Hierarchical Gaussian Markov Random Field for Image Denoising

Yuki MONMA†, Kan ARO††, Nonmembers, and Muneki YASUDA†*, Member

SUMMARY In this study, Bayesian image denoising, in which the prior distribution is assumed to be a Gaussian Markov random field (GMRF), is considered. Recently, an effective algorithm for Bayesian image denoising with a standard GMRF prior has been proposed, which can help implement the overall procedure and optimize its parameters in $O(n)$-time, where $n$ is the size of the image. A new GMRF-type prior, referred to as a hierarchical GMRF (HGMRF) prior, is proposed, which is obtained by applying a hierarchical Bayesian approach to the standard GMRF prior; in addition, an effective denoising algorithm based on the HGMRF prior is proposed. The proposed HGMRF method can help implement the overall procedure and optimize its parameters in $O(n)$-time, as well as the previous GMRF method. The restoration quality of the proposed method is found to be significantly higher than that of the previous GMRF method as well as that of a non-local means filter in several cases. Furthermore, numerical evidence implies that the proposed HGMRF prior is more suitable for the image prior than the standard GMRF prior.

key words: Bayesian image denoising, Gaussian Markov random field, hierarchical Bayes method, linear-time algorithm

1. Introduction

Bayesian image processing [1] based on a graphical model has been used extensively for decades [2]. In Bayesian image denoising, a restored image is obtained from a posterior distribution. In the formulation of the posterior distribution, a corresponding prior distribution that captures the statistical properties of the images is needed. One of the major challenges of Bayesian image processing is the development of an effective prior for the images. A Gaussian Markov random field (GMRF) (or Gaussian graphical model) is a viable choice of a prior. Because a GMRF is equivalent to a multi-dimensional Gaussian distribution, its mathematical treatment is tractable; therefore, it can be used for various purposes other than Bayesian image denoising, e.g., image inpainting [3], traffic reconstruction [4], complex networks [5], sparse graphical modeling [6], and earth science [7], [8]. Therefore, the development of the GMRF prior is important for Bayesian image denoising and various fields.

In this study, we focus on Bayesian image denoising based on the GMRF [9], [10], which is divided into two stages: parameter estimation and restoration. In the parameter-estimation stage, all parameters in the model are optimized using a maximum likelihood (more precisely, a maximum marginal-likelihood), whereas in the restoration stage, the restored image is obtained by maximizing the posterior distribution with the optimal parameters estimated in the parameter-estimation stage. However, the processes used in the two stages (particularly, in the parameter-estimation stage) have a computational problem; they require the inverse operation of an $n \times n$ covariance matrix, the computational time of which is $O(n^3)$, where $n$ is the size of the image. Hence, to reduce the computational time, some approximations are employed. In the usual setting for Bayesian image denoising, the graph structure of GMRF is assumed to be a square grid graph according to the configuration of the pixels. By adding the periodic boundary condition to the original square grid graph, the authors of [9], [10] constructed $O(n^2)$-time algorithms. This approximation is referred to as a torus approximation because a square grid graph with a periodic boundary condition can be regarded as a torus graph. This approximation allows the graph Laplacian of the model to be diagonalized by using a discrete Fourier transformation, and the diagonalization can reduce the computational time to $O(n^2)$. More recently, this was further reduced to $O(n \ln n)$ by using fast Fourier transformation [11]. However, the computational cost of Bayesian image denoising based on GMRF without an approximation remains $O(n^3)$. Recently, an efficient algorithm for Bayesian image denoising based on GMRF was proposed [12], in which the denoising procedure, including parameter optimization, can be completed without any approximations in $O(n)$-time, which is the same cost as some well-known practical denoising filters, such as non-local means (NLM) [13] and block-matching and 3D (BM3D) filters [14].

The aim of this study is to improve the previous GMRF method [12]. The major merits of the previous GMRF method are as follows: (i) all parameters in the model can be automatically optimized; therefore users do not need to suffer from the parameter settings, and (ii) the overall procedure can be completed in $O(n)$-time. The former property can be the most important advantage for practical applications. However, it has two drawbacks, one of which is the limitation for the image size; the previous GMRF method can be applicable to $p \times p$ square images owing to the limitation in the eigenvalue analysis of the graph Laplacian [15], [16] employed in this study. Fortunately, this problem can be immediately resolved; the eigenvalue anal-
alysis provided in [17] allows the application to \( p \times q \) rectangular images, which is one of the contributions of this study. The other drawback is the restoration quality. There exist several high-quality denoising-filters such as NLM and BM3D filters. The restoration quality of the previous GMRF method is equal to or less than that of the NLM filter and is much lower than that of the BM3D filter in terms of the peak signal-to-noise ratio (PSNR) or structural similarity index measure (SSIM) [18]. However, the NLM and BM3D filters have several parameters that should be carefully tuned by users, e.g., the smoothness parameter \( h \). In both filters, to obtain a good restoration, the smoothness parameter \( h \) has to be set to a value close to the true standard deviation of noise to obtain good restorations.

This study contributes to an improvement of the (standard) GMRF prior. It focuses on the bias vector in the GMRF prior (i.e., \( \mathbf{b} \) in Eq. (3)). Several studies have considered the bias vector to be unimportant, and the bias vector has been frequently fixed to zero. However, it has an explicit role in a prior, which should not be neglected. The result of a least-mean-square analysis presented in Sect. 4 clarifies that the optimal \( \mathbf{b} \) is generally not zero. We propose a hierarchical prior, referred to as a hierarchical GMRF (HGMRF) prior, by introducing a hyperprior for the bias vector, and then propose a denoising procedure based on the HGMRF prior. The proposed method inherits the aforementioned two merits of the previous GMRF method, i.e., (i) and (ii). As demonstrated in Sect. 5.1, the restoration quality of the proposed method is significantly better than that of the previous GMRF method; moreover, it overcomes the NML filter in several cases, although it cannot reach the BM3D filter. In Sect. 5.2, the adequacies of GMRF and HGMRF priors as image priors are compared. The numerical results imply that the HGMRF prior is more suitable for an image prior.

As mentioned above, a GMRF prior is important for various problems other than Bayesian image denoising; therefore, an improvement of the GMRF prior is beneficial. The hyperprior employed in this study is simple and general (i.e., a simple Gaussian), and it is not specialized in image processing. This implies that the proposed HGMRF prior can be used for other problems, which could be an additional advantage of the proposed prior.

Image denoising has also developed in the field of deep learning [19]. Bayesian image processing focused on this study is different from such a machine learning approach. In a machine learning approach, a model is trained using a significantly large number of images; whereas, Bayesian image processing approaches such as GMRF and HGMRF do not need such expensive training, and the parameters in the models are optimized using only input degraded images.

The remainder of this paper is organized as follows: Sect. 2 describes the framework of Bayesian image denoising based on additive white Gaussian noise (AWGN). Section 3 briefly describes the denoising based on the standard GMRF prior [12]. The proposed hierarchical prior, i.e., the HGMRF prior, is introduced in Sect. 4; Sections 4.1 and 4.2 discuss the maximum a posterior (MAP) estimation and parameter estimation based on the proposed model, respectively, and Sect. 4.3 describes the pseudo-code of the proposed denoising procedure. Section 5 demonstrates the validity of the proposed method using numerical experiments. Section 6 summarizes this study and discusses the area of future research.

2. Framework of Bayesian Image Denoising

We consider the problem of image denoising as described below. Given an original image comprising \( p \times q \) pixels, a degraded image is generated by adding AWGN to the original image. Suppose that \( K \) degraded images independently generated from the same stochastic process are obtained, where \( K \geq 1 \) [12], [20]. The goal of Bayesian image denoising is to infer the original image from the given data, i.e., the \( K \) degraded images. The original image is denoted by the vector \( \mathbf{x} := (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \) (\( x_i \) expresses the intensity of the \( i \)th pixel), where the image is vectorized by raster scanning (or row-major scanning) on the image, and \( n := pq \) is the number of pixels in the image; the \( k \)th degraded image is denoted by the vector \( \mathbf{y}^{(k)} := (y_{1}^{(k)}, y_{2}^{(k)}, \ldots, y_{n}^{(k)}) \in \mathbb{R}^n \). The degraded images are vectorized by the same procedure as the original image. The AWGN is assumed, and the noise distribution for the \( K \) degraded images \( Y_{K} := \{y_{i}^{(k)} | k = 1, 2, \ldots, K \} \) is defined as

\[
P_{\text{noise}}(Y_{K} | \mathbf{x}, \sigma^2) := \prod_{k=1}^{K} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - y_{i}^{(k)})^2}{2\sigma^2}\right),
\]

where \( \sigma > 0 \) is the standard deviation (or the noise level) of the AWGN. The framework of the Bayesian image denoising is illustrated in Fig. 1.

3. Basic Prior: Gaussian Markov Random Field

One basic choice of a prior distribution is a GMRF [12]: a GMRF based on an undirected graph \( G(V, E) \) is defined as

\[
P_{\text{pri}}(\mathbf{x} | \theta_{\text{pri}}) \propto \exp\left(-E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}})\right),
\]

where \( E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}}) \) is the energy function. The energy function is defined as

\[
E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}}) := \sum_{(i,j) \in E} \theta_{ij} \phi_{ij}(x_i - x_j),
\]

where \( \phi_{ij}(x_i - x_j) \) is a neighbourhood function and \( \theta_{ij} \) is a parameter that represents the interaction between the \( i \)th and \( j \)th pixels. The parameter estimation based on the proposed model, respectively, and Sect. 4.3 describes the pseudo-code of the proposed denoising procedure. Section 5 demonstrates the validity of the proposed method using numerical experiments. Section 6 summarizes this study and discusses the area of future research.

2. Framework of Bayesian Image Denoising

We consider the problem of image denoising as described below. Given an original image comprising \( p \times q \) pixels, a degraded image is generated by adding AWGN to the original image. Suppose that \( K \) degraded images independently generated from the same stochastic process are obtained, where \( K \geq 1 \) [12], [20]. The goal of Bayesian image denoising is to infer the original image from the given data, i.e., the \( K \) degraded images. The original image is denoted by the vector \( \mathbf{x} := (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \) (\( x_i \) expresses the intensity of the \( i \)th pixel), where the image is vectorized by raster scanning (or row-major scanning) on the image, and \( n := pq \) is the number of pixels in the image; the \( k \)th degraded image is denoted by the vector \( \mathbf{y}^{(k)} := (y_{1}^{(k)}, y_{2}^{(k)}, \ldots, y_{n}^{(k)}) \in \mathbb{R}^n \). The degraded images are vectorized by the same procedure as the original image. The AWGN is assumed, and the noise distribution for the \( K \) degraded images \( Y_{K} := \{y_{i}^{(k)} | k = 1, 2, \ldots, K \} \) is defined as

\[
P_{\text{noise}}(Y_{K} | \mathbf{x}, \sigma^2) := \prod_{k=1}^{K} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - y_{i}^{(k)})^2}{2\sigma^2}\right),
\]

where \( \sigma > 0 \) is the standard deviation (or the noise level) of the AWGN. The framework of the Bayesian image denoising is illustrated in Fig. 1.

3. Basic Prior: Gaussian Markov Random Field

One basic choice of a prior distribution is a GMRF [12]: a GMRF based on an undirected graph \( G(V, E) \) is defined as

\[
P_{\text{pri}}(\mathbf{x} | \theta_{\text{pri}}) \propto \exp\left(-E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}})\right),
\]

where \( E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}}) \) is the energy function. The energy function is defined as

\[
E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}}) := \sum_{(i,j) \in E} \theta_{ij} \phi_{ij}(x_i - x_j),
\]

where \( \phi_{ij}(x_i - x_j) \) is a neighbourhood function and \( \theta_{ij} \) is a parameter that represents the interaction between the \( i \)th and \( j \)th pixels. The parameter estimation based on the proposed model, respectively, and Sect. 4.3 describes the pseudo-code of the proposed denoising procedure. Section 5 demonstrates the validity of the proposed method using numerical experiments. Section 6 summarizes this study and discusses the area of future research.

Fig. 1 Framework of Bayesian image denoising. \( P_{\text{pri}}(\mathbf{x}) \), \( P_{\text{noise}}(Y_{K} | \mathbf{x}, \sigma^2) \), and \( P_{\text{post}}(\mathbf{x} | Y_{K}) \) denote the prior distribution, noise distribution, and posterior distribution, respectively. \( Y_{K} \) is the set of \( K \) degraded images.
with the energy function

$$E_{\text{pri}}(\mathbf{x}; \theta_{\text{pri}}) := - \sum_{v \in V} b_i x_i + \frac{1}{2} \sum_{v \in V} x_i^2 + \frac{\alpha}{2} \sum_{\{i,j\} \in E} (x_i - x_j)^2. \tag{3}$$

Here, $V := \{1, 2, \ldots, n\}$ and $E := \{\{i,j\} \mid i, j \in V \text{ are adjacent}\}$ are the sets of vertices and undirected edges in $G(V, E)$, respectively; because an undirected graph is considered, $\{i, j\}$ and $\{j, i\}$ indicate the same edge. Here, $\theta_{\text{pri}} := \{b, \lambda, \alpha\}$ is the set of parameters in the prior distribution, i.e., $b, \lambda, \alpha \in \mathbb{R}^n$ is the bias vector that controls the brightness (or the contrast), $\lambda > 0$ controls the variance of the intensity, and $\alpha > 0$ controls the smoothness. In this study, $G(V, E)$ is presupposed to be a $p \times q$ rectangular grid graph corresponding to the structure of the pixels, in which the edges exist only between the nearest-neighbors. However, almost all analyses described below will be validated for any undirected graphs.

The energy function of the prior distribution in Eq. (3) can be rewritten as

$$E_{\text{pri}}(\mathbf{x}; \theta) = -b^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{S}_{\text{pri}} \mathbf{x}, \tag{4}$$

where

$$\mathbf{S}_{\text{pri}} := \lambda \mathbf{I}_n + \alpha \mathbf{L}_n \in \mathbb{R}^{n \times n} \tag{5}$$

is the precision matrix (the inverse of the covariance matrix) of the prior distribution; here, $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ is the $n$-dimensional identity matrix, and $\mathbf{L}_n \in \mathbb{R}^{n \times n}$ is the graph Laplacian of $G(V, E)$ defined by

$$L_{i,j} = \begin{cases} \partial(i) & i = j \\ -1 & \{i, j\} \in E \\ 0 & \text{others} \end{cases} \tag{6}$$

where $\partial(i) \subset V$ is the set of vertices connected to vertex $i$; because $G(V, E)$ is a $p \times q$ rectangular grid graph, $|\partial(i)| \leq 4$ for $i \in V$. Note that the positivity of $\lambda$ is important for the mathematical treatment of the prior distribution because when $\lambda = 0$, the precision matrix $\mathbf{S}_{\text{pri}}$ is not positive definite; therefore, Eq. (2) is not a well-defined probabilistic distribution in this case.

Using Eqs. (1) and (2) and the Bayesian theorem, the posterior distribution is expressed as

$$P_{\text{post}}(\mathbf{x} \mid Y_K, \theta_{\text{post}}) \propto P_{\text{noise}}(Y_K \mid \mathbf{x}, \sigma^2)P_{\text{pri}}(\mathbf{x} \mid \theta_{\text{pri}}) \propto \exp \left( \mathbf{c}^T \mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{S}_{\text{post}} \mathbf{x} \right), \tag{7}$$

where $\theta_{\text{post}} := \{b, \lambda, \alpha, \sigma^2\}$ and

$$\mathbf{c} := b + \frac{K}{\sigma^2} \hat{y} \in \mathbb{R}^n, \tag{8}$$

$$\mathbf{S}_{\text{post}} := \frac{K}{\sigma^2} \mathbf{I}_n + \mathbf{S}_{\text{pri}} \in \mathbb{R}^{n \times n} \tag{9}$$

are the bias vector and precision matrix of the posterior, respectively. Here,

$$\hat{y} := \frac{1}{K} \sum_{k=1}^{K} y^{(k)} \in \mathbb{R}^n \tag{10}$$

is the average image of $K$ degraded images. The restored image is obtained using the MAP estimation according to Eq. (7), and the MAP solution is

$$\mathbf{m} := \mathbf{S}^{-1}_{\text{post}} \mathbf{c} \in \mathbb{R}^n, \tag{11}$$

which can be efficiently obtained by solving the mean-field equation [12], [21], [22], i.e.,

$$m_i = \frac{b_i + (K/\sigma^2)\hat{y}_i + \alpha \sum_{j \in \partial(i)} m_j}{\lambda + K/\sigma^2 + \alpha |\partial(i)|}. \tag{12}$$

This is known as a Gauss–Seidel method or Jacobi method. By solving this equation using a successive substitution method, $\mathbf{m}$ can be obtained in $O(n)$-time because the sum on the right hand side of this equation includes at most four terms (because $|\partial(i)| \leq 4$) and $m_i$ on the left hand side of this equation can then be computed in $O(1)$-time; here, the number of iterations is considered a constant. For a detailed derivation of Eq. (12), refer to [12].

When $\mathbf{b}$ is a constant vector, i.e., $\mathbf{b} = \{b, b, \ldots, b\}$, an efficient algorithm for the aforementioned denoising scheme was proposed in [12], in which the overall procedure, including the optimization of all parameters, is conducted in $O(n)$-time. Moreover, it was proved that $b = 0$ (i.e., $\mathbf{b} = \mathbf{0}$) is optimal from the perspective of the maximum likelihood, when $\hat{y}$ is zero-centered during the following preprocessing:

$$\hat{y}_i \leftarrow \hat{y}_i - \bar{y}_{\text{ave}}, \quad i \in V, \tag{13}$$

where $y_{\text{ave}} := \frac{1}{n} \sum_{i \in V} \hat{y}_i$ is the average intensity of $\hat{y}$. Although the linear-time algorithm and the optimality analysis for $b$ presented in [12] are proposed for $p \times p$ square images, they can be naturally expanded to $p \times q$ rectangular images (the details of which is described in Sect. 4.2).

In integrations defined later, the explicit description of the integral interval is omitted; they should be read as integrations over the corresponding real space, i.e., for a scalar $z \in \mathbb{R}$, the integration $\int f(z)dz$ should be read as $\int_{-\infty}^{\infty} f(z)dz$. Moreover, for a vector (or set) $\mathbf{z}$, the integration $\int f(\mathbf{z})d\mathbf{z}$ should be read as a multiple integration for all elements in $\mathbf{z}$.

4. Hierarchical Gaussian Markov Random Field

Consider a mean squared error (MSE) between an original image $\mathbf{x}^{\text{true}}$ and the corresponding restored image $\mathbf{m}$ obtained using Eq. (11), as

$$\int ||\mathbf{x}^{\text{true}} - \mathbf{m}||_2^2 P_{\text{noise}}(Y_K \mid \mathbf{x}^{\text{true}}, \sigma^2)dy_K, \tag{14}$$

where $\| \cdot \|_2$ denotes the $L_2$ norm (or the Euclidean norm). For fixed $\lambda, \alpha,$ and $\sigma^2$, Eq. (14) is minimized when

$$\mathbf{b} = \mathbf{S}_{\text{pri}}^{-1} \mathbf{x}^{\text{true}}. \tag{15}$$
The detailed derivation of Eq. (15) is described in Appendix A. This result is trivial because $x^\text{true}$ yields the maximum probability of the prior distribution when $b = S_{\text{pri}}x^\text{true}$. This implies that $b = 0$ is not always optimal when $b$ is not a constant vector. However, the setting of Eq. (15) is not practical because it needs the original image. Alternatively, we introduce a hyperprior for $b$, and regard $b$ as random variables, which is called the hierarchical Bayesian approach [23].

The appropriate setting of the hyperprior is important and should be carefully considered; however, a simple Gaussian

$$P_{\text{bias}}(b \mid \gamma^2) \propto \exp \left( -\frac{1}{2\gamma^2}b'b \right)$$

(16)
is considered in this study to simplify the following analyses. The proposed hierarchical prior is defined by

$$Q_{\text{pri}}(x \mid h_{\text{pri}}) := \int P_{\text{pri}}(x \mid \theta_{\text{pri}})P_{\text{bias}}(b \mid \gamma^2)db$$

$$\propto \exp \left( -\frac{1}{2}x'H_{\text{pri}}x \right).$$

(17)

where $h_{\text{pri}} := \{\lambda, \alpha, \gamma^2\}$ are the parameters of the hierarchical prior and

$$H_{\text{pri}} := S_{\text{pri}} - \gamma^2(I_n + \gamma^2S_{\text{pri}}^{-1})^{-1} \in \mathbb{R}^{n \times n}$$

(18)
is the precision matrix. Equation (17) is a GMRF, and we refer to it as a HGMRF prior. The GMRF prior in Eq. (2) is log-linear, whereas the HGMRF prior in Eq. (17) is not.

4.1 MAP Estimation for HGMRF

Using the HGMRF of Eq. (17) instead of Eq. (2), the posterior distribution is obtained as

$$Q_{\text{post}}(x \mid Y_K, h_{\text{post}}) \propto P_{\text{noise}}(Y_K \mid x, \sigma^2)Q_{\text{pri}}(x \mid h_{\text{pri}})$$

$$\propto \exp \left( g'x - \frac{1}{2}x'H_{\text{post}}x \right).$$

(19)

where $h_{\text{post}} := \{\lambda, \alpha, \gamma^2, \sigma^2\}$ are the parameters of this posterior, and

$$g := \frac{K}{\sigma^2} \hat{y} \in \mathbb{R}^n,$$

(20)

$$H_{\text{post}} := \frac{K}{\sigma^2}I_n + H_{\text{pri}} \in \mathbb{R}^{n \times n}.$$  

(21)

The restored image is obtained using the MAP estimation according to Eq. (19). The MAP solution is

$$u := H_{\text{post}}^{-1}g \in \mathbb{R}^n.$$  

(22)

The MAP solution $u$ can be efficiently obtained as follows. Multiplying $H_{\text{post}}$ to both sides of Eq. (22) leads to

$$u = S_{\text{post}}^{-1}(g + \gamma^2v),$$

(23)

where

$$v := (I_n + \gamma^2S_{\text{pri}}^{-1})^{-1}u \in \mathbb{R}^n.$$  

(24)

Equations (23) and (24) can be regarded as the simultaneous equations for $u$ and $v$. The solution to these simultaneous equations can be obtained by solving the following mean-field equations:

$$u_i = \frac{(K/\sigma^2)\hat{y}_i + \gamma^2v_i + \alpha \sum_{j \neq i} u_j}{\lambda + K/\sigma^2 + \alpha}$$

(25)

$$v_i = \frac{(\lambda + \alpha)\hat{y}_i u_i + \alpha \sum_{j \neq i} (v_j - u_j)}{\lambda + \alpha + \lambda}.$$  

(26)

By solving these equations using a successive substitution method, $u$ and $v$ can be obtained in $O(n)$-time; the reason of the linear cost is the same as that in Eq. (12). When $\gamma^2 \to 0$, $u$ obtained from Eqs. (25) and (26) coincides with $m$ in Eq. (11). Note that Eqs. (25) and (26) (also, Eq. (41) appearing later) can be obtained in a similar manner as that of Eq. (12).

4.2 Parameter Estimation

For a fixed $h_{\text{post}}$, the restored image can be obtained by solving Eqs. (25) and (26). The denoising quality significantly depends on the values of the parameters. In this section, the parameter optimization based on the maximum marginal-likelihood estimation is considered.

For the given $Y_K$, consider the marginal distribution

$$Q_{\text{mar}}(Y_K \mid h_{\text{post}})$$

$$= \int P_{\text{noise}}(Y_K \mid x, \sigma^2)Q_{\text{pri}}(x \mid h_{\text{pri}})dx$$

$$= \exp \left( -\frac{1}{2\sigma^2} \sum_{k=1}^K \|y^{(k)}\|^2 \right) Z_{\text{post}}(Y_K, h_{\text{post}}),$$

(27)

where $Z_{\text{pri}}(h_{\text{pri}})$ and $Z_{\text{post}}(Y_K, h_{\text{post}})$ are the normalization constants for $Q_{\text{pri}}(x \mid h_{\text{pri}})$ and $Q_{\text{post}}(x \mid Y_K, h_{\text{post}})$, respectively, which are

$$Z_{\text{pri}}(h_{\text{pri}}) := \sqrt{2\pi\sigma^2} \det H_{\text{pri}}^{-1},$$

(28)

$$Z_{\text{post}}(Y_K, h_{\text{post}}) := \exp \left( \frac{1}{2}g'H_{\text{post}}^{-1}g \right).$$

(29)

From the perspective of the maximum likelihood, the optimal $h_{\text{post}}$ is obtained through the maximization of the log-marginal-likelihood:

$$\ell(h_{\text{post}}) := \frac{1}{nk} \ln Q_{\text{mar}}(Y_K \mid h_{\text{post}}).$$

(30)

Using Eqs. (27) and (30), the expression

$$\ell(h_{\text{post}}) = -\frac{1}{2n\sigma^2} \sum_{k=1}^K \|y^{(k)}\|^2 + \frac{1}{2nK} g'u$$

$$+ \frac{1}{2nK} \ln \det H_{\text{post}} - \frac{1}{2} \ln (2\pi\sigma^2)$$

(31)
is obtained, where Eq. (22) is used. Suppose that the
eigenvalues of the graph Laplacian \( \Lambda \) are denoted by \( e_1, e_2, \ldots, e_n \). The eigenvalues of \( H_{\text{pri}} \) are expressed as

\[
\phi_i := \lambda + \alpha e_i - \gamma^2 (1 + \frac{\gamma^2}{\lambda + \alpha e_i})^{-1} = \frac{(\lambda + \alpha e_i)^2}{\gamma^2 + \lambda + \alpha e_i}
\]  
(32)

for \( i \in V \), and the eigenvalues of \( H_{\text{post}} \) are

\[
\chi_i := \frac{K}{\sigma^2} + \phi_i
\]  
(33)

for \( i \in V \). Using these eigenvalues, Eq. (31) is rewritten as

\[
\ell(h_{\text{post}}) = -\frac{1}{2n\sigma^2 K} \sum_{k=1}^{K} ||y^{(k)}||_2^2 + \frac{1}{2nK} g'u + \frac{1}{2nK} \sum_{i \in V} \ln \chi_i - \frac{1}{2} \ln (2\pi \sigma^2).
\]  
(34)

For the maximization of the log-marginal-likelihood based on a gradient ascent method, the gradients of \( \ell(h_{\text{post}}) \) with respect to the parameters in \( h_{\text{post}} \) are required, which are obtained as follows:

\[
\nabla \lambda = -\frac{1}{2nK} ||u||_2^2 + \frac{\gamma^4 - 1}{2nK ||w||_2^2} \sum_{i \in V} \frac{1}{\chi_i} \left( \lambda + \alpha e_i - \frac{\gamma^2 + \lambda + \alpha e_i}{\gamma^2 + \lambda + \alpha e_i} \right),
\]  
(35)

\[
\nabla \alpha = -\frac{1}{2nK} \text{Sm}(u) + \frac{\gamma^4}{2nK} \text{Sm}(w) + \frac{1}{2nK} \sum_{i \in V} \frac{e_i}{\chi_i} \left( \lambda + \alpha e_i - \frac{\gamma^2 + \lambda + \alpha e_i}{\gamma^2 + \lambda + \alpha e_i} \right),
\]  
(36)

\[
\nabla \gamma^2 = \frac{1}{2nK} ||v||_2^2 - \frac{1}{2nK} \sum_{i \in V} \frac{1}{\chi_i} \left( \gamma^2 + \lambda + \alpha e_i \right),
\]  
(37)

\[
\nabla \sigma^2 = \frac{1}{2n\sigma^2 K} \sum_{k=1}^{K} ||y^{(k)} - u||_2^2 + \frac{1}{2n\sigma^2} \sum_{i \in V} \frac{1}{\chi_i} - \frac{1}{2\sigma^2}.
\]  
(38)

Here, \( \nabla z \) denotes the gradient with respect to \( z \), i.e., \( \nabla z := \frac{\partial \ell(h_{\text{post}})}{\partial z} \), and the vector \( w \) is defined as

\[
w := S_{\text{pri}}^{-1} v \in \mathbb{R}^n.
\]  
(39)

In these gradients, \( u \) and \( v \) are the solutions to the mean-field equations in Eqs. (25) and (26), and for \( a \in \mathbb{R}^n \), the function \( \text{Sm}(a) \) is defined as

\[
\text{Sm}(a) := a^t \Lambda a = \sum_{i,j \in E} (a_i - a_j)^2,
\]  
(40)

which corresponds to the smoothness term in the energy function. The detailed derivation of Eqs. (35)–(38) is presented in Appendix B. The vector \( w \) defined in Eq. (39) can be obtained by solving the mean-field equation:

\[
w_i = v_i + \alpha \sum_{j \in \partial(i)} w_j / \lambda + \alpha |\partial(i)|.
\]  
(41)

Using a successive substitution method, \( w \) can be obtained in \( O(n) \)-time; the reason for the linear cost is the same as that in Eqs. (12), (25), and (26).

The vectors \( u \), \( v \), and \( w \) can be obtained in \( O(n) \)-time, and \( \text{Sm}(a) \) can be computed in \( O(n) \)-time because \( |E| < 4n \); therefore, all gradients in Eqs. (35)–(38) can be obtained in \( O(n) \)-time if the \( n \) eigenvalues of the graph Laplacian \( [e_i] \) are obtained. In general, for a large matrix, the computation of the eigenvalues is intractable. However, fortunately, the eigenvalues of the graph Laplacian are known when \( G(V, E) \) is a \( p \times q \) rectangular grid graph [17], i.e.,

\[
\{4 \sin^2 \frac{\pi(i - 1)}{2p} + 4 \sin^2 \frac{\pi(j - 1)}{2q}\}
\]  
(42)

for \( i = 1, 2, \ldots, p \) and \( j = 1, 2, \ldots, q \).

At the maximum point of \( \ell(h_{\text{post}}) \), the relation

\[
\sigma^2 = \frac{1}{nK} \sum_{k=1}^{K} ||y^{(k)} - u||_2^2 + \frac{1}{n} \sum_{i \in V} \frac{1}{\chi_i}
\]  
(43)

is satisfied, which is obtained by setting \( \nabla \sigma^2 = 0 \) in Eq. (38).

4.3 Proposed Denoising Algorithm

The pseudo-code of the proposed denoising procedure is shown in Algorithm 4.1. From Steps 11–18, the vectors \( u \), \( v \), and \( w \) are computed by solving the mean-field equations using a successive substitution method with a warm restart, for which an asynchronous updating is recommended. In the experiments mentioned in the following sections, the process was stopped after just one and two iterations, i.e., \( T_{\text{mf}}^{(1)} = 2 \) and \( T_{\text{mf}}^{(2)} = 1 \). This early-stopping negligibly impacts the final results. The initial values of the parameters and step sizes of the gradient ascent are quite important because they affect the restoration performance and the stability of convergence. The experiments described in this and the following sections set the parameters as \( \lambda_0 = 10^{-7} \), \( \alpha_0 = 10^{-4} \), \( \gamma_0^2 = 10^{-3} \), \( \sigma_0^2 = 10^3 \), \( \eta_1 = 10^{-12} \), \( \eta_2 = 10^{-6} \), and \( \eta_3 = 5 \times 10^{-3} \); in addition, the convergence criteria \( \varepsilon \) was set to \( 10^{-3} \). Empirically, the above setting stabilized the convergences and allowed the high-performance restorations for various images; however, a more appropriate setting might exist.

For a colored image (an image having multiple channels), denoising is achieved by separately applying this method to each channel. Additionally, Algorithm 4.1 is essentially the same as the GMRF method proposed in [12] when \( \gamma^2 \) is fixed to zero; note that, \( v \) and \( w \) can be eliminated in the case of \( \gamma^2 = 0 \). For the GMRF method used in all experiments, we applied Algorithm 4.1 under the condition of \( \gamma^2 = 0 \), in which the setting of the initial values and step sizes are the same as those of the HGMRF method except for \( r_a = 5 \times 10^{-7} \).

The real computational time of the proposed HGMRF method depends on several factors, such as the size, structure of the \( K \) degraded images, and noise level. Experimentally, in almost all cases, the proposed method is completed
Algorithm 4.1 Denoising based on HGMRF

1: Input $K$ degraded images $Y_K$
2: // zero-centering (mean-shifting)
3: for $i \in V$ do
4: \( \hat{y}_i \leftarrow y_i - y_{ave} \)
5: end for
6: Initialize: $\lambda = \lambda_0$, $\alpha = \alpha_0$, $\gamma^2 = \gamma^2_0$, and $\sigma^2 = \sigma^2_0$
7: Initialize: $u = v = w = \hat{y}$
8: repeat
9: $u^{old} = u$
10: // solve the mean-field equations
11: for $t = 1,2,\ldots,T^{(1)}$ do
12: for $i \in V$ do
13: \( u_i \leftarrow \frac{(K/\sigma^2)\hat{y}_i + \gamma^2 y_i + \alpha \sum_{j \in n(i)} u_j}{\lambda + K/\sigma^2 + \alpha |\partial \hat{y}(i)|} \)
14: \( v_i \leftarrow \frac{(\lambda + \alpha |\partial \hat{y}(i)|)u_i + \alpha \sum_{j \in n(i)} (v_i - u_j)}{\lambda + \gamma^2 + \alpha |\partial \hat{y}(i)|} \)
15: end for
16: end for
17: // parameters update
18: Update $\sigma^2$ using Eq. (43)
19: $\lambda \leftarrow \lambda + \eta_\lambda \nabla \lambda$ using Eq. (35)
20: $\alpha \leftarrow \alpha + \eta_\alpha \nabla \alpha$ using Eq. (36)
21: $\gamma^2 \leftarrow \gamma^2 + \eta_\gamma \nabla \gamma^2$ using Eq. (37)
22: until $\|u - u^{old}\|_2/n < \epsilon$
23: // reverse-mean-shifting
24: for $i \in V$ do
25: \( u_i \leftarrow u_i + y_{ave} \)
26: end for
27: Output $u$

within a few seconds. For example, when $K$ degraded images are generated from the image in Fig. 2 (c) with a noise level of $\sigma = 20$, the real computational times of the proposed method for several $K$ are as listed in Table 1. For comparison, the computational times of the GMRF method are also listed. The proposed method is several times slower than the GMRF method because the HGMRF method has more estimates than the GMRF method. The GMRF and HGMRF methods are faster with an increase in $K$; this is the standard observation in Bayesian inferences, i.e., an increase of data points helps a fast convergence of the algorithm. For reference, the computational times of the NLM and BM3D filters in the restoration of Fig. 2 (c) with $\sigma = 20$ were 646.0 [ms] and 3578.0 [ms], respectively, for which the smoothness parameter $h$ of both filters was set to $h = \sigma^{1+}$.

5. Numerical Experiment

In this section, numerical results are demonstrated using the 8-bit images shown in Fig. 2.

5.1 Denoising with One Degraded Image

In this section, the denoising results in the case of $K = 1$, the most practical case, are presented. The results obtained from the GMRF and proposed HGMRF methods are listed in Tables 2–5 for several noise levels $\sigma$, in which the PSNR and SSIM are used as restoration measures. The values in these tables are the average values obtained from 20 experiments. During these experiments, degraded and restored images were quantized into 8-bit images. For comparison, the results of the NLM filter are presented. In the NLM filter, its smooth parameter $h$ was set to the “true” noise level of AWGN, i.e., non-blind denoising. By contrast, the GMRF and HGMRF methods are blind denoising.

Based on these results, it is found that the HGMRF method is significantly superior to the GMRF method in terms of PSNR and SSIM. Moreover, the HGMRF method overcomes the NLM filter in several cases, in particular, high-noise-level cases. Figure 3 shows examples of the denoising results of Figs. 2 (b) and (d). In the GMRF results, fine noise remains; however, they are removed in the HGMRF results. By contrast, in the low-noise-level cases

---

Table 1 The real computation times for the $K$ degraded images generated from the image in Fig. 2 (c) with $\sigma = 20$. The values in this table are the average times obtained from 20 experiments.

| $K$ | GMRF [ms] | HGMRF [ms] |
|-----|-----------|------------|
| 1   | 276.7     | 1196.5     |
| 2   | 82.9      | 414.1      |
| 3   | 73.0      | 288.2      |
| 4   | 69.4      | 216.0      |
| 5   | 68.0      | 127.8      |

††The NLM and BM3D filters were implemented using OpenCV, for which the parameters of both filters, other than the smoothness parameter $h$, were the default values of OpenCV.
The results of the HGMRF method can be lower when the parameter setting is not appropriate. In Table 7, the denoising results of the NLM and BM3D filters for the images in Figs. 2 (b) and (d), along with several values of their smoothness parameter $h$ are shown, for which the true noise level is $\sigma = 20$. In this table, the results of $h = 20$ are the same as those shown in Tables 3, 5, and 6. When $h \neq \sigma$, the denoising quality of the BM3D filter (as well as that of the NLM filter) can significantly degrade; therefore, the appropriate setting of the smoothness parameter is very important; however, the true noise level is unknown. By contrast, such a problem does not exist in the current Bayesian approach because all parameters, including the noise level, can be estimated from given degraded images.

Next, we conducted a denoising experiment for the gray-scaled version of the image in Fig. 2(a) using the GMRF and HGMRF methods. The normalized histograms of the original and restored images are shown in Fig. 4, for which a noise level of $\sigma = 30$ was used. The histogram of the HGMRF method is more similar to that of the original image, particularly, in the tails; in other words, the HGMRF method maintains the lighting contrast of the original image owing to the effect of the hyperprior for $b$.

### Table 2: Denoising results of Fig. 2(a).

| Noise level | PSNR       | SSIM       |
|-------------|------------|------------|
| $\sigma = 10$ | 28.16 29.71 28.44 | 0.665 0.874 0.676 | 0.719 |
| $\sigma = 20$ | 22.25 26.91 25.34 | 0.416 0.801 0.553 | 0.695 |
| $\sigma = 30$ | 18.86 24.95 24.79 | 0.290 0.740 0.535 | 0.679 |
| $\sigma = 40$ | 16.54 23.55 24.22 | 0.216 0.694 0.522 | 0.662 |

### Table 3: Denoising results of Fig. 2(b).

| Noise level | PSNR       | SSIM       |
|-------------|------------|------------|
| $\sigma = 10$ | 28.16 29.74 28.38 | 0.821 0.875 0.826 | 0.849 |
| $\sigma = 20$ | 22.22 25.70 24.59 | 0.593 0.689 0.680 | 0.756 |
| $\sigma = 30$ | 18.82 23.46 23.82 | 0.472 0.549 0.625 | 0.672 |
| $\sigma = 40$ | 16.49 22.23 23.15 | 0.324 0.471 0.579 | 0.600 |

### Table 4: Denoising results of Fig. 2(c).

| Noise level | PSNR       | SSIM       |
|-------------|------------|------------|
| $\sigma = 10$ | 28.44 28.79 28.62 | 0.734 0.796 0.739 | 0.760 |
| $\sigma = 20$ | 22.68 25.00 24.73 | 0.508 0.673 0.585 | 0.682 |
| $\sigma = 30$ | 19.36 23.97 23.84 | 0.360 0.596 0.539 | 0.611 |
| $\sigma = 40$ | 17.06 21.72 22.93 | 0.280 0.538 0.498 | 0.556 |

### Table 5: Denoising results of Fig. 2(d).

| Noise level | PSNR       | SSIM       |
|-------------|------------|------------|
| $\sigma = 10$ | 28.17 28.38 28.28 | 0.816 0.899 0.818 | 0.821 |
| $\sigma = 20$ | 22.23 21.77 22.79 | 0.624 0.655 0.642 | 0.726 |
| $\sigma = 30$ | 18.91 19.37 21.93 | 0.491 0.487 0.594 | 0.656 |
| $\sigma = 40$ | 16.68 18.46 21.43 | 0.396 0.414 0.557 | 0.587 |

(i.e., the cases of $\sigma = 10$), the results of the HGMRF method can be worse than those of the NLM filter due to over-smoothing caused by the prior.

In addition, we compared the HGMRF method with the BM3D filter using the same experiments. The results obtained from the BM3D filter were better than those of the HGMRF method in terms of PSNR and SSIM (Table 6); for example, the PSNR results of BM3D were 1–2 [dB] higher on average. With the BM3D method, non-blind denoising was used (i.e., the smoothness parameter of the BM3D filter was set to the true noise level). The denoising quality of the BM3D filter can be lower when the parameter setting is not appropriate. In Table 7, the denoising results of the NLM and BM3D filters with several values of $h$. The values in this table are the average values obtained from 20 experiments. The true noise level was $\sigma = 20$.

### Table 6: Denoising results of non-blind BM3D filter. The values in this table are the average values obtained from 20 experiments.

| Noise level | PSNR       | SSIM       |
|-------------|------------|------------|
| $\sigma = 10$ | 34.11 0.920 30.60 0.892 | 31.51 0.892 30.74 0.918 |
| $\sigma = 20$ | 30.52 0.850 27.01 0.777 | 27.89 0.780 26.71 0.830 |
| $\sigma = 30$ | 28.32 0.780 25.27 0.692 | 25.91 0.693 24.54 0.750 |
| $\sigma = 40$ | 26.58 0.715 23.99 0.619 | 24.37 0.619 22.98 0.674 |

### Table 7: Denoising results of the NLM and BM3D filters with several values of $h$. The values in this table are the average values obtained from 20 experiments. The true noise level was $\sigma = 20$.

| Noise level | PSNR       | SSIM       |
|-------------|------------|------------|
| $h = 10$ | 22.32 0.598 24.37 0.731 | 24.75 0.701 24.63 0.725 |
| $h = 20$ | 25.70 0.689 21.77 0.655 | 27.01 0.777 26.71 0.830 |
| $h = 30$ | 23.21 0.515 19.27 0.470 | 25.96 0.715 25.34 0.781 |
| $h = 40$ | 22.15 0.442 18.47 0.401 | 25.33 0.673 24.26 0.733 |
5.2 GMRF Versus HGMRF Priors

The difference between the GMRF and HGMRF methods is the form of the prior. The numerical results presented in Sect. 5.1 show that the denoising quality of the HGMRF method is higher than that of the GMRF method; which implies that the HGMRF prior defined in Eq. (17) is more appropriate than the GMRF prior defined in Eq. (2). Both priors are compared by considering an alternative viewpoint.

If a perfect prior can be obtained, the estimate of $\sigma^2$ obtained from the maximization of the log-marginal-likelihood discussed in Sect. 4.2 should perfectly agree with the true noise variance. This can be explained as follows: when original images $x$ are generated from the true prior $P_{\text{pri}}(x)$ and $Y_K$ are $K$ degraded images with AWGN with $\sigma^2_{\text{true}}$, consider the following log-marginal-likelihood:

$$\ell(\sigma^2) \propto \ln \int P_{\text{noise}}(Y_K \mid x, \sigma^2) P_{\text{pri}}(x) dx,$$

and its statistical average over $Y_K$, i.e.,

$$\mathbb{E}[\ell(\sigma^2)] := \int \int \ell(\sigma^2) P_{\text{noise}}(Y_K \mid x, \sigma^2_{\text{true}}) P_{\text{pri}}(x) dY_K dx.$$

The maximization of $\mathbb{E}[\ell(\sigma^2)]$ with respect to $\sigma^2$ leads to $\sigma^2 = \sigma^2_{\text{true}}$, which can be immediately obtained from the principle of cross-entropy maximization. Therefore, a similarity between the true and estimated noise variances can be expected to be an indicator of the quality of the prior.

5.3 Denoising with Multiple Degraded Images

In this section, the dependence of the GMRF and HGMRF methods on $K$ is considered.

In general, the effect of the GMRF and HGMRF priors becomes small as $K$ increases because, for $K \to +\infty$, the MAP solutions do not depend on their priors and converge to $\hat{y}$ (when all other parameters are finite). This can be easily checked by considering $K \to +\infty$ limit in Eqs. (12) and (25). Therefore, it is inferred that the difference between both methods decreases as $K$ increases. This behavior can be justified because the MSE between $\hat{y}$ and the original image $x_{\text{true}}$ can be evaluated based on the law of large numbers, leading to [12]

$$\frac{1}{n}\|\hat{y} - x_{\text{true}}\|_2^2 \approx \frac{\sigma^2}{K},$$

where $\sigma^2$ is the (finite) true noise variance; therefore, in the $K \to +\infty$ limit, the MSE goes to zero, and $\hat{y}$ then becomes the perfect reconstruction of $x_{\text{true}}$. Thus, the average image $\hat{y}$ can be regarded as a restored image when $K$ is large; however, it is known that the quality of restoration of the average image of $K$ degraded images is lower than that of the GMRF method using the same $K$ degraded images [12].

Figure 5 shows the PSNRs of the restored images obtained from both methods for several values of $K$. In these experiments, degraded and restored images were quantized.

The values of $\sigma$-estimates in the GMRF and HGMRF methods for the images in Figs. 2(b) and (c) are listed in Table 8; for the colored image, the average value of $\sigma$ over different channels is listed. The $\sigma$-estimates in the HGMRF method are closer to the true noise level, which supports the validity of the HGMRF prior. For the images in Figs. 2(a) and (d), similar results were obtained.
into 8-bit images. The difference of both the methods monotonically decreases with an increase in $K$, as expected in the above discussion. The PSNRs of the HGMRF method are higher than those of the GMRF method for all values of $K$; however, the differences are extremely small for $K > 2$.

6. Summary and Future Studies

In this study, a hierarchical prior based on a GMRF, called as HGMRF prior, is proposed by introducing a hyperprior for the bias vector, i.e., the hierarchical Bayesian approach. In the proposed prior, the resulting joint and posterior distributions are no longer a log-linear model. Based on the HGMRF prior, an effective algorithm for Bayesian image denoising is proposed; it can complete the overall procedure, including estimation of all parameters in $O(n)$-time. The numerical results described in Sect. 5 support the validation of the proposed method.

The proposed method overcomes the NLM filter in several cases, particularly in high noise-level cases (Sect. 5.1). However, it cannot reach the level of a BM3D filter for which additional improvements are required. For example, a hierarchical treatment for other parameters (i.e., $\lambda$ and $\alpha$) can be considered. In particular, $\alpha$ would be promising. In the present prior (as well as in the previous GMRF prior), the smoothness constraint is uniformly imposed on all neighboring pairs of pixels. However, this constraint is mismatched around the edges. Increasing the flexibility of $\alpha$ based on the hierarchical Bayesian treatment (e.g., based on the line process [1]) might relax this mismatching.

As mentioned in Sect. 1, a GMRF prior can be used for various purposes other than Bayesian image denoising. In most cases, an HGMRF prior can be used instead of a GMRF prior. The application of an HGMRF prior to other problems is an interesting area of future study.

Acknowledgments

This work was supported by JSPS KAKENHI (grant numbers 18K11459, 18H03303, and 21K11778) and JST CREST (grant number JPMJCR1402). We would like to thank S. Ikegaya for the helpful discussions.

References

[1] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intell., vol.PAMI-6, no.6, pp.721–741, 1984.
[2] A. Blake, P. Kohli, and C. Rother, Markov Random Fields for Vision and Image Processing, The MIT Press, 2011.
[3] M. Yasuda, J. Oklubo, and K. Tanaka, “Digital image inpainting based on Markov random field,” Proc. International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, vol.2, pp.747–752, 2005.
[4] S. Kataoka, M. Yasuda, C. Furtlehner, and K. Tanaka, “Traffic data reconstruction based on Markov random field modeling,” Inverse Problems, vol.30, no.2, 025003, 2014.
[5] K. Tanaka, T. Usui, and M. Yasuda, “Statistical inferences by Gaussian Markov random fields on complex networks,” Proc. 2008 International Conference on Computational Intelligence for Modelling Control Automation, pp.214–219, 2008.
[6] O. Banerjee, L.E. Ghaoui, and A. d’Aspremont, “Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data,” Journal of Machine Learning Research, vol.9, pp.485–516, 2008.
[7] T. Kuwatani, K. Nagata, M. Okada, and M. Toriumi, “Markov random field modeling for mapping geofluid distributions from seismic velocity structures,” Earth, Planets and Space, vol.66, Article no.5, pp.1–9, 2014.
[8] T. Kuwatani, K. Nagata, M. Okada, and M. Toriumi, “Markov-random-field modeling for linear seismic tomography,” Physical Review E, vol.90, no.4, 042137, Oct. 2014.
[9] H. Nishimoto, Bussei Kenkyu (in Japanese), vol.73, no.5, pp.850–857, 2000.
[10] K. Tanaka, “Statistical-mechanical approach to image processing,” Journal of Physics A: Mathematical and General, vol.35, no.37, pp.R81–R150, 2002.
[11] N. Morin, “Study on probabilistic image processing by extended Gauss-Markov random fields,” Master Thesis in Tohoku University (unpublished), 2010.
[12] M. Yasuda, J. Watanabe, S. Katoaka, and K. Tanaka, “Linear-time algorithm in Bayesian image denoising based on Gaussian Markov random field,” IEICE Trans. Inf. & Syst., vol.E101-D, no.6, pp.1629–1639, June 2018.
[13] A. Buades, B. Coll, and J.-M. Morel, “A non-local algorithm for image denoising,” 2005 IEEE Computer Vision and Pattern Recognition, vol.2, pp.60–65, 2005.
[14] K. Dobov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” IEEE Trans. Image Process., vol.16, no.8, pp.2080–2095, 2007.
[15] G. Fracastoro, S.M. Fossom, and E. Magli, “Steerable discrete cosine transform,” IEEE Trans. Image Process., vol.26, no.1, pp.303–314, 2017.
[16] S. Katakami, H. Sakamoto, S. Murata, and M. Okada, “Gaussian Markov random field model without boundary conditions,” Journal of the Physical Society of Japan, vol.86, no.6, 064801, 2017.
[17] A. Pothen, H.D. Simon, and K.-P. Liou, “Partitioning sparse matrices with eigenvectors of graphs,” SIAM Journal on Matrix Analysis and Applications, vol.11, no.3, pp.430–452, 1990.
[18] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Trans. Image Process., vol.13, no.4, pp.600–612, 2004.
[19] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, and L. Shao, “Multi-stage progressive image restoration,” Proc. Conference on Computer Vision and Pattern Recognition, pp.14816–14826, 2021.
[20] S. Kataoka and M. Yasuda, “Bayesian image denoising with multiple noisy images,” The Review of Socionetwork Strategies, vol.13, no.2, pp.267–280, 2019.
[21] Y. Weiss and W.T. Freeman, “Correctness of belief propagation in Gaussian graphical models of arbitrary topology,” Neural Computation, vol.13, no.10, pp.2173–2200, 2001.
[22] O. Shental, P.H. Siegel, J.K. Wolf, D. Bickson, and D. Dolev, “Gaussian belief propagation solver for systems of linear equations,” Proc. IEEE International Symposium on Information Theory, pp.1863–1867, 2008.
[23] C.M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
[24] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.5, pp.898–916, 2011.
Appendix A: Minimization of Mean-Squared Error in Eq. (14)

In this appendix, the minimization of

\[ E(\mathbf{b}) := \int \|\mathbf{x}_{\text{true}} - \mathbf{m}\|^2 \rho_{\text{noise}}(Y_K | \mathbf{x}_{\text{true}}, \sigma^2) dY_K \]

(i.e., Eq. (14)) with respect to \( \mathbf{b} \) is considered for a fixed \( \lambda \), \( \alpha \), and \( \sigma^2 \). Equation (11) leads to

\[
E(\mathbf{b}) = -2(\mathbf{x}_{\text{true}})^T S^{-1}_{\text{post}} \int \mathbf{c}_{\text{noise}}(Y_K | \mathbf{x}_{\text{true}}, \sigma^2) dY_K \\
+ \int \mathbf{c}^T S^{-2}_{\text{post}} \mathbf{c}_{\text{noise}}(Y_K | \mathbf{x}_{\text{true}}, \sigma^2) dY_K + \text{constant.}
\]

(A-1)

Here, from Eq. (8),

\[
\int \mathbf{c}_{\text{noise}}(Y_K | \mathbf{x}_{\text{true}}, \sigma^2) dY_K = \frac{K}{\sigma^2} \mathbf{x}_{\text{true}} + \mathbf{b}
\]

and

\[
\int \mathbf{c}^T S^{-2}_{\text{post}} \mathbf{c}_{\text{noise}}(Y_K | \mathbf{x}_{\text{true}}, \sigma^2) dY_K \\
= \frac{K^2}{\sigma^4} \int (\hat{\mathbf{y}})^T S^{-1}_{\text{post}} \hat{\mathbf{y}} \rho_{\text{noise}}(Y_K | \mathbf{x}_{\text{true}}, \sigma^2) dY_K \\
+ \frac{2K}{\sigma^2} \mathbf{b}^T S^{-1}_{\text{post}} \mathbf{x}_{\text{true}} + \mathbf{b}^T S^{-1}_{\text{post}} \mathbf{b}
\]

are obtained. Using these equations, Eq. (A-1) is expressed as

\[
E(\mathbf{b}) = -2(\mathbf{x}_{\text{true}})^T S^{-1}_{\text{post}} (\mathbf{I}_n - \frac{K}{\sigma^2} S^{-1}_{\text{post}}) \mathbf{b} + \mathbf{b}^T S^{-2}_{\text{post}} \mathbf{b} \\
+ \text{constant.}
\]

(A-2)

The minimal condition,

\[
\frac{\partial E(\mathbf{b})}{\partial \mathbf{b}} = -2 S^{-1}_{\text{post}} (\mathbf{I}_n - \frac{K}{\sigma^2} S^{-1}_{\text{post}}) \mathbf{x}_{\text{true}} + 2 S^{-2}_{\text{post}} \mathbf{b} = \mathbf{0},
\]

and the relation of Eq. (9) lead to Eq. (15).

Appendix B: Computation of Gradients

In this appendix, the gradients of \( f(h_{\text{post}}) \) in Eq. (34) with respect to the parameters in \( h_{\text{post}} \) are considered.

In this derivation, the differentiation of the second term is rather complicated, whereas the differentiations of the other terms are straightforward. Here, the differentiation of \( f(h_{\text{post}}) := \mathbf{g}^T \mathbf{u} \) is therefore considered. For \( h \in h_{\text{post}} \), because \( \mathbf{u} = H_{\text{post}} \mathbf{g} \),

\[
\frac{\partial f(h_{\text{post}})}{\partial h} = \left( \frac{\partial g}{\partial h} \right)^T \mathbf{u} - \mathbf{g}^T H_{\text{post}}^{-1} \frac{\partial H_{\text{post}}}{\partial h} H_{\text{post}}^{-1} \mathbf{g} \\
+ \mathbf{g}^T H_{\text{post}}^{-1} \left( \frac{\partial g}{\partial h} \right)
\]

is obtained, where \( \gamma^2 \mathbf{v} = \Sigma^{-1} \mathbf{u} \) is used (Eq. (24)). Finally,

\[
\mathbf{v}^T \frac{\partial \Sigma}{\partial h} \mathbf{v} = \left( \frac{\partial \gamma^2}{\partial h} \right) \mathbf{v}^T \mathbf{v} - \mathbf{w}^T \frac{\partial \Sigma_{\text{pri}}}{\partial h} \mathbf{w}
\]

is obtained, where \( \mathbf{w} \) is defined in Eq. (39). From Eqs. (A-3)–(A-5), the following differentiations are obtained:

\[
\frac{\partial f(h_{\text{post}})}{\partial \lambda} = -\mathbf{u}^T \mathbf{u} + \gamma^4 \mathbf{w}^T \mathbf{w},
\]

\[
\frac{\partial f(h_{\text{post}})}{\partial \alpha} = -\mathbf{u}^T \mathbf{A} \mathbf{u} + \gamma^4 \mathbf{w}^T \mathbf{A} \mathbf{w},
\]

\[
\frac{\partial f(h_{\text{post}})}{\partial \gamma^2} = \mathbf{v}^T \mathbf{v},
\]

\[
\frac{\partial f(h_{\text{post}})}{\partial \sigma^2} = -2 \frac{K}{\sigma^2} \mathbf{v}^T \mathbf{u} + \frac{K}{\sigma^2} \mathbf{u}^T \mathbf{u}.
\]

Using these differentiations, the gradients in Eqs. (35)–(38) are immediately obtained.

Yuki Monma received a bachelor’s degree (in engineering) from Yamagata University in 2021, and is currently pursing a Master’s degree from the same university. Her research interests are statistical machine learning and signal/image processing.

Kan Aro received a bachelor’s degree (in engineering) from Yamagata University in 2021. He now works in Diamond Head Co., Ltd. His research interests are statistical machine learning and signal/image processing.
Muneki Yasuda  received a doctoral degree (in information science) from Tohoku University in 2007. He worked at Tohoku University from 2007 to 2013 as an Assistant Professor. From 2013 to 2021, he was an Associate Professor at Yamagata University, and is currently a professor at the same university. His research interests include statistical machine learning, data mining, AI, signal processing, and statistical mechanics.