# COLLABORATIVE COMPRESSIVE X-RAY IMAGE RECONSTRUCTION

*Jiaji Huang, Xin Yuan, and Robert Calderbank*

Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708, U.S.A

## ABSTRACT

The Poisson Factor Analysis (PFA) is applied to recover signals from a Poisson compressive sensing system. Motivated by the recently developed compressive X-ray imaging system, Coded Aperture Coherent Scatter Spectral Imaging (CACSSI) [1], we propose a new Bayesian reconstruction algorithm. The proposed Poisson-Gamma (PG) approach uses *multiple* measurements to *refine* our knowledge on both sensing matrix and background noise to overcome the uncertainties and inaccuracy of the hardware system. Therefore, a collaborative compressive X-ray image reconstruction algorithm is proposed under a Bayesian framework. Experimental results on real data show competitive performance in comparison with point estimation based methods.

***Index Terms*—** Poisson inversion, sensing matrix, compressive sensing, X-ray imaging, Bayesian inference

## 1. INTRODUCTION

Photon (energy) limited imaging systems collect Poisson distributed measurements, whose inverse problem has received lots of attention recently [2, 3, 4, 5]. In a Poisson imaging system, the measurements are

$$\mathbf{y} \sim \text{Pois}(\mathbf{Hf} + \boldsymbol{\mu}), \tag{1}$$

where $\mathbf{H} \in \mathbb{R}_+^{M\times N}$ is the sensing matrix and $\boldsymbol{\mu} \in \mathbb{R}_+^M$ is the "dark current" (background noise). A Poisson inverse problem is the task of recovering $\mathbf{f} \in \mathbb{R}_+^N$ given $\mathbf{y} \in \mathbb{Z}_+^M$, $\mathbf{H}$ and $\boldsymbol{\mu}$. While the Maximum Likelihood Estimator (MLE) [6] has been shown effective in solving Poisson inverse problems [7], incorporating prior information of the scene, *e.g.*, sparsity, structures, significantly improves the recovery performance. For example, [8] learns a data dependent dictionary and denoises images based on dictionary representation. [9, 10] assume the piece-wise smooth nature of images and impose TV semi-norm regularization. [11] further exploits the smoothness of images and proposes the Hessian-Schatten norm regularizor. In the Poisson inverse regime, those problems associated with compressive Poisson imaging system are quite new and appealing. [12] is among the very few works that give the performance bounds on compressive Poisson inverse problem. In this work, we are interested in the signal reconstruction problem based on a compressive Poisson imaging system, Coded Aperture Coherent Scatter Spectral Imaging (CACSSI) [1].

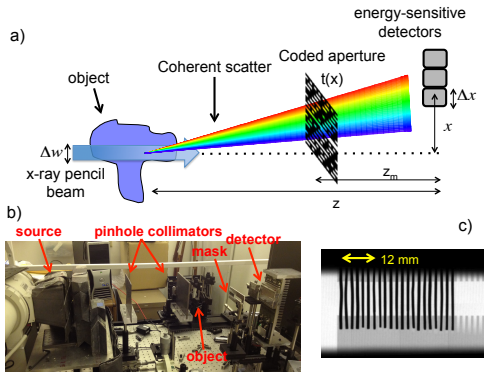



**Fig. 1**: CACSSI setup: a) Schematic and b)Photograph of the CACSSI setup c) X-ray transmission image of the coded aperture [1]

## 2. HARDWARE AND EXISTING APPROACH

An illustration of the CACSSI [1] is shown in figure 1. By placing a coded aperture in the path of the scattered X-rays, CACSSI enhances the throughput of the system without comprising the imaging performance. CACSSI allows one to measure quickly both the location and molecular signature of a target object placed in the path of a pencil beam. It is faster than systems that requires multiple times of measurements [13], and is more compact (compressive) compared with systems that uses an array of detectors [7, 14].

The imaging mechanism of CACSSI is the fact that X-ray scatters at different angles and energies after hitting different materials. This fact is described by the Bragg's law, $q = \frac{1}{2d} = \frac{E}{hc}\sin\frac{\delta}{2}$, where $q$ is the momentum transfer. $d$ is the effective lattice spacing of the material. $E$ is the X-ray energy. $h$ is the Planck constant. $c$ is the light speed and $\delta$ is the angle between the incident and scattered X-ray. Resorting to forward model, the imaging mechanism boils down to equation (1), where the sensing matrix $\mathbf{H}$ can be estimated by the forward model given CACSSI's physical parameters. The problem we are interested in is the imperfect knowledge of $\mathbf{H}$ during the spectrum recovery step. Namely, the $\mathbf{H}$ estimated

by forward model is not exactly the truly underlying sensing matrix. While our motivation and experiments are based on CACSSI, we would like to point out that this problem does not lose generality in face of other real world imaging systems.

Herein, we consider that the underlying signal $\mathbf{f}$ is a two dimensional image as in [1]. More specifically, suppose the space in the X-ray body is quantized into $Z$ spatial grids, and the material with spectral signature $\mathbf{s} \in \mathbb{R}_+^Q$ occupies a subset $\Omega \subset \{1, 2, \ldots, Z\}$ of these grids. Then the input signal induced by this material is an image $\mathbf{I} \in \mathbb{R}_+^{Q \times Z}$, whose $z$-th column is $\mathbf{s}$ if $z \in \Omega$ and zero otherwise. Denote $\mathbf{f} = \text{vec}(\mathbf{I}) \in \mathbb{R}_+^N$ as the resultant vector by stacking $\mathbf{I}$'s columns on one another. Then the measurements at the detector can be modeled by equation (1).

In [7], the authors take an optimization based approach. They assume that their knowledge of the sensing matrix, $\widehat{\mathbf{H}}$, is perfect. To estimate $\boldsymbol{\mu}$, they first measure the background by placing no material in the X-ray body, which results in measurements $\hat{\boldsymbol{\mu}} \sim \text{Pois}(\boldsymbol{\mu})$. Then they estimate $\boldsymbol{\mu}$ by the Maximum Posterior estimator with Total Variation Penalty (MAP-TV),

$$\bar{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \left( -\log \text{Pois}(\hat{\boldsymbol{\mu}}|\boldsymbol{\mu}) + \lambda \|\boldsymbol{\mu}\|_{\text{TV}} \right), \tag{2}$$

where $\text{Pois}(\hat{\boldsymbol{\mu}}|\boldsymbol{\mu})$ is understood as the Poisson probability density at $\hat{\boldsymbol{\mu}}$ given the rate parameter $\boldsymbol{\mu}$.

Substituting the $\boldsymbol{\mu}$ in equation (1) with $\bar{\boldsymbol{\mu}}$, again using MAP-TV estimator, they get the estimate of signal,

$$\mathbf{f}_{\text{MAP}} = \arg\min_{\mathbf{f}} \left( -\log \text{Pois}\left( \mathbf{y} | \widehat{\mathbf{H}}\mathbf{f} + \bar{\boldsymbol{\mu}} \right) + \tau \|\mathbf{f}\|_{\text{TV}} \right) \tag{3}$$

An MLE estimator, in contrast, sets $\tau = 0$ in equation (3).

However, there are a few problems with the aforementioned approach. First, the knowledge of the sensing matrix, $\widehat{\mathbf{H}}$, is an estimate by the forward model and in general not accurate. Second, for different materials at different locations, the environment in the X-ray body changes, and background noise varies accordingly. In fact, for the $k$-th sample put in the X-ray body, we have

$$\mathbf{y}_k \sim \text{Pois}(\mathbf{H}\mathbf{f}_k + \boldsymbol{\mu}_k), k = 1, \ldots, K \tag{4}$$

Before proceeding to our method, we define the notations here. Denote $\mathbf{F} = \left[\mathbf{f}_1, \ldots, \mathbf{f}_K\right] \in \mathbb{R}_+^{N \times K}$ as the concatenation of $K$ input signals induced by $K$ materials. Similarly, $\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_K\right] \in \mathbb{Z}_+^{M \times K}$ and $\mathbf{U} = \left[\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K\right] \in \mathbb{R}_+^{M \times K}$. A more compact form of equation (4) is

$$\mathbf{Y} \sim \text{Pois}(\mathbf{H}\mathbf{F} + \mathbf{U}) \tag{5}$$

## 3. PROPOSED APPROACH

### 3.1. Poisson-Gamma Model

Since a piece of material typically occupies a small part of the entire space, it is thus reasonable to assume that the $\mathbf{f}_k$'s are sparse. For those sparse nonnegative variables, a natural idea is to impose Gamma distribution as its prior. Gamma is the conjugate prior of Poisson, thus analytical form of the posterior is available. Also, it is easy to encourage sparseness by setting the shape parameter of Gamma no bigger than 1.

We thus impose the following model

$$\begin{aligned}
\mathbf{F}_{n,k} &\sim \text{Gamma}(\alpha_f, \beta_f) \\
\mathbf{H}_{m,n} &\sim \text{Gamma}(\beta_h \cdot \widehat{\mathbf{H}}_{m,n},\ \beta_h) \\
\mathbf{U}_{m,k} &\sim \text{Gamma}(\beta_\mu \cdot \hat{\boldsymbol{\mu}}_m,\ \beta_\mu) \\
\mathbf{Y}_{m,k} &\sim \text{Pois}\left( \sum_{n=1}^{N} \mathbf{H}_{m,n}\mathbf{F}_{n,k} + \mathbf{U}_{m,k} \right)
\end{aligned} \tag{6}$$

Notice that according to the second and third equation of the above model, the mean of the priors on $\mathbf{H}$ and $\boldsymbol{\mu}_k$ equals $\widehat{\mathbf{H}}$ and $\hat{\boldsymbol{\mu}}$. Thus, pre-computed sensing matrix and pre-measured background are utilized.

### 3.2. Inference

The inference follows the Poisson Factor Analysis (PFA) [15]. To start, it is necessary to introduce the latent variables that composes $\mathbf{Y}_{m,k}$.

1. Latent variables $\mathbf{Y}_{m,k}^{(n)}$

Define

$$\begin{aligned}
\mathbf{Y}_{m,k}^{(n)} &\sim \text{Pois}(\mathbf{H}_{m,n}\mathbf{F}_{n,k}),\ \ n = 1, \ldots, N \\
\mathbf{Y}_{m,k}^{(N+1)} &\sim \text{Pois}(\mathbf{U}_{m,k})
\end{aligned} \tag{7}$$

Assuming the independency among the $\mathbf{Y}_{m,k}^{(n)}$'s, and due to the additivity of independent poisson variables, we have

$$\mathbf{Y}_{m,k} = \sum_{n=1}^{N+1} \mathbf{Y}_{m,k}^{(n)} \tag{8}$$

To estimate $\mathbf{F}$, it is necessary to sample each latent variable $\mathbf{Y}_{m,k}^{(n)}$. The following lemma states that given $\mathbf{Y}_{m,k}$, $\mathbf{Y}_{m,k}^{(n)}$ can be sampled from a multinomial distribution.

**Lemma 1.** *[15] Let independent variables $x_1, \ldots, x_n$ be drawn from Poisson distribution with rate parameters $\lambda_i, i = 1, \ldots, n$ and $x = \sum_{i=1}^{n} x_i$. Then the conditional distribution of $(x_1, \ldots, x_n)$ given $x$ is*

$$(x_1, \ldots, x_n)|x \sim \textit{Mult}\left( x; \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}, \ldots, \frac{\lambda_n}{\sum_{i=1}^n \lambda_i} \right) \tag{9}$$

Here we omit the proof, which is straightforward by applying the Bayesian rule.

Applying Lemma 1, we can sample the latent variables by

$$(\mathbf{Y}_{m,k}^{(1)},\ldots,\mathbf{Y}_{m,k}^{(N)},\mathbf{Y}_{m,k}^{(N+1)})|\mathbf{Y}_{m,k}\sim \text{Mult}\left(\mathbf{Y}_{m,k};\xi_{m,k}^{(1)},\ldots,\xi_{m,k}^{(N)},\xi_{m,k}^{(N+1)}\right)$$

where

$$\xi_{m,k}^{(n)}=\frac{\mathbf{H}_{m,n}\mathbf{F}_{n,k}}{\sum_{n=1}^{N}\mathbf{H}_{m,n}\mathbf{F}_{n,k}+\mathbf{U}_{m,k}},\quad n=1,\ldots,N$$

and

$$\xi_{m,k}^{(N+1)}=\frac{\mathbf{U}_{m,k}}{\sum_{n=1}^{N}\mathbf{H}_{m,n}\mathbf{F}_{n,k}+\mathbf{U}_{m,k}} \tag{10}$$

2. Infer $\mathbf{F}_{n,k}$

Obtaining the $\mathbf{Y}_{m,k}^{(n)}$'s, we are able to derive a closed form posterior of $\mathbf{F}_{n,k}, n=1,\ldots,N$.

$$\begin{aligned} p(\mathbf{F}_{n,k}|-) &\propto \prod_{m=1}^{M}\text{Pois}(\mathbf{Y}_{m,k}^{(n)}|\mathbf{H}_{m,n}\mathbf{F}_{n,k}) \\ &\quad\times\text{Gamma}(\mathbf{F}_{n,k}|\alpha_f,\beta_f) \\ &= \text{Gamma}\left(\alpha_f+\sum_{m=1}^{M}\mathbf{Y}_{m,k}^{(n)},\beta_f+\sum_{m=1}^{M}\mathbf{H}_{m,n}\right) \end{aligned} \tag{11}$$

2. Infer $\mathbf{U}_{m,k}$

Similar to equation (11), the posterior of $\mathbf{U}_{m,k}$ is

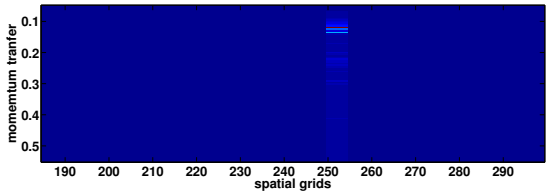
$$\begin{aligned} p(\mathbf{U}_{m,k}|-) &\propto \text{Pois}\left(\mathbf{Y}_{m,k}^{(N+1)}|\mathbf{U}_{m,k}\right) \\ &\quad\times\text{Gamma}(\mathbf{U}_{m,k}|\beta_\mu\cdot\hat{\boldsymbol{\mu}}_m,\ \beta_\mu) \\ &= \text{Gamma}\left(\beta_\mu\cdot\hat{\boldsymbol{\mu}}_m+\mathbf{Y}_{m,k}^{(N+1)},\beta_\mu+1\right) \end{aligned} \tag{12}$$

3. Infer $\mathbf{H}_{m,n}$

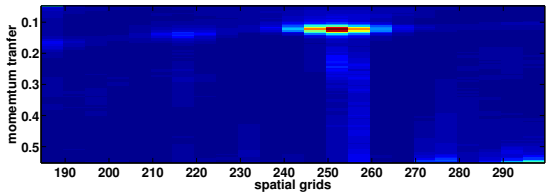
The posterior for $\mathbf{H}_{m,n}$ can be derived as

$$\begin{aligned} p(\mathbf{H}_{m,n}|-) &= \prod_{k=1}^{K}\text{Pois}\left(\mathbf{Y}_{m,k}^{(n)}|\mathbf{H}_{m,n}\mathbf{F}_{n,k}\right)\times \\ &\quad\text{Gamma}(\mathbf{H}_{m,n}|\beta_h\cdot\widehat{\mathbf{H}}_{m,n},\ \beta_h) \\ &= \text{Gamma}\left(\beta_h\cdot\widehat{\mathbf{H}}_{m,n}+\sum_{k=1}^{K}\mathbf{Y}_{m,k}^{(n)},\right. \\ &\quad\left.\beta_h+\sum_{k=1}^{K}\mathbf{F}_{n,k}\right) \end{aligned} \tag{13}$$

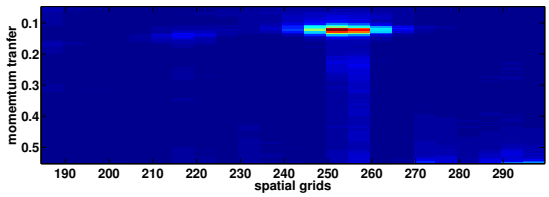
Closed-from posterior distributions in (10)-(13) provide fast MCMC (Markov Chain Monte Carlo) inference and the variational Bayesian (VB) [16] inference. For the running time consideration, we adopt the mean-field VB inference here. A VB approach attempts to approximate the posterior distribution by a simpler distribution, $p(\boldsymbol{\Theta}|\mathbf{Y})\approx q(\boldsymbol{\Theta})$, where $\mathbf{Y}$ is the observed data matrix and $\boldsymbol{\Theta}$ is the set of independent latent variables in the model [16]. VB assumes a complete factorization across latent variables, $q(\boldsymbol{\Theta})=\prod_i q_i(\Theta_i)$. We define $\boldsymbol{\Theta}=\{\mathbf{F},\mathbf{H},\mathbf{U}\}$ for the purpose of this work. Solving



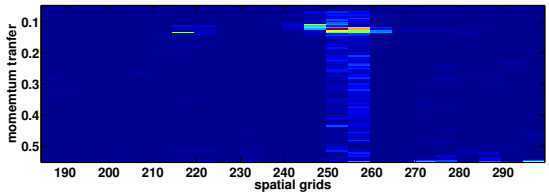

(a) HDPE input image (reference)



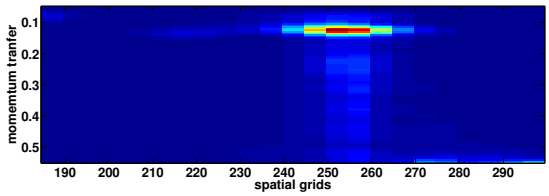

(b) by PG-updateH




(c) by PG-fixH




(d) by MLE




(e) by MAP-TV

**Fig. 2**: Estimated input image

for the optimal distribution $q^\star(\boldsymbol{\Theta})$ that minimizes the distance between $p$ and $q$ effectively estimates the conditional posterior distribution $p(\boldsymbol{\Theta}|\mathbf{Y})$. Commonly used distance metric between the two distribution function is the Kullback-Leibler (KL) divergence [17]. We write the KL-divergence of $p$ from $q$ as follows:

$$\text{KL}(q\|p)=\int_{\boldsymbol{\Theta}}q(\boldsymbol{\Theta})\ln\frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta}|\mathbf{Y})}d\boldsymbol{\Theta}, \tag{14}$$

which can be simplified to

$$\begin{aligned} \ln p(\mathbf{Y}) &= \text{KL}(q\|p)+\mathcal{L}(q), \\ \mathcal{L}(q) &= -\int_{\boldsymbol{\Theta}}q(\boldsymbol{\Theta})\ln\frac{q(\boldsymbol{\Theta})}{p(\boldsymbol{\Theta},\mathbf{Y})}d\boldsymbol{\Theta}. \end{aligned} \tag{15}$$

Now, it is easy to notice that $\ln p(\mathbf{Y})$ is fixed with respect to the variations in $q(\mathbf{\Theta})$. Therefore, maximizing the Evidence Lower Bound (ELBO), $\mathcal{L}(q)$, is equivalent to minimizing the KL-divergence between the two distributions. This minimal distance occurs when

$$\ln q^{\star}(\mathbf{\Theta}) = \mathbb{E}[\ln p(\mathbf{Y}, \mathbf{\Theta})] + \text{const.} \tag{16}$$

Assuming a complete factorization cross the latent variables, $q(\mathbf{\Theta}) = \prod_i q_i(\Theta_i)$, each parameter in a variational Bayes model is independently updated according to

$$q_j^{\star}(\Theta_j) \propto \exp\{\mathbb{E}_{i \neq j}[\ln p(\mathbf{Y}, \mathbf{\Theta})]\}. \tag{17}$$

By applying the update rule in [16], equation (17) associated with (10)-(13) gives the inference in Algorithm 1.

**Algorithm 1** Poisson-Gamma Inference with an updated $\mathbf{H}$

**Input:** $\widehat{\mathbf{H}}$, background measurement $\hat{\boldsymbol{\mu}}$, measurements $\mathbf{Y}$
**Output:** Signal estimates $\mathbf{F}$, background estimates $\mathbf{U}$ and updated $\mathbf{H}$

1: **for** loop=1,. . . ,MAX_LOOPS **do**
2: Compute $\xi_{m,k}^{(n)}$ as in equation (10)
3: Estimate latent variables

$$\mathbf{Y}_{m,k}^{(n)} \leftarrow \mathbf{Y}_{m,k} \cdot \xi_{m,k}^{(n)} \tag{18}$$

4: Estimate $\mathbf{F}_{n,k}$

$$\mathbf{F}_{n,k} \leftarrow \left(\alpha_f + \sum_{m=1}^{M} \mathbf{Y}_{m,k}^{(n)}\right) / \left(\beta_f + \sum_{m=1}^{M} \mathbf{H}_{m,n}\right) \tag{19}$$

5: Estimate $\mathbf{U}_{m,k}$

$$\mathbf{U}_{m,k} \leftarrow \left(\beta_\mu \cdot \hat{\boldsymbol{\mu}}_m + \mathbf{Y}_{m,k}^{(N+1)}\right) / (\beta_\mu + 1) \tag{20}$$

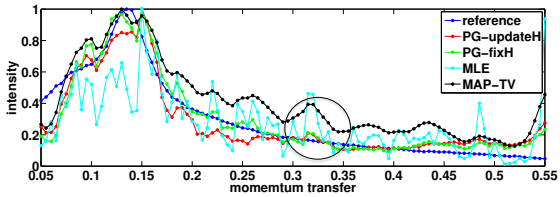
6: Estimate $\mathbf{H}_{m,n}$

$$\mathbf{H}_{m,n} \leftarrow \left(\beta_h \cdot \widehat{\mathbf{H}}_{m,n} + \sum_{k=1}^{K} \mathbf{Y}_{m,k}^{(n)}\right) / \left(\beta_h + \sum_{k=1}^{K} \mathbf{F}_{n,k}\right) \tag{21}$$

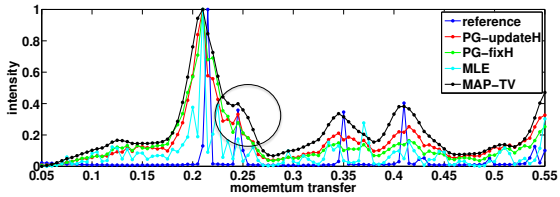
7: **end for**

## 4. EXPERIMENTS

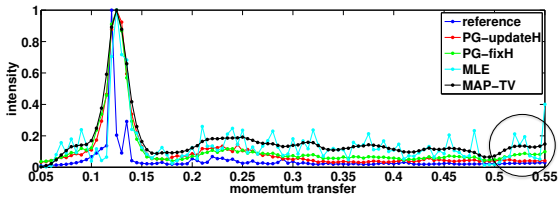
We test our algorithm on the 25 measurements (*i.e.*, $K = 25$) collected by CACSSI. The comparator algorithms are MLE and MAP-TV, both of them have been shown effective for spectrum recovery from CACSSI measurements [7, 1]. We also demonstrate the results with fixed $\mathbf{H}$, *i.e.*, removing step 6 in algorithm 1. Due to the limited space, we demonstrate several representative materials. The parameter settings are $\alpha_f = 1$, $\beta_f = 0.1$, $\beta_\mu = 1$, $\beta_h = 1$ and MAX_LOOPS$= 20$. For MAP-TV, the TV regularizer has a values of 3000.



(a) Methanol spectrum



(b) Al spectrum



(c) HDPE spectrum

**Fig. 3**: Estimated spectrums

Remember that the input signal $\mathbf{f}$ is an image. We demonstrate the image induced by a piece of teflon put at grid 252. An ideal recovery would yield an image that only has one non-zero column, corresponding to the 252-th spatial grid, as shown in 2a. Figure 2 (b) through (e) are the recovered images by all methods. All of them are able to correctly locate the material. Then we extract the strongest column in these images and normalize it to compare with the reference spectrum. Figure 3 shows the extracted spectrums for three representative materials. MLE yields very jagged estimate of the spectrums whereas MAP-TV tends to over-smooth the peaks (see the highlighted region of figure 3b). In all cases, the two Poisson-Gamma model based approaches are very competitive. And with refined sensing matrix, the estimated spectrum is even closer to the reference.

## 5. CONCLUSION

We propose a collaborative compressive X-ray image reconstruction algorithm based on the Poisson-Gamma Bayesian framework. By refining the sensing matrix and the background noise through joint inversion of multiple measurements, the uncertainties of noise and sensing matrix introduced by the hardware are weakened. Experimental results on real data demonstrate the superior performance of the proposed approach.

## 6. REFERENCES


[1] J. A. Greenberg, K. Krishnamurthy, and D. J. Brady, "Snapshot molecular imaging using coded energy- sensitive detection," *Optics Express*, vol. 21, no. 21, pp. 25480–25491, October 2013.

[2] A. Antoniadis and J. Bigot, "Poisson inverse problems," *The Annals of Statistics*, vol. 34, no. 5, pp. 2132–2158, 2006.

[3] K. Krishnamurthy, M. Raginsky, and R. Willett, "Multiscale photon-limited spectral image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 619–645, September 2010.

[4] Z. T. Harmany, F. M. Roummel, and R. Willett, "This is spiral-tap: sparse poisson intensity reconstruction algorithms—theory and practice," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1084–1096, 2012.

[5] J. Salmon, C. A. Deledalle, R. Willett, and Z. T. Harmany, "Poisson noise reduction with non-local pca," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1109–1112.

[6] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JOSA*, vol. 62, no. 1, pp. 55–59, 1972.

[7] K. MacCabe, K. Krishnamurthy, A. Chawla, D. Marks, E. Samei, and D. Brady, "Pencil beam coded aperture x-ray scatter imaging," *Optics Express*, vol. 20, no. 15, pp. 16310–16320, 2012.

[8] L. Ma, L. Moisan, J. Yu, and T. Zeng, "A dictionary learning approach for poisson image deblurring," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, 2013.

[9] L.Rudin, S.Osher, and E.Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[10] P. Getreuer, "Rudin-osher-fatemi total variation denoising using split bregman," *Image Processing On Line*, vol. 10, 2012.

[11] S. Lefkimmiatis and M. Unser, "Poisson image reconstruction with hessian schatten-norm regularization," *IEEE transactions on image processing*, vol. 22, no. 11, 2013.

[12] M. Raginsky, R. Willett, Z. T. Harmany, and R. F. Marcia, "Compressed sensing performance bounds under poisson noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 3990–4002, 2010.

[13] A. Dicken, K. Rogers, P. Evans, J. W. Chan, J. Rogers, and S. Godber, "Combined x-ray diffraction and kinetic depth effect imaging," *Optics Express*, vol. 19, no. 7, pp. 6406–6413, 2011.

[14] K. P. MacCabe, A. D. Holmgren, M. P. Tornai, and D. J. Brady, "Snapshot 2d tomogra- phy via coded aperture x-ray scatter imaging," *Applied Optics*, vol. 52, no. 19, pp. 4582–4589, 2013.

[15] M. Zhou, L. Hannah, D. Dunson, and L. Carin, "Beta-negative binomial process and poisson factor analysis," in *International Conference on Artificial Intelligence and Statistics*, 2012.

[16] M. J. Beal, *Variational algorithms for approximate Bayesian inference*, Ph.D. thesis, University College London, 2003.

[17] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.