$p_0(x) = \prod_{s \in S_0} g(P_s x) \qquad p_0(x) = \prod_{s \in S_1} g(P_s x) \qquad p_{M^2}(x) = \prod_{s \in S_{M^2}} g(P_s x)$$

- Form a single distribution using the "Product of Experts"

$$p(x) = \frac{1}{z}\left(\prod_{i=0}^{M^2} p_i\right)^{1/M^2} = \frac{1}{z}\left(\prod_{s \in S} g(P_s x)\right)^{1/M^2}$$

Model 2:
Non-Overlapping Tiling with GM Patches

- M^2 different tilings with non-overlapping patches



$$p_0(x) = \prod_{s \in S_0} g\left(P_s x\right) \qquad p_0(x) = \prod_{s \in S_1} g\left(P_s x\right) \qquad p_{M^2}(x) = \prod_{s \in S_{M^2}} g\left(P_s x\right)$$

- "Product of Experts" energy function

$$u(x) = \frac{1}{M^2} \sum_{s \in S} V\left(P_s x\right)$$

Final GM-MRF Model

- Prior model

$$p(x) = \frac{1}{z} \exp\{-u(x)\}$$

- Energy function

corrects for patch overlap

$$u(x) = \frac{1}{M^2} \sum_{s \in S} V(P_s x)$$

sums over all patches

- Log GMM

$$V(P_s x) = -\log\left(\sum_{k=0}^{K-1} \pi_k \frac{1}{(2\pi)^{p/2}} |B_k|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \| P_s x - \mu_k \|_{B_k}^2 \right\} \right)$$

Final GM-MRF Model

- GM-MRF prior model

$$p(x) = \frac{1}{z} \exp\left\{ -\frac{1}{M^2} \sum_{s \in S} V(P_s x) \right\}$$

$$V(P_s x) = -\log\left(\sum_{k=0}^{K-1} \pi_k \frac{1}{(2\pi)^{p/2}} |B_k|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \|P_s x - \mu_k\|_{B_k}^2 \right\} \right)$$

- OK, but …
  - Is this really an MRF?
    - Yes, with an *(2M-1)x(2M-1)* neighborhood.
  - How do I train the model?
    - Just use your favorite GMM app to fit to patch data.
  - How do I use this?
    - Hmm, good point. We’ll give you a surrogate function.

MAP Estimation with GM-MRF Model

- MAP estimate

$$\hat{x} = \arg\min_x \left\{ -\log p(y|x) + u(x) \right\}$$

- MAP estimate with surrogate prior

x' is the current state of x

$$\hat{x} = \arg\min_x \left\{ -\log p(y|x) + u(x;x') \right\}$$

where

$$u(x') = u(x';x')$$

$$u(x) \geq u(x;x')$$

Perform surrogate optimization iteratively, updating x' with each iteration

How do we find $u(x';x')$?

Lemma: Surrogate Functions for Logs of Exponential Mixtures

Lemma: surrogate functions for logs of exponential mixtures
Let $f : \Re^N \to \Re$ be a function of the form,

$$f(x) = \sum_k w_k \exp\{-v_k(x)\} \tag{13}$$

where $w_k \in \Re^+$, $\sum_k w_k > 0$, and $v_k : \Re^N \to \Re$. Furthermore $\forall (x, x') \in \Re^N \times \Re^N$ define the function

$$q(x; x') \triangleq -\log f(x') + \sum_k \tilde{\pi}_k (v_k(x) - v_k(x')) \tag{14}$$

where $\tilde{\pi}_k = \frac{w_k \exp\{-v_k(x')\}}{\sum_j w_j \exp\{-v_j(x')\}}$. Then $q(x; x')$ is a surrogate function for $-\log f(x)$, and $\forall (x, x') \in \Re^N \times \Re^N$,

$$\begin{aligned} q(x'; x') &= -\log f(x') & (15) \\ q(x; x') &\geq -\log f(x) & (16) \end{aligned}$$

- Each $v_k(x)$ is quadratic, so the resulting surrogate function, $q(x; x')$, is also quadratic

Surrogate Prior for GM-MRF Model

- Original energy function

$$u(x) = \frac{1}{\eta} \sum_{s \in S} V(P_s x)$$

$$V(P_s x) = -\log\left(\sum_{k=0}^{K-1} \pi_k \frac{1}{(2\pi)^{p/2}} |B_k|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2} \| P_s x - \mu_k \|_{B_k}^2 \right\} \right)$$

- Surrogate energy function

$$u(x; x') = \frac{1}{2\eta} \sum_{s \in S} \sum_{k=0}^{K-1} \tilde{w}_k \| P_s x - \mu_k \|_{B_k}^2 + c(x')$$

where the weights are given by

$$\tilde{w}_k = \frac{\pi_k |B_k|^{1/2} \exp\left\{ -\frac{1}{2} \| P_s x' - \mu_k \|_{B_k}^2 \right\}}{\sum_{j=0}^{K-1} \pi_j |B_j|^{1/2} \exp\left\{ -\frac{1}{2} \| P_s x' - \mu_j \|_{B_j}^2 \right\}} \qquad x' : \text{ current state of } x$$

- The weights, $\tilde{w}_k$, are soft classifications into the GMM classes

Experiments

- Denoising experiments with the GM-MRF model
    1. high dosage CT images with artificially added white noise;
    2. low dosage CT images, containing real reconstruction noise.

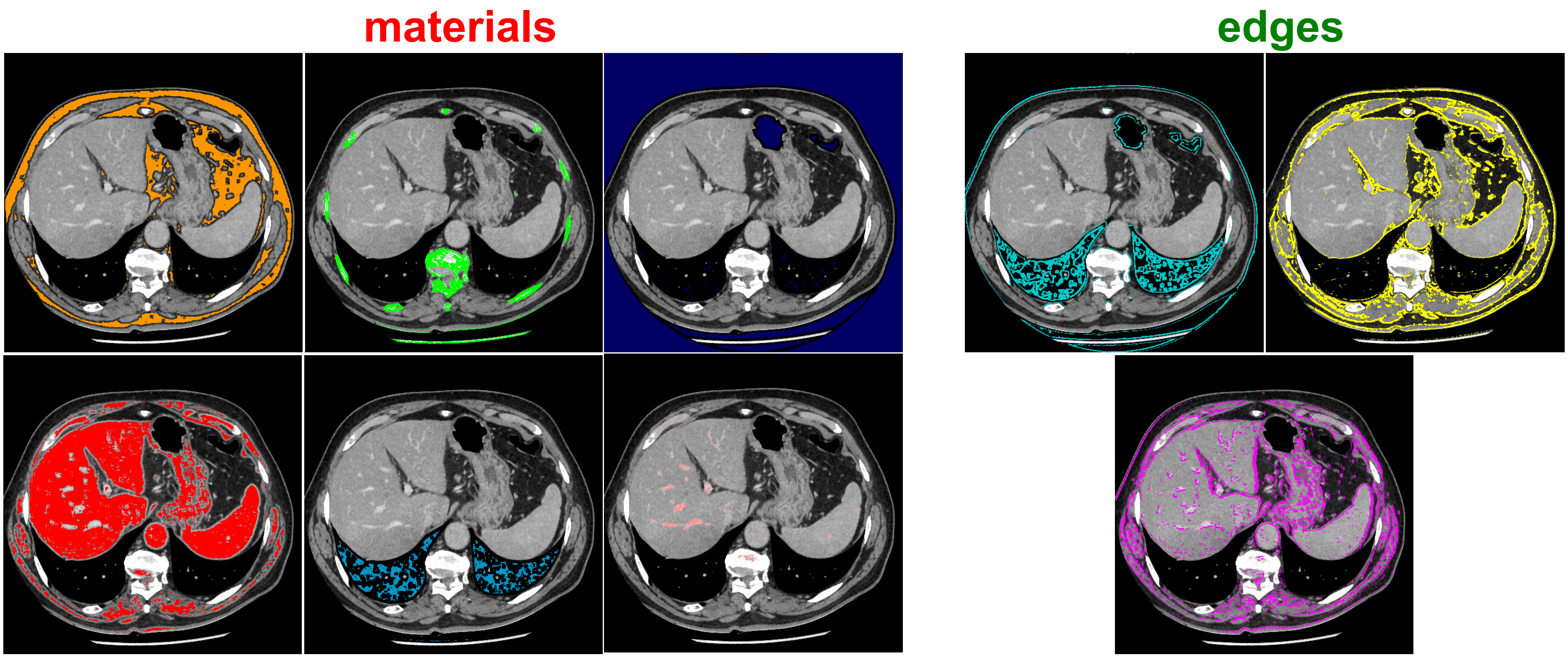
- Compared with the following methods
    - q-GGMRF model (8−point neighborhood, p=2, q=1.2, c=10)
    - K-SVD method (7x7 patch, 512 dictionary entries)
    - BM3D method (8x8 patch)

- The GM-MRF model was trained from clean high dosage CT images, with 30 subclasses and patch size 5x5.

- Parameters adjusted for lowest RMSE values (Experiment 1) and comparable noise level in homogeneous region (Experiment 2) for all methods.

MAP classification with Learned GM-MRF

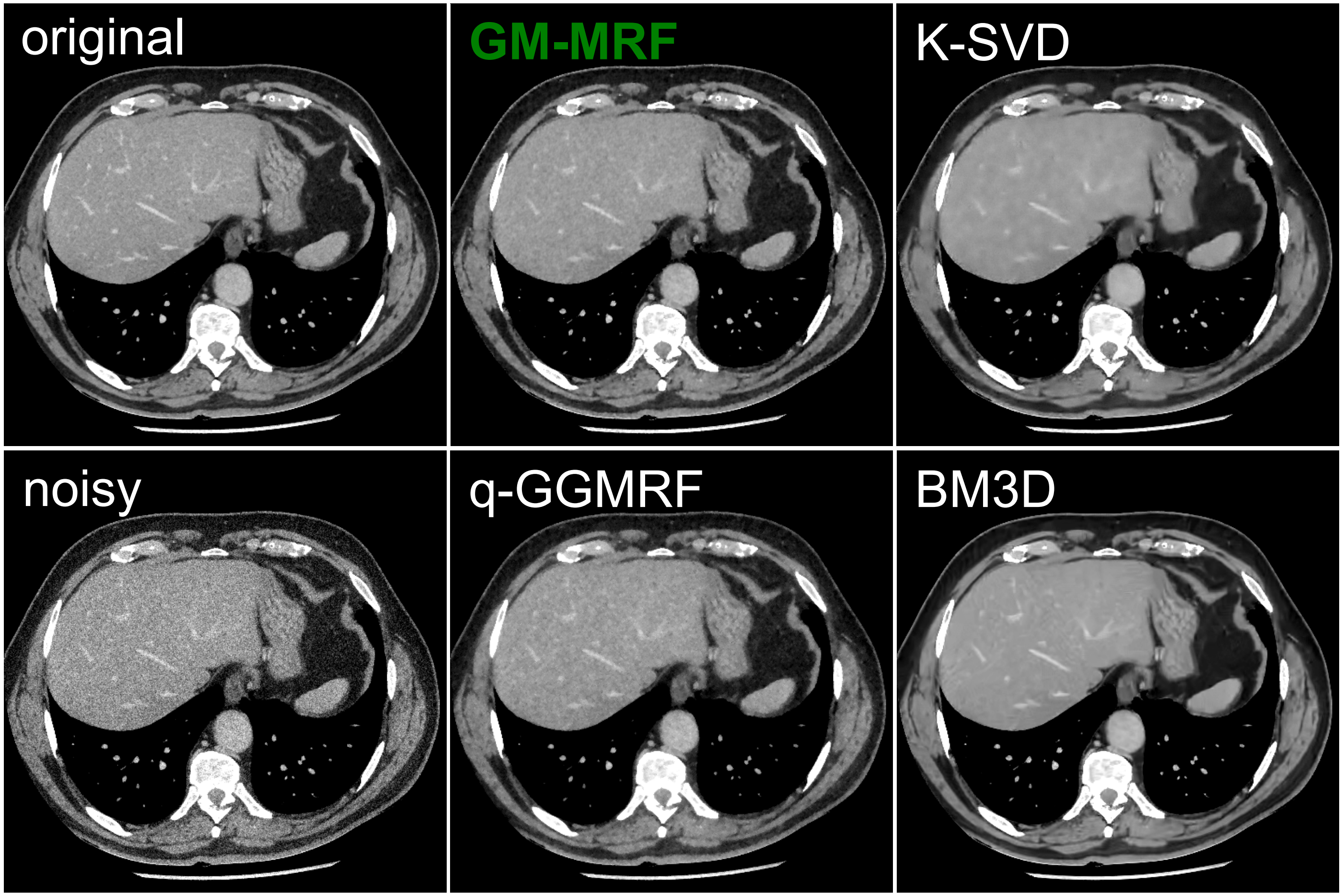
- Color-codes the most probable subclass for each patch with the learned GMM parameters
- Shows that the GMM parameters capture different materials along with different edges



Experiment 1: High Dosage CT Images

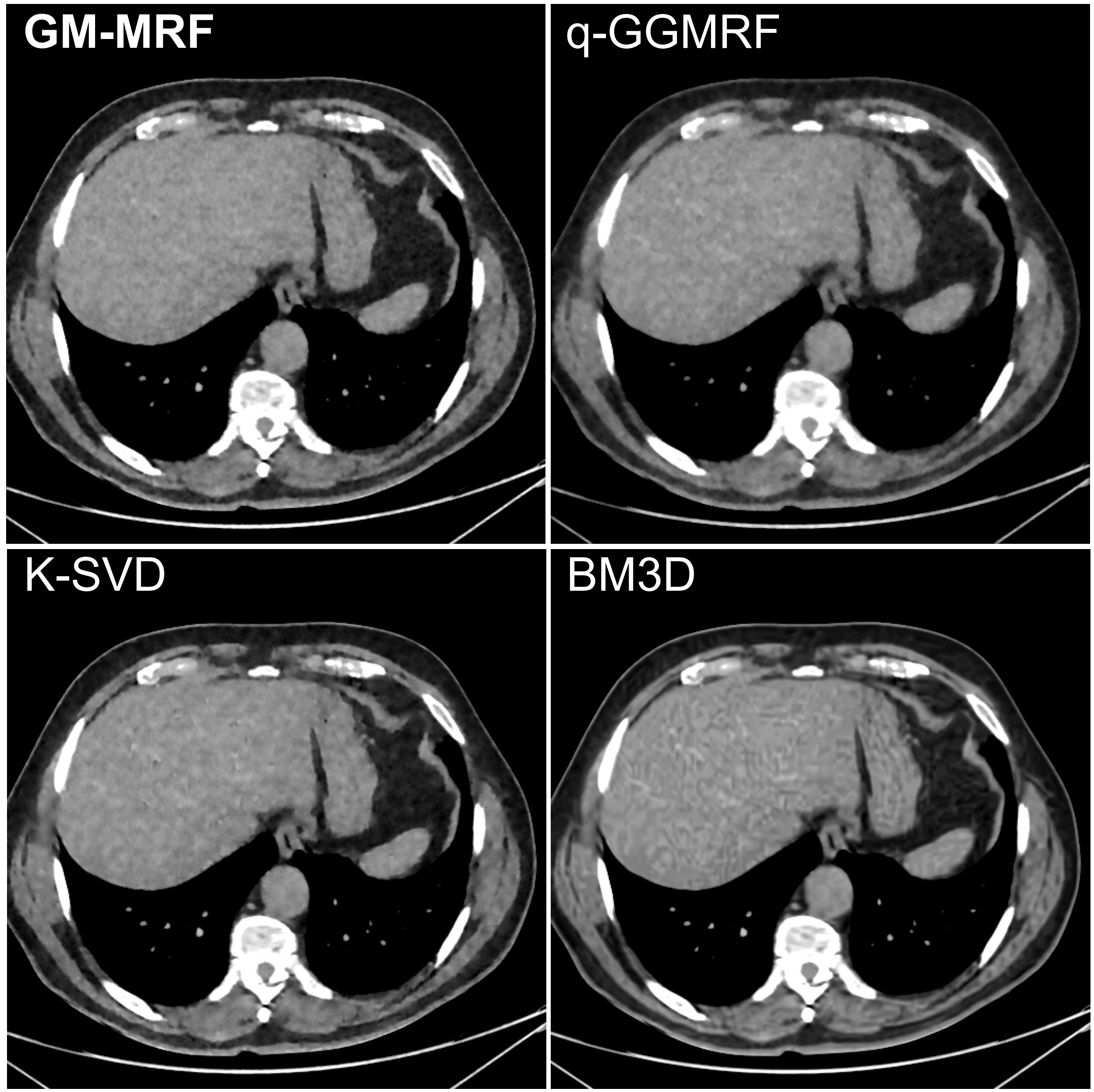
- GM-MRF model achieves
  - lowest RMSE
  - less salt/pepper noise and sharper edges than q-GGMRF model
  - less aggressive and preserves more details in soft tissues than K-SVD and BM3D

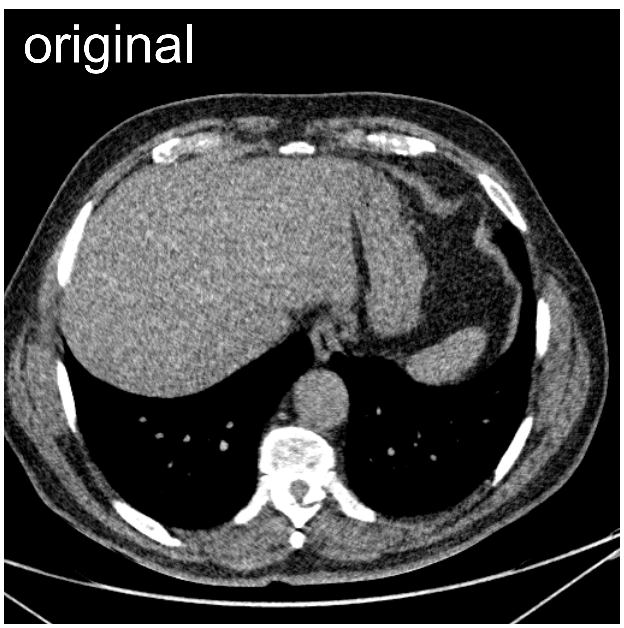
Methods	RMSE (HU)
noisy	40.05
GM-MRF	**17.02**
Q-GGMRF	20.48
K-SVD	18.68
BM3D	17.25



Experiment 2: Low Dosage CT Images

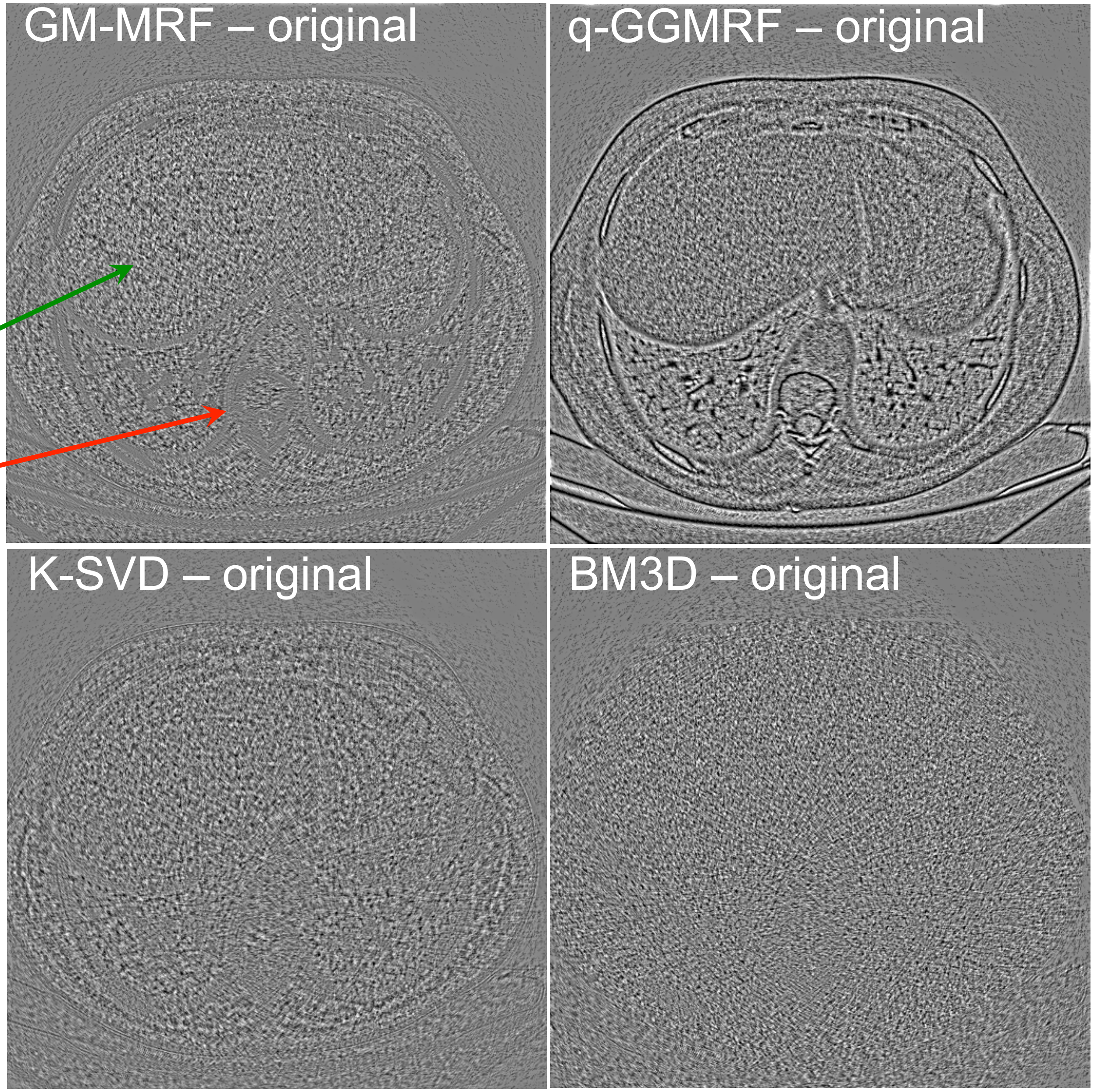
- GM-MRF achieves
  - sharper edges than q-GGMRF model
  - less artifacts and better texture in soft tissues than K-SVD and BM3D





Experiment 2: Low Dosage Difference Images

- GM-MRF model shows the ability to regularize different materials/structures differently:
  - more regularization in soft tissue
  - less regularization in bone/lung tissue



Conclusions

- GM-MRF (Gaussian Mixture MRF)
  - Is and MRF
  - Can be trained for any image
  - Captures full multivariate distribution of image
- How is the GM-MRF used?
  - Is constructed with POE trick (geometric mean of distributions)
  - Surrogate function for an mixture distribution
- Medical applications
  - It can capture both mean and texture characteristics for medical applications
  - MAP optimization looks like it uses an adaptive quadratic prior