Poisson Observed Image Restoration using a Latent Variational Approximation with Gaussian MRF

SHOUNO HAYARU[1,a)] OKADA MASATO[2,,b)]

Abstract: We treat an image restoration problem throughout a Poisson noise channel. The Poisson randomness might be appeared in observation of low contrast object, and its variable takes discrete and positive value. The Poisson noise observation is often hard to treat in a theoretical analysis. In our formulation, we interpret the Poisson noise channel observation as a Bernoulli process, and apply a latent variable method to transform the observation as a Gaussian process with single latent variable. We formulate the image restoration problem as a Bayesian approach, and introduce a Gaussian Markov random field as its prior. The latent parameters and Poisson parameters are treated as hyper-parameters, and we infer them in the expectation maximization framework.

1. Introduction

The techniques of the noise reducing, which is called image restoration in the field of digital image processing, is an important in the meaning of the pre-processing.

From the theoretical view of Bayesian image restoration, additive white Gaussian noise (AWGN) was mainly discussed as the image corrupting process. However, in the real world, the noise corruption process could not be described as such Gaussian process. For example, night photograph must treat low contrast object observation as a Poisson process. In this study, we treat image restoration with the Poisson corruption process in the manner of the Bayesian approach.

Assuming Gaussian Markov random field as a prior of Bayes inference, Poisson corruption process makes difficult to derive posterior probability in analytic form, since the Poisson variable take discrete and non-negative value. Thus, we introduce a latent variational approximation in the inference derivation [1][2][3]. In this study, we transform the Poisson corruption process as the corresponding Bernoulli process, and introduce the latent variable to approximate the Poisson process as the Gaussian process[3]. Once, we get the corresponding Gaussian process, we can infer the posterior probability easily. In this formulation, we introduce several latent parameters, so that, we should infer them. In order to solve the problem, we introduce a expectation maximization (EM) algorithm as a inference engine.

2. Formulation

The digital image is defined by the 2-dimensional array of pixels, and each pixel has some value. Considering Poisson noise corruption means the pixel have some parameter, and the observation is obtained by stochastic process under the parameter. Let us consider to assign the parameter $\rho\Delta$ and derive the Poisson random variable z can be denoted as

$$p(z \mid \rho\Delta) = \frac{(\rho\Delta)^z}{z!} \exp(-\rho\Delta),$$

which means the number of observed photons dropping into the pixel. The Poisson noise corruption process appears in the low contrast object observation such like night photograph, and some kind of computed tomography such like positron emission tomography (PET). Assuming large parameter of $\rho\Delta$, this corruption process can be approximated by the additive white Gaussian noise (AWGN) corruption process, whose average and variance are both $\rho\Delta$, that is $\mathcal{N}(z \mid \rho\Delta, \rho\Delta)$ where $\mathcal{N}\left(x \mid m, \sigma^2\right)$ means x is generated by the normal distribution whose mean and standard deviation are m and σ respectively.

However, in the the small $\rho\Delta$ area, the approximation is not good enough since the negative observation value z is sometimes appeared. Watanabe *et al.* treat the Poisson corruption process of firing neuron as a Bernoulli process, which counts the number of on-off event in the proper time bins [3]. In the manner with the Watanabe's method, we should consider the dividing of the pixel area with miniregions. Fig.1 shows the configuration of the dividing system. The thick large rectangle shows the pixel area size. On the other hand, the small rectangles show the miniregions which have $\Delta_x\Delta_y = \Delta$ area sizes, so that one pixel include several miniregions. We assume each miniregion has only information of on-off event which means a photon drop into the miniregion or not. We consider the image have M pixels and denote the pixel index as m. Each pixel has K miniregions whose index is denoted as k. And (m, k)-th miniregion have an event value $\zeta_{mk} = \pm 1$. The $\zeta_{mk} = +1$ means the photon count is on, and $\zeta_{mk} = -1$ is off. Assuming the on-event probability in the pixel m is uniform as $\rho_m\Delta$, we can define the observation probability as

1 Graduate School of Informatics and Engineering, University of Electro-Communications, Chofugaoka 1-5-1, Chofu, 182–8585, Japan
2 Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, 277–8561, Japan
†1 Presently with RIKEN BSI, 2-1 Hirosawa, Wako, 351-0198, Japan
a) shouno@uec.ac.jp
b) okada@k.u-tokyo.ac.jp

$$p(\zeta_{mk} \mid \rho_m) = (\rho_m\Delta)^{\frac{1+\zeta_{mk}}{2}}(1-\rho_m\Delta)^{\frac{1-\zeta_{mk}}{2}}. \quad (1)$$

In the observation, each pixel value z_m can be defined as the sum of the whole on-events in the pixel:

$$p(z_m \mid \{\zeta_{mk}\}) = \delta\left(z_m - \sum_k \frac{\zeta_{mk}+1}{2}\right), \quad (2)$$

where $\delta(\cdot)$ denotes Kronecker's function. Our interest in the causality of the parameter ρ_m for the observed value z_m. Hence, we derive the probability by marginalization:

$$p(z_m \mid \rho_m) = \sum_{\{\zeta_{mk}\}} p(z_m \mid \{\zeta_{mk}\}) \prod_{k=1}^{K} p(\zeta_d mk \mid \rho_m\Delta) \quad (3)$$

$$= \binom{K}{z_m}(\rho_m\Delta)^{z_m}(1-\rho_m\Delta)^{K-z_m}. \quad (4)$$

In this formulation, we can confirm the eq.(4) converges to the Poisson process,

$$p(z_m \mid \rho_m) = \frac{(\rho_m\Delta)^{z_m}}{z_m!}\exp(-\rho_m\Delta), \quad (5)$$

in the limit of $K \to \infty$ with keeping $\rho_m\Delta \ll 1$.

2.1 Image Observation process

In the following analysis, the non-negative parameter ρ_m is not enough tractable, so that we apply the logit transform in the manner of the Watanabe *et al.*[3]. The logit transform from ρ_m to x_m is denoted as

$$x_m = \frac{1}{2}\ln\frac{\rho_m\Delta}{1-\rho_m\Delta}, \quad (6)$$

and we can also denote the inverted transform from x_m to ρ_m as

$$\rho_m\Delta = \frac{e^{x_m}}{2\cosh(x_m)}. \quad (7)$$

Thus, we can rewrite the Poisson noise corruption $p(z_m \mid \rho_m) = \frac{(\rho_m\Delta)^{z_m}}{z_m!}e^{-\rho_m\Delta}$ as the conditional probability

$$p(z_m|x_m) = \sum_{\{\zeta_{mk}\}} p(z_m|\{\zeta_{mk}\}) \prod_k p(\zeta_{mk} \mid x_m) \quad (8)$$

$$= \binom{K}{z_m}\exp((2z_m - K)x_m - K\ln 2\cosh x_m). \quad (9)$$

Thus, the image restoration problem can be interpreted as the inference from the observation z_m to the parameter x_m.

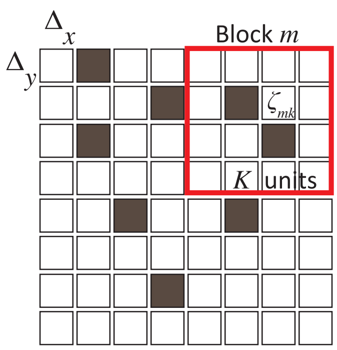

Fig. 1: Schematic diagram of the relationship between Poisson and Bernoulli observation. The large rectangles shows the pixel block and small indicates the minimum event unit which can describe only two-state, that is on-off events.

2.2 Prior probability

Introducing the Bayesian inference requires several prior probability for the image. In this study, we assume some kinds of Gaussian Markov random field (GMRF). GMRF, which is defined in collection of the neighborhoods pixel value pairs, can be denoted as the multidimensional Gaussian distribution, so that it can be applied in the theoretic analysis[4]. Usually, GMRF is defined by the sum of neighborhood differential square $\sum \|\rho_m - \rho_n\|^2$ where ρ_m and ρ_n are neighborhood pixel values. Thus, the only difference of these values $\rho_m - \rho_n$ are effective, but the absolute values are not effective in the GMRF. In this study, however, we define the parameter of the prior as $\{x_m\}$, which are logit-transformed values for $\{\rho_m\}$. Hence the absolute value of x_m are effective when we define the GMRF with the difference $x_m - x_n$ where x_m and x_n are neighbor parameter values. Thus we introduce a compensate value u_{mn} for each neighborhood. We define the energy function of the prior as

$$H_{\mathrm{pri}}(\boldsymbol{x}) = \frac{1}{2}\sum_{(m,n)}((x_m - x_n) - u_{mn})^2 \quad (10)$$

$$p(\boldsymbol{x}) = \exp\left(-\alpha H_{\mathrm{pri}}(\boldsymbol{x})\right), \quad (11)$$

where (m, n) means the neighborhood pixel indices. The prior corresponding to the energy function can denote as

$$p(\boldsymbol{x} \mid \alpha, \boldsymbol{\mu}) \propto \exp\left(-\frac{\alpha}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\Lambda(\boldsymbol{x}-\boldsymbol{\mu})\right), \quad (12)$$

, where Λ is a correlation matrix, since eq.(11) is a quadratic form of $\{x_m\}$.

Considering the prior eq.(12), the inference sometimes unstable since the determinant of the accuracy matrix $\alpha Lambda$ is 0. So that, we introduce some diagonal matrix hI where $h > 0$ and I means the unit matrix for the accuracy matrix. Thus, we rewrite the prior distribution as

$$p(\boldsymbol{x} \mid \alpha, h, \boldsymbol{\mu}) = \mathcal{N}\left(\boldsymbol{x} \mid \boldsymbol{\mu}, (\alpha\Lambda + hI)^{-1}\right), \quad (13)$$

and add h as inference hyper-parameter as well as α and $\boldsymbol{\mu}$

2.3 Image restoration with Latent variable approximation

Eq.(9) shows an exponential expression and it can be as Gauss distribution family when we can approximate the argument as the quadratic form. In this study, we introduce a latent variable approximation [2][3]. Palmer *et al.* proposed the super-Gaussian distribution can be approximated as multiplied form of the Gaussian distribution and concave parameter function[2], that is, any distribution function, which denote as $p(u) = \exp(-g(u^2))$ where $g(\cdot)$ is a concave, can be described as

$$p(u) = \exp(-g(u^2)) \quad (14)$$

$$= \sup_{\eta>0} \varphi(\eta)\,\mathcal{N}\left(u \mid 0, \eta^{-1}\right), \quad (15)$$

$$\varphi(\eta) = \sqrt{\frac{2\pi}{\eta}}\exp\left(g^*\left(\frac{\eta}{2}\right)\right). \quad (16)$$

The function pair $g(u)$ and $g^*(\eta)$ is a convex conjugate relationship which is derived from Legendre's transform

$$g(u) = \inf_{\eta>0} \eta u - g^*(\eta), \quad (17)$$

$$g^*(\eta) = \inf_{u>0} \eta u - g(u). \quad (18)$$

In the eq(15), the stochastic value u is included in the Gaussian distribution part, and non-Gaussian part is driven into the $\varphi(\eta)$ with latent-parameter η. Thus, ignoring the $\sup_{\eta>0}$ operator, we can treat eq.(15) as the Gaussian form. Using this approximation form, we derive the observation process defined by eq.(9) as the Gaussian form with latent-parameter. In the eq.(9), the untractable term is $\ln 2\cosh(\cdot)$. When we introduce the latent parameter form, we obtain the upper limit:

$$\ln 2\cosh x \leq \frac{\tanh\xi}{2\xi}(x^2 - \xi^2) + \ln 2\cosh\xi. \quad (19)$$

Thus, we introduce it into the eq.(9), we obtain

$$p_{\xi}(z \mid \boldsymbol{x}) = \prod_m \binom{K}{z_m} \exp\left(-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\Xi\boldsymbol{x} + z^{\mathrm{T}}\boldsymbol{x}\right) \exp\left(\frac{1}{2}\boldsymbol{\xi}^{\mathrm{T}}\Xi\boldsymbol{\xi} - K\sum_m \ln 2\cosh\xi_m\right), \quad (20)$$

where z means observation vector

$$z = (2z_1 - K, \cdots, 2z_m - K, \cdots, 2z_M - K)^{\mathrm{T}}, \quad (21)$$

$\boldsymbol{\xi}$ means the collection of latent parameter $\{\xi_m\}$, and matrix Ξ means a diagonal matrix whose components are $\{\frac{\tanh\xi_m}{\xi_m}\}$.

From the observation (9) and the prior (12), we can derive posterior as

$$p_{\xi}(\boldsymbol{x} \mid z, \alpha, \boldsymbol{\mu}) \propto p_{\xi}(z \mid \boldsymbol{x})\, p(\boldsymbol{x} \mid \alpha, h, \boldsymbol{\mu}), \quad (22)$$

and the observation can be approximated by the latent-valued form:

$$p_{\xi}(\boldsymbol{x} \mid z, \alpha, \boldsymbol{\mu}) \sim \mathcal{N}\left(\boldsymbol{x} \mid \boldsymbol{m}, (\Xi + \alpha\Lambda + hI)^{-1}\right), \quad (23)$$

$$\boldsymbol{m} = (\Xi + \alpha\Lambda + hI)^{-1}(z + (\alpha\Lambda + hI)\boldsymbol{\mu}). \quad (24)$$

Considering the inference parameter of $\boldsymbol{x}$ as the posterior mean of the $\boldsymbol{x}$, that is $\hat{\boldsymbol{x}} = \langle \boldsymbol{x} \rangle$, we can obtain the inference parameter explicitly:

$$\langle \boldsymbol{x} \rangle = \sum_{\boldsymbol{x}} \boldsymbol{x}\, p_{\xi}(\boldsymbol{x} \mid z, \alpha, \boldsymbol{\mu}) = \boldsymbol{m}. \quad (25)$$

However, in this form, the variable parameter $\boldsymbol{\xi}$ and the hyper-parameters α and h is undefined, so that, we introduce expectation-maximization (EM) algorithm to infer these parameters. For convenient in the following we derive these inference parameter as $\theta = \{\alpha, h, \boldsymbol{\mu}, \boldsymbol{\xi}\}$. The marginal log-likelihood of $p(z, \boldsymbol{x}; \mid \theta)$ is extracted as

$$\ln p(z \mid \theta) = \ln \sum_{\boldsymbol{x}} p(\boldsymbol{x}, z \mid \theta) \geq Q(\theta \mid \theta^{(t)}) + S[p(\boldsymbol{x} \mid z, \theta^{(t)})], \quad (26)$$

where $Q(\theta \mid \theta^{(t)})$ and entropy function $S[p(\boldsymbol{x} \mid z, \theta^{(t)})]$ can be described as

$$Q(\theta \mid \theta^{(t)}) = \sum_{\boldsymbol{x}} p(\boldsymbol{x} \mid z, \theta^{(t)}) \ln p(\boldsymbol{x}, z \mid \theta), \quad (27)$$

$$S[p(\boldsymbol{x} \mid z, \theta^{(t)})] = -\sum_{\boldsymbol{x}} p(\boldsymbol{x} \mid z, \theta^{(t)}) \ln p(\boldsymbol{x} \mid z, \theta^{(t)}). \quad (28)$$

In the marginal log-likelihood (26), the inference parameter is denoted as θ under the some fixed parameter $\theta^{(t)}$. The entropy function $S[p(\boldsymbol{x} \mid z, \theta^{(t)})]$ does not include the inference parameter θ, so that we should find the theta which maximize the function $Q(\theta|\theta^{(t)})$. Then, we can consider the iteration algorithm called EM algorithm with substituting the obtained parameter θ into $\theta^{(t+1)}$ in the eq.(26). From the EM algorithm, we can derive update equations as following:

$$\boldsymbol{\mu}^{(t+1)} = \langle \boldsymbol{x} \rangle_{\theta^{(t)}}, \quad (29)$$

$$(\xi_m^2)^{(t+1)} = \left\langle x_m^2 \right\rangle_{\theta^{(t)}}, \quad (30)$$

$$\mathrm{Tr}\left((\alpha^{(t+1)}\Lambda + h^{(t+1)}I)^{-1}\Lambda\right) = \mathrm{Tr}\Lambda(\Xi^{(t)} + \alpha^{(t)}\Lambda + h^{(t)}I)^{-1} + (\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)})^{\mathrm{T}}\Lambda(\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}) \quad (31)$$

$$\mathrm{Tr}\left((\alpha^{(t+1)}\Lambda + h^{(t+1)}I)^{-1}\right) = \left\|\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}\right\|^2 \quad (32)$$

In order to obtain the inference parameter θ, we should solve the equation eq.(31) and (32).

3. Computer Simulation

In order to solve the eqs. (31) and (32), we introduce eigenvalue extraction of the determinant. The left hand of the equations come from the partial derivation for the $\ln|\alpha\Lambda + hI|$. Thus, we assume $\{\lambda_i\}$ as the eigenvalues of the matrix Λ and obtain the relationship

$$\ln|\alpha\Lambda + hI| = M\ln\alpha + \sum_i \ln\left(\lambda_i + \frac{h}{\alpha}\right), \quad (33)$$

where M is the size of the matrix Λ. Then, the simultaneous equations (31) and (32) can be denoted as

$$\frac{M}{\alpha} - \sum_i \frac{h}{\alpha\lambda_i + h} = \mathrm{Tr}\Lambda(\Xi^{(t)} + \alpha^{(t)}\Lambda + h^{(t)}I)^{-1} + (\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)})^{\mathrm{T}}\Lambda(\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}), \quad (34)$$

$$\sum_i \frac{\alpha}{\lambda_i + h} = \left\|\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}\right\|^2. \quad (35)$$

We solve the simultaneous equations for the α and h and assign them as $\alpha^{(t+1)}$ and $h^{(t+1)}$ respectively.

In the computer simulation, we use 32×32 pixels image for

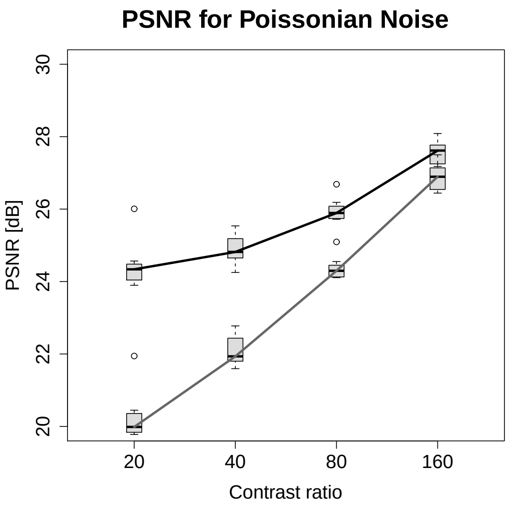

Fig. 3: Quantitative evaluation of image restoration: The horizontal axis shows the contrast ratio, and the vertical shows the peak signal to noise ratio (PSNR). The thick line shows the restoration result and the thin shows the observed one.

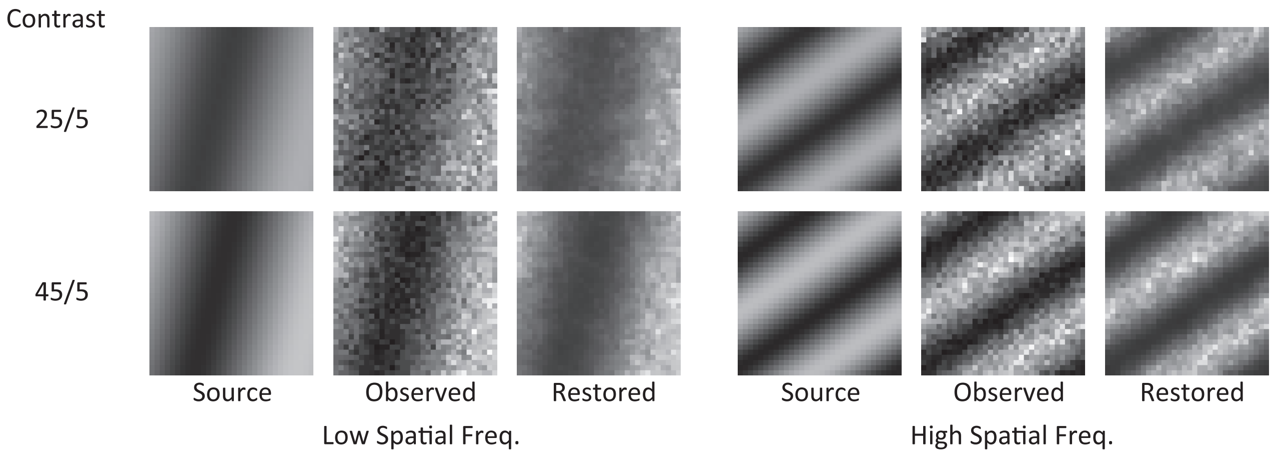

Fig. 2: Reconstruction result: The Left part shows the image with low spatial frequency and the right part shows the high frequency. In each part, the left shows the source image defined by $\rho_m\Delta$, the middle shows the Poisson observed image, and the right shows the restored image. The top row shows the case of low contrast image which means the minimum and the maximum of the $\rho_m\Delta$ is 5 and 25 respectively, which denote as $5/25$. The bottom row shows the high contrast case with $5/45$.

evaluation, and assume the original image consists of a spatial frequency. Each pixel is assigned event rate $\rho_m\Delta$, which controls the on-off event described in the eq.(1). We denote the lowest and highest event rates as $L_{\min}$ and $L_{\max}$ respectively, and define the contrast as $L_{\max}/L_{\min}$. In the simulation, each pixel divided into the $K = 10000$ miniregions.

In the initial value of the restoration image $\boldsymbol{\mu}^0$ as the observed image z, and the hyper-parameters $\alpha^0 = h^0$ is assumed as K. We use the relative errors of the restoration image and hyper-parameters as the convergence condition of the EM-algorithm, that is, $\frac{\sum_i |\mu_i^{t+1}-\mu_i^t|}{\sum_i |\mu_i|} < 10^{-4}$, $\frac{\alpha^{t+1}-\alpha^t}{\alpha^t} < 10^{-3}$, and $\frac{h^{t+1}-h^t}{h^t} < 10^{-3}$. In typical iteration requires about 200 times for convergence.

4. Results

Fig. 2 shows the restoration result of two spatial frequencies. The left part shows the result for the low spatial frequency image and the right one shows the high spatial frequency. In each part, the left, middle, and right columns show the original image $\{\rho_m\Delta\}$, observed image $\{z_m\}$, and restored image respectively. The top row shows the result for low contrast image which ratio $L_{\max}/L_{\min}$ equals $25/5$. The bottom shows the high contrast image with $45/5$. In the Poisson observation, the image contrast ratio controls the noise strength. Thus, the top row corresponds to the low S/N ratio, and the bottom shows the just higher than the top. Even though the corruption is high in the top row, the restoration image is just smoothed. On the contrary, the bottom row shows just lower corruption, so that the observation and the restoration images looks similar.

In order to evaluate restoration quantitatively, we introduce the peak signal noise to ratio (PSNR). Fig.3 shows the PSNR between original image $\rho_m\Delta$ and restored image with inverse logit transform. The horizontal axis shows the image contrast $L_{\max}/L_{\min}$, and the vertical shows the PSNR values. The right side of the plot means the high contrast that means the low observation noise area and the left means high observation noise. The evaluation is carried out with 10 times trials and plot with median with quantile deviation. From the plot, we can see the improvement by the restoration in the high observation noise area. On the contrary, in the low observation noise area, the restoration does not reduce the image quality.

5. Summary & Conclusion

In this study, we propose a image restoration method for the Poisson noised observation. For the inference of the Poisson parameters, we introduce a logit-transform and latent-valued form, and treat the observation process as a Gaussian function with latent-parameters. By this transformation, the observation process becomes tractable in the meaning of the Bayesian approach. By use of the GMRF like prior, obtaining the posterior mean is easy to derive. The induced latent-parameters and hyper-parameters are inferred by the EM algorithm. Thus, our algorithm can infer whole unknown parameters from the observation data.

In the computer simulation, we carry out several contrast ratio examples with two spatial frequencies cases that are defined on the small image plane 32×32 pixels. In the low contrast case that means high observation noise, we confirm the smoothness operation works well by visual comparison. For the quantitative evaluation, we introduce PSNR as the measure for the restoration. We also confirm our restoration algorithm works well within the low contrast area, and does not reduce image quality in within the high contrast area. Thus, we conclude our inference method work well with adapting in the noise strength automatically.

References

[1] Seeger, M. W.: Bayesian Inference and Optimal Design for the Sparse Linear Model, *Journal of Machine Learning Research*, Vol. 9, pp. 759–813 (2008).

[2] Palmer, J. A., Kerutz-Delgado, K., Wipf, D. P. and Rao, B.: Variational EM Algorithm for Non-Gaussian Latent Variable Models, *Advances in Neural Information Processing Systems 18*, Vol. 18, MIT Press, pp. 1059–1066 , (2005).

[3] Watanabe, K. and Okada, M.: Approximatate Bayesian Estimation of varying binomial process., *IEICE Transactions on Information and Systems*, Vol.E94-A, pp. 2879-2885 (2011).

[4] Tanaka, K.: Statistical-Mechanical Approach to Image Processing, *Journal of Physics A: Mathematical and General*, Vol. 35, No. 37, pp. R81–R150 , (2002).

