Computational Intelligence Seminar (8 November, 2005, Waseda University, Tokyo, Japan)

Probabilistic image processing and Bayesian network

Kazuyuki Tanaka [1]

Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan

Abstract

The basic frameworks and practical schemes of the Bayesian network and the belief propagation to the probabilistic image processing are reviewed. The probabilistic image processing is formulated by means of Bayesian statistics and Markov random fields. The system is regarded as one of Bayesian networks. In general, the Bayesian network has serious computational complexity because the probabilistic models include a number of random variables for nodes or pixels. Recently, many researchers in the intermediate region of the mathematical sciences and the computer sciences are interested in the belief propagations, which is one of powerful approximate methods for probabilistic inference. In the present tutorial lecture note, we briefly explain the formulation of the probabilistic image processing and the theoretical structure of the belief propagation. As an examples of the applications of the probabilistic image processing, we review a noise reduction by means of Gaussian graphical model as Markov random fields.

1 Introduction

Probabilistic information processing based on statistical sciences and statistical mechanics is applicable to computer sciences. Conversely, many problems in probabilistic information processing have created research ideas in statistical science and statistical mechanics. Many schemes in probabilistic information processing is based on the Bayesian statistics and are reduced to probabilistic models. The probabilistic models can be represented by graphs including nodes and directed line segments and are sometimes called *graphical models*. The system which consists of such a graphical model is called Bayesian network.

Probabilistic image processing based on the Bayes' formula is one of the useful applications in the Bayesian network. By using the Bayes' formula and assuming an *a priori* probability for the original image, one creates probabilistic image processing in the form of an *a posteriori* probability for the original image when the observed image is given[1]. Image processing by means of probabilistic models and the Bayes statistics are usually based on Markov random fields. In the Markov random fields, the state of a pixel is dependent only on the states of its neighbouring pixels. Many kinds of Markov random field models which are applicable to the practical image processing, image restorations, segmentations, edge detection, image compressions and motion detections, were proposed[2, 3, 4]. Comparing the conventional techniques in the image processing, the probabilistic image processing is expected to give good performance even for large noise and to construct the robust systems for various data.

As for most of familiar Markov random fields as well as Bayesian networks, it is hard to achieve the criteria to obtain the estimated image in the statistical frameworks exactly except some special cases[5]. In the artificial intelligence, the belief propagation was proposed in the middle of 1980s[6],

[1]E-mail: kazu@smapip.is.tohoku.ac.jp, Webpage: http://www.smapip.is.tohoku.ac.jp/~kazu/

as an efficient algorithm of statistical inference, which aims at implementing probability-based artificial intelligence capable of handling uncertainties. The belief propagation is proved to give exact results in the special cases where the graphical representation of the underlying probability model has no loops in its graphical representation. Although the belief propagation is not guaranteed to give exact results for general graphical models with loops, it works reasonably well in various cases, and furthermore, in some specific cases it turns out to work excellently. In response to such empirically demonstrated good performance in the belief propagation, a number of researchers in the mathematical science have attempt to achieve the theoretical understanding of the performance and the frameworks in the algorithm[7, 8]. It is worthy of notice that the belief propagation has a link with statistical mechanics[9, 10, 11] and has been equivalent to a Bethe approximation and a cluster variation method, which are ones of the advanced mean-field methods in statistical mechanics. Moreover, some interpretations for the belief propagation have been given also from the information geometrical point of view[12, 13]. The belief propagation has been applied to many practical problems in the computer vision[14, 15, 16, 17]. Theoretical study of the application of the belief propagation to image processing has been done in the statistical mechanical point of view[18, 19].

In this paper, the basic frameworks and its applications of the Bayesian network and the belief propagation in the probabilistic image processing are reviewed. As examples of the probabilistic image processing, we introduce a noise reduction and a segmentation and extend the segmentation to a motion detection. Section 2 describes the formulation of the probabilistic image processing as Bayesian networks. In Section 3, we explain the belief propagation which is reduced to the numerical iterative calculations of the simultaneous fixed point equations with respect to messages. In Section 4, we explain the probabilistic image restorations and noise reductions by using the Gaussian graphical model. Section 5 provides concluding remarks.

2 Formulation of image processing by Bayesian network

We consider an image on a square lattice $\Omega \equiv \{1, 2, \cdots, L\}$. The states at pixel $i (\in \Omega)$ in the original image and the observed image are regarded as random variables denoted by A_i and D_i, respectively. In the noise reduction, the observed image corresponds to a degraded image and the original image corresponds to an natural image before degraded by noise. In the segmentation, a given natural image is regarded as the observed image and a segmented image with a few kinds of regions is regarded as the original image. Then the random fields of states in the original image and the observed image are represented by $\vec{A} = (A_1, A_2, \cdots, A_L)$ and $\vec{D} = (D_1, D_2, \cdots, D_L)$. The actual original image and the observed image are denoted by $\vec{a} = (a_1, a_2, \cdots, a_L)$ and $\vec{d} = (d_1, d_2, \cdots, d_L)$, respectively. The probability that the original image is $\vec{a}$, $\mathcal{P}(\vec{A} = \vec{a}|\vec{\alpha})$, is called the *a priori* probability of the image. Here, $\vec{\alpha}$ is not a set of random variables but a set of *hyperparameters*, which specify the function to represent the *a priori* probability.

In the Bayes formula, the *a posteriori* probability $\mathcal{P}(\vec{A} = \vec{a}|\vec{D} = \vec{d})$, that the original image is $\vec{a}$ when

the given observed image is $\vec{d}$, is expressed as

$$\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{d},\vec{\alpha},\vec{\beta})=\frac{\mathcal{P}(\vec{D}=\vec{d}\big|\vec{A}=\vec{a},\vec{\beta})\mathcal{P}(\vec{A}=\vec{a},\vec{\alpha})}{\sum_{\vec{z}}\mathcal{P}(\vec{D}=\vec{d}|\vec{A}=\vec{z},\vec{\beta})\mathcal{P}(\vec{A}=\vec{z},\vec{\alpha})}, \tag{1}$$

where the summation $\sum_{\vec{z}}=\sum_{z_1}\sum_{z_2}\cdots\sum_{z_L}$ is taken over all possible configurations of images $\vec{z}=(z_1,z_2,\cdots,z_L)$. If z_i takes any real number in the interval $(-\infty,+\infty)$, where the summation $\sum_{\vec{z}}[\cdots]$ should be replaced by the integral $\int d\vec{z}\equiv\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}[\cdots]dz_1dz_2\cdots dz_L$ and the probability should be regarded as a probability density function. The probability $\mathcal{P}\big(\vec{D}=\vec{d}\big|\vec{A}=\vec{a},\vec{\beta}\big)$ is the conditional probability that the observed image is $\vec{d}$ when the original image is $\vec{a}$ and describes the generating process of data. Here $\vec{\beta}$ is a set of hyperparameters in the conditional probability. In the Bayesian statistics, the estimate $\widehat{a}_i$ of the state at each pixel i in the original image are determined so as to maximize the posterior marginal probability:

$$\mathcal{P}(A_i=a_i|\vec{D}=\vec{d},\vec{\alpha},\vec{\beta})=\sum_{\vec{z}}\delta_{a_i,z_i}\mathcal{P}(\vec{A}=\vec{z}|\vec{D}=\vec{d},\vec{\alpha},\vec{\beta}). \tag{2}$$

In the present framework, we have to estimate not only the estimate $(\widehat{a}_1,\widehat{a}_2,\cdots,\widehat{a}_L)$ but also the hyperparameters $\vec{\alpha}$ and $\vec{\beta}$. We apply the maximum likelihood estimation to the estimation of hyperparameters from an observed image, which corresponds to data, in the statistical sciences as follows:

$$(\widehat{\vec{\alpha}},\widehat{\vec{\beta}})=\max_{(\vec{\alpha},\vec{\beta})}\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta}) \tag{3}$$

$$\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})=\sum_{\vec{z}}\mathcal{P}(\vec{A}=\vec{z},\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})=\sum_{\vec{z}}\mathcal{P}(\vec{D}=\vec{d}|\vec{A}=\vec{z},\vec{\beta})\mathcal{P}(\vec{A}=\vec{z},\vec{\alpha}). \tag{4}$$

In this framework, the probability $\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})$ is given by marginalizing the joint probability $\mathcal{P}(\vec{A}=\vec{a},\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})$ over all the possible images $\vec{a}$ and can be regarded as a likelihood for $\vec{\alpha}$ and $\vec{\beta}$ when the observed image $\vec{d}$ is given in statistics. $\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})$ is referred to as *evidence* or *marginal likelihood*[21, 22]. In order to maximize the marginal likelihood, the expectation maximization (EM) algorithm is often employed. In the EM algorithm, we introduce a Q-function defined by

$$Q(\vec{\alpha}',\vec{\beta}'|\vec{\alpha},\vec{\beta},\vec{d})\equiv\sum_{\vec{z}}\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{d},\vec{\alpha},\vec{\beta})\ln\mathcal{P}(\vec{A}=\vec{a},\vec{D}=\vec{d}|\vec{\alpha}',\vec{\beta}'). \tag{5}$$

The extremum conditions of marginal likelihood

$$\frac{\partial}{\partial\vec{\alpha}}\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})=\vec{0},\qquad\frac{\partial}{\partial\vec{\beta}}\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})=\vec{0}, \tag{6}$$

are equivalent to the following equalities:

$$\Big[\frac{\partial}{\partial\vec{\alpha}'}Q(\vec{\alpha}',\vec{\beta}'|\vec{\alpha},\vec{\beta},\vec{d})\Big]_{\vec{\alpha}'=\vec{\alpha},\vec{\beta}'=\vec{\beta}}=\vec{0},\qquad\Big[\frac{\partial}{\partial\vec{\beta}'}\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})\Big]_{\vec{\alpha}'=\vec{\alpha},\vec{\beta}'=\vec{\beta}}=\vec{0}, \tag{7}$$

respectively. As an algorithm to calculate the set of estimates, $(\widehat{\vec{\alpha}},\widehat{\vec{\beta}})$, which satisfies the extremum conditions (6), we iterate the following EM-steps until convergence:

E-Step: $Q(\vec{\alpha},\vec{\beta}|\vec{\alpha}(t),\vec{\beta}(t),\vec{g})\leftarrow\sum_{\vec{z}}\mathcal{P}(\vec{A}=\vec{z}|\vec{D}=\vec{d},\vec{\alpha}(t),\vec{\beta}(t))\ln\mathcal{P}(\vec{A}=\vec{z},\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})$.

M-Step: $(\vec{\alpha}(t+1),\vec{\beta}(t+1))\leftarrow\underset{(\vec{\alpha},\vec{\beta})}{\mathrm{argmax}}Q(\vec{\alpha},\vec{\beta}|\vec{\alpha}(t),\vec{\beta}(t),\vec{g})$.

Though the above scheme can give us the solution for the extremum conditions (6), we have to remark that it does not necessarily provide the global maximum for the marginal likelihood $\mathcal{P}(\vec{D}=\vec{d}|\vec{\alpha},\vec{\beta})$.

We denote a set of links between some pairs of pixels by $\mathcal{N}$ and assign a function $\Phi_{ij}(a_i,a_j)$ at each pair of links ij belonging to the set $\mathcal{N}$. The *a posteriori* probabilities $\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{d},\vec{\alpha},\vec{\beta})$ and the *a priori* probabilities, $\mathcal{P}(\vec{A}=\vec{a}|\vec{\alpha})$ in the Bayesian image analysis, are often expressed in terms of the following function $P(\vec{a})$:

$$P(\vec{a})=\frac{\prod_{ij\in\mathcal{N}}\Phi_{ij}(a_i,a_j)}{\sum_{\vec{z}}\prod_{ij\in\mathcal{N}}\Phi_{ij}(z_i,z_j)}. \tag{8}$$

In the case, the random field $\vec{A}$ is the set of random variables in which the state of each pixel i depends only on the configuration of all the pixels linked to the pixel i, $\mathcal{N}_i\equiv\{j|ij\in\mathcal{N}\}$ and is called the Markov random field[1].

3 Belief propagation

Except some special cases, it is hard to calculate the marginalization in Eqs.(2) and (4) exactly, and we have to employ an approximate algorithm. In the present section, we use the belief propagation as one of approximate algorithms to calculate approximate values of statistical quantities numerically.

For a probability given in Eq.(8), the approximate form of marginal probabilities in the belief propagation is assumed to be $P(a_i|\vec{D}=\vec{d})$ to the posterior marginal probability:

$$P_i(a_i)\equiv\sum_{\vec{z}}\delta_{z_i,a_i}P(\vec{a})\simeq\frac{\prod_{k\in\mathcal{N}_i}\mathcal{M}_{k\to i}(a_i)}{\sum_{z_i}\prod_{k\in\mathcal{N}_i}\mathcal{M}_{k\to i}(z_i)}, \tag{9}$$

$$P_{ij}(a_i,a_j)\equiv\sum_{\vec{z}}\delta_{z_i,a_i}\delta_{z_j,a_j}P(\vec{a})\simeq\frac{\big(\prod_{k\in\mathcal{N}_i}\mathcal{M}_{k\to i}(a_i)\big)\Phi_{ij}(a_i,a_j)\big(\prod_{l\in\mathcal{N}_j}\mathcal{M}_{l\to j}(a_j)\big)}{\sum_{z_i}\sum_{z_j}\big(\prod_{k\in\mathcal{N}_i}\mathcal{M}_{k\to i}(z_i)\big)\Phi_{ij}(z_i,z_j)\big(\prod_{l\in\mathcal{N}_j}\mathcal{M}_{l\to j}(z_j)\big)}. \tag{10}$$

Here the notation $\mathcal{N}_i\equiv\{j|ij\in\mathcal{N}\}$ represents the set of all the nearest-neighbor pairs of pixel i. The quantity $\mathcal{M}_{j\to i}(a_i)$ in Eqs.(9) and (10) is called a message propagated from j to i Both quantities $\mathcal{M}_{j\to i}(a_i)$ and $\mathcal{M}_{i\to j}(a_j)$ are assigned at each link ij and are determined so as to satisfy the following simultaneous fixed point equations:

$$\mathcal{M}_{j\to i}(a_i)=\frac{\sum_{z_j}\Phi_{ij}(z_i,a_j)\prod_{k\in\mathcal{N}_j\setminus\{i\}}\mathcal{M}_{k\to j}(z_j)}{\sum_{z_i}\sum_{z_j}\Phi_{ij}(z_i,z_j)\prod_{k\in\mathcal{N}_j\setminus\{i\}}\mathcal{M}_{k\to j}(z_j)}\ (i\in\Omega\ j\in\mathcal{N}_i). \tag{11}$$

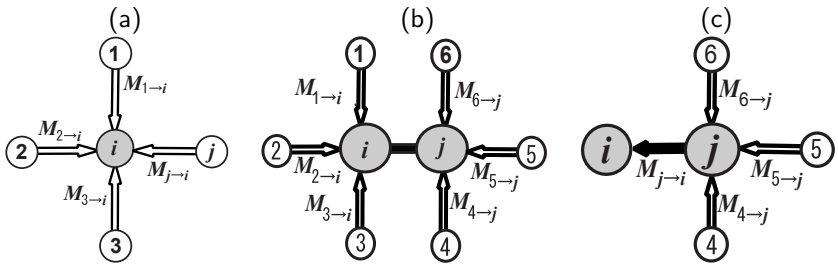

Figure 1: Graphical representations of the belief propagation. (a) Eq.(9). (b) Eq.(10). (c) Eq.(11).

The graphical representations of Eqs.(9)-(11) are shown in Figure 1.

The simultaneous fixed point equations (11) are derived by substituting Eqs.(9) and (10) to the reducibility conditions:

$$P_i(a_i) = \sum_{z_j} P_{ij}(a_i, z_j) \ \ (i\in\Omega,\ \ j\in\mathcal{N}_i). \tag{12}$$

We can calculate the numerical solutions of the simultaneous fixed point equations (11), numerically, by using the iteration method. By applying the numerical solutions to Eqs.(9) and (10), we obtain the marginal probability $P_i(a_i)$ and $P_{ij}(a_i, a_j)$.

The approximate expressions of marginal probabilities in Eqs.(9) and (10) can be obtained by means of Bethe approximation in the statistical mechanics. In Bethe approximation, we introduce an approximate free energy for every probabilistic model in Eq.(8) which is defined by

$$\mathcal{F}_{\rm Bethe}[\{\rho_i, \rho_{ij} | i\in\Omega, ij\in\mathcal{N}\}] \equiv \sum_{i\in\Omega} \mathcal{F}_i[\rho_i] + \sum_{ij\in\mathcal{N}} \Big(\mathcal{F}_{ij}[\rho_{ij}] - \mathcal{F}_i[\rho_i] - \mathcal{F}_j[\rho_j] \Big), \tag{13}$$

where

$$\mathcal{F}_i[\rho_i] \equiv \sum_{z_i} \rho_i(\zeta) \ln\big(\rho_i(z_i)\big), \ \ \mathcal{F}_{ij}[\rho_{ij}] \equiv \sum_{z_i}\sum_{z_j} \rho_{ij}(\zeta,\zeta') \ln\Big(\frac{\rho_{ij}(z_i,z_j)}{\Phi_{ij}(z_i,z_j)}\Big). \tag{14}$$

Here $\mathcal{F}_{\rm Bethe}[\{\rho_i, \rho_{ij} | i\in\Omega, ij\in B\}]$ is often referred to as *Bethe free energy* and is regarded as a good approximation for the true free energy of probabilistic model in Eq.(8) which is defined by $\mathcal{F}[\rho] = -\ln\Big(\sum_{\vec{z}} \prod_{ij\in\mathcal{N}} \Phi_{ij}(z_i, z_j)\Big)$ in the statistical mechanics. The simultaneous fixed-point equations in Eqs.(9)-(11) are equivalent to the extremum conditions of $\mathcal{F}_{\rm Bethe}[\{\rho_i, \rho_{ij} | i\in\Omega, ij\in B\}]$ with respect to the marginal probability distributions $\{\rho_i, \rho_{ij} | i\in\Omega, ij\in B\}$ under the normalization conditions, $\sum_{z_i} \rho_i(z_i) = 1$ $(i\in\Omega)$ and $\sum_{z_i}\sum_{z_j} \rho_{ij}(z_i, z_j) = 1$ $(ij\in\mathcal{N})$, and the reducibility conditions (12). However, it is known that the Bethe free energy $\mathcal{F}_{\rm Bethe}[\{\rho_i, \rho_{ij} | i\in\Omega, ij\in B\}]$ does not provide any bounds for the true free energy $\mathcal{F}[\rho]$, while a mean-field free energy is a bound for the true free energy[10]. Furthermore, in some cases the solution of the simultaneous fixed point equations (9)-(11) corresponds not to a local minimum but to a saddle point of the Bethe free energy[20].

4 Noise Reduction by Gaussian Graphical Model

As an example for the applications of image processing, we show the noise reduction from the degraded

image which is corrupted by the additive white Gaussian noise with the mean 0 and the variation σ^2. In this case, the random fields of the original image and of the degraded image are denoted by $\vec{A}$ and $\vec{D}$, respectively. The degradation process is expressed in terms of the following conditional probability:

$$\mathcal{P}(\vec{D}=\vec{d}|\vec{A}=\vec{a},\beta=\sigma^{-2})=\prod_{i\in\Omega}\frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{1}{2\sigma^2}(a_i-d_i)^2\Big), \tag{15}$$

$$\mathcal{P}(\vec{A}=\vec{a}|\alpha)=\frac{\displaystyle\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(a_i-a_j)^2\Big)}{\displaystyle\int\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(a_i-a_j)^2\Big)d\vec{z}}\ \ (\alpha>0). \tag{16}$$

By substituting Eqs.(15) and (16) to the Bayes formula, the *a posteriori* probability is expressed as follows:

$$\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{a},\alpha,\beta=\sigma^{-2})=\frac{\Big(\displaystyle\prod_{i\in\Omega}\exp\big(-\tfrac{1}{2\sigma^2}(a_i-d_i)^2\big)\Big)\Big(\displaystyle\prod_{ij\in\mathcal{N}}\exp\big(-\tfrac{1}{2}\alpha(a_i-a_j)^2\big)\Big)}{\displaystyle\int\Big(\prod_{i\in\Omega}\exp\big(-\tfrac{1}{2\sigma^2}(z_i-d_i)^2\big)\Big)\Big(\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(z_i-z_j)^2\Big)\Big)d\vec{z}}. \tag{17}$$

By setting

$$\Phi_{ij}(a_i,a_j)\equiv\exp\Big(-\frac{1}{8\sigma^2}(a_i-d_i)^2-\frac{1}{8\sigma^2}(a_j-d_j)^2-\frac{1}{2}\alpha(a_i-a_j)^2\Big), \tag{18}$$

Eq.(17) can be reduced to

$$\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{a},\alpha,\beta=\sigma^{-2})=\frac{\displaystyle\prod_{ij\in\mathcal{N}}\Phi_{ij}(a_i,a_j)}{\displaystyle\int\Big(\prod_{ij\in\mathcal{N}}\Phi_{ij}(z_i,z_j)\Big)d\vec{z}}. \tag{19}$$

By applying the belief propagation to these expressions of the probabilities, we can achieve the maximization of the marginal likelihood $\mathcal{P}(\vec{D}=\vec{d}|\alpha,\beta=\sigma^{-2})\equiv\int\mathcal{P}(\vec{D}=\vec{d}|\vec{A}=\vec{z},\beta=\sigma^{-2})\mathcal{P}(\vec{A}=\vec{z}|\alpha)d\vec{z}$ with respect to the hyperparameters (α,σ), and can calculate the posterior marginal probability $\mathcal{P}(A_i=a_i|\vec{D}=\vec{d},\alpha,\beta=\sigma^{-2})$[19]. Now we assume that $\mathcal{M}_{j\to i}(\xi)$ can be approximately expressed as

$$\mathcal{M}_{j\to i}(\xi)\simeq\sqrt{\frac{\lambda_{j\to i}}{2\pi}}\exp\Big(-\frac{\lambda_{j\to i}}{2}(\xi-\mu_{j\to i})^2\Big). \tag{20}$$

The simultaneous fixed-point equation (11) can be reduced to the following fixed point equations for $\{\lambda_{j\to i},\lambda_{i\to j},\mu_{j\to i},\mu_{i\to j}|j\in\mathcal{N}_i,\ i\in\Omega\}$:

$$\frac{1}{\lambda_{j\to i}}=\frac{1}{\alpha}+\frac{1}{\beta+\displaystyle\sum_{k\in\mathcal{N}_j\setminus i}\lambda_{k\to j}},\ \ \mu_{j\to i}=\frac{\beta g_j+\displaystyle\sum_{k\in\mathcal{N}_j\setminus i}\mu_{k\to j}\lambda_{k\to j}}{\beta+\displaystyle\sum_{k\in\mathcal{N}_j\setminus i}\lambda_{k\to j}}\ \ (i\in\Omega\ j\in\mathcal{N}_i). \tag{21}$$

If we substitute Eq. (20) into Eqs. (9) and (10), the one- and two-body marginal probability densities $\rho_i(a_i)$ and $\rho_{ij}(a_i,a_j)$ are approximately obtained.

The simultaneous fixed-point equations (21) are solved by the following iterative algorithm:

Algorithm LBP$[\vec{g}, \alpha, \beta]$

Step 1: Set $r \leftarrow 0$ as an initial value.

Step 2: Update $r \leftarrow r+1$ and

$$a_{j\to i}(r) \leftarrow \left(\frac{1}{\alpha} + \frac{1}{\beta + \sum_{k \in \mathcal{N}_j \backslash i} a_{k\to j}(r-1)} \right)^{-1} \quad (i \in \Omega,\ j \in \mathcal{N}_i). \tag{22}$$

$$b_{j\to i}(r) \leftarrow \frac{\beta g_j + \sum_{k \in \mathcal{N}_j \backslash i} b_{k\to j}(r-1) a_{k\to j}(r-1)}{\beta + \sum_{k \in \mathcal{N}_j \backslash i} a_{k\to j}(r-1)} \quad (i \in \Omega,\ j \in \mathcal{N}_i). \tag{23}$$

Step 3: Update $R \leftarrow r$, $\lambda_{j\to i} \leftarrow a_{j\to i}(r)$ and $\mu_{j\to i} \leftarrow b_{j\to i}(r)$ $(i \in \Omega)$. Go to **Step 4** if, for prespecified ε,

$$\sum_{i \in \Omega} \sum_{j \in \mathcal{N}_i} \left(\left| a_{j\to i}(r) - a_{j\to i}(r-1) \right| + \left| b_{j\to i}(r) - b_{j\to i}(r-1) \right| \right) < \varepsilon, \tag{24}$$

and go to **Step 2** otherwise.

Step 4: Substitute $\{\lambda_{j\to i}, \mu_{j\to i} | j \in \mathcal{N}_i,\ i \in \Omega\}$ into Eq. (20), and calculate $\rho_i(\xi | \vec{g}, \alpha, \beta)$ and $\rho_{ij}(\xi, \xi' | \vec{g}, \alpha, \beta)$ by using Eqs.(9) and (10). Stop after $\rho_i(\xi | \vec{g}, \alpha, \beta)$ and $\rho_{ij}(\xi, \xi' | \vec{g}, \alpha, \beta)$ are set as outputs in the present algorithm.

Again, it is usually adequate to set $\varepsilon = 10^{-6}$. In the denominators of Eqs.(22) and (23), the summations $\sum_{k \in \mathcal{N}_j \backslash i} a_{k\to j}(r-1)$ and $\sum_{k \in \mathcal{N}_j \backslash i} b_{k\to j}(r-1) a_{k\to j}(r-1)$ can be evaluated in $\mathcal{O}(1)$ time per pair of pixels i and j, because the number of elements in the set $\mathcal{N}_j \backslash i$ is equal to 3 per pair of pixels. Hence the iterative algorithm for solving the simultaneous fixed-point equations (21) requires a total of $\mathcal{O}(|\Omega|)$ computations per update.

For fixed values of α' and σ', the extremum conditions of $Q(\alpha', \sigma' | \alpha, \sigma, \vec{g})$ with respect to σ and α are reduced to the following equations:

$$\sum_{i \in \Omega} \int_{-\infty}^{+\infty} (\xi - g_i)^2 \rho_i(\xi | \vec{g}, \alpha, \sigma^{-2}) d\xi = |\Omega| \sigma'^2, \tag{25}$$

$$\sum_{ij \in \mathcal{N}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi - \xi')^2 \rho_{ij}(\xi, \xi' | \vec{g}, \alpha, \sigma^{-2}) d\xi d\xi' = \sum_{ij \in \mathcal{N}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi - \xi')^2 \rho_{ij}(\xi, \xi' | \vec{g}, \alpha', 0) d\xi d\xi', \tag{26}$$

respectively. The marginal probability density function $\rho_i(\xi | \vec{g}, \alpha, \sigma^{-2})$, $\rho_{ij}(\xi, \xi' | \vec{g}, \alpha, \sigma^{-2})$ and $\rho_{ij}(\xi, \xi' | \vec{g}, \alpha', 0)$ are approximately obtained as outputs of the loopy belief propagation algorithms LBP$[\vec{g}, \alpha, \sigma^{-2}]$ and LBP$[\vec{g}, \alpha', 0]$. Therefore, we can give the EM algorithm for the maximization of marginal likelihood in Eq.(3) by using the loopy belief propagation as follows:

EM Algorithm in Loopy Belief Propagation

Step 1: Set $\big(\alpha(0), \sigma(0)\big)$ and $t \leftarrow 0$.

Step 2: Run the algorithm LBP$[\vec{g}, \alpha(t), \sigma(t)^{-2}]$ and update $\sigma(t)$ and A as follows:

$$\sigma(t+1) \leftarrow \frac{1}{|\Omega|} \sum_{i \in \Omega} \int_{-\infty}^{+\infty} (\xi - g_i)^2 \rho_i(\xi | \vec{g}, \alpha(t), \sigma(t)^{-2}) d\xi, \tag{27}$$

$$A \leftarrow \sum_{ij \in \mathcal{N}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi - \xi')^2 \rho_{ij}(\xi, \xi' | \vec{g}, \alpha(t), \sigma(t)^{-2}) d\xi d\xi'. \tag{28}$$

Step 3: Run the algorithm LBP$[\vec{g}, \alpha, 0]$ for various positive values of α and set $\alpha(t+1)$ to the value of α which satisfies the follows equation:

$$\sum_{ij \in \mathcal{N}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\xi - \xi')^2 \rho_{ij}(\xi, \xi' | \vec{g}, \alpha, 0) d\xi d\xi' = A, \tag{29}$$

and $t \leftarrow t+1$.

Step 4: Update

$$\big(\widehat{\alpha}, \widehat{\sigma}\big) \leftarrow \big(\alpha(t), \sigma(t)\big). \tag{30}$$

Stop if the values of $\widehat{\alpha}$ and $\widehat{\sigma}$ converge, and return to **Step 2** otherwise.

Because both probabilities in Eqs.(16) and (17) are multi-dimensional Gaussian distribution, we can calculate the some statistical quantities and the marginal likelihood, exactly, by means of the multi-dimensional Gaussian integral formula and the discrete Fourier transformation[22]. Hence we can check the accuracy of the belief propagation for the probabilistic image processing based on the Bayesian statistics and the maximum likelihood estimation. We show in Figure 2 and Table 1 results of numerical experiments of noise reduction from the degraded image which is generated by the additive white Gaussian noise with the mean 0 and the variation 40^2. The estimates of $\widehat{\alpha}$ and $\widehat{\sigma}$ given in Table 1 are obtained by using the EM algorithm. The process for the convergence of $(\alpha(t), \sigma(t))$ $(\sigma(t) \equiv \sqrt{\beta(t)})$ in the EM algorithm is shown in Figure 3. The open and the solid circles are corresponding to the results in the EM algorithm for the loopy belief propagation and the exact calculation, respectively.

Figure 2: Noise reduction by using the Gaussian graphical model and the belief propagation. (a) Original image. (b) Degraded image by the additive white Gaussian noise with the mean 0 and the variance 40^2. (c) Restored image in the loopy belief propagation. (d) Restored image in the generalized belief propagation. (e) Restored image in the exact calculation by means of multi-dimensional Gaussian integral formula and the discrete Fourier transformation.

The loopy belief propagation can be extended to a generalized belief propagation by means of the cluster variation method which is one of familiar advanced mean-field methods in the statistical mechanics[11]. The estimates of hyperparameters, $\widehat{\alpha}$ and $\widehat{\sigma}$, and the restored image obtained by replacing the loopy belief

Table 1: Values of estimates $(\widehat{\alpha}, \widehat{\sigma})$, mean-square error between the original image $\vec{a}$ and the restored image $\vec{a}$, $\mathrm{MSE}(\vec{a}, \widehat{\vec{a}}) \equiv \frac{1}{L}||\vec{a} - \widehat{\vec{a}}||^2$ and the improvement of signal to noise ratio, $\Delta_{\mathrm{SNR}} \equiv 10 \log_{10}\left(\frac{||\vec{a}-\vec{d}||^2}{||\vec{a}-\widehat{\vec{a}}||^2}\right)$ (dB), in the noise reductions by means of the Gaussian graphical model given in Figure 2.

	$\widehat{\alpha}$	$\widehat{\sigma}$	$\mathrm{MSE}(\vec{a}, \widehat{\vec{a}})$	Δ_{SNR} (dB)
Figure 2(c)	0.000483	31.094	331	6.2848
Figure 2(d)	0.000516	31.749	313	6.5321
Figure 2(e)	0.000524	31.910	308	6.5905

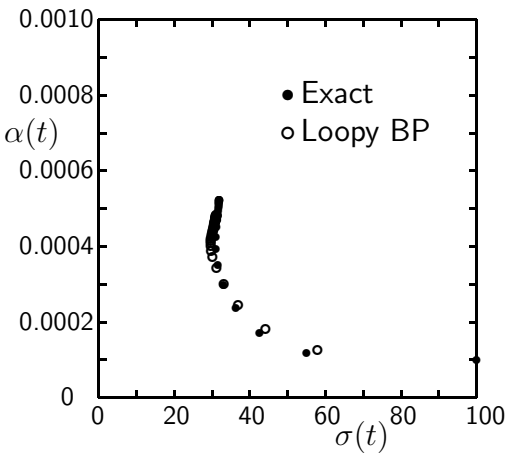

Figure 3: Convergence of $(\alpha(t), \sigma(t))$ $(\sigma(t) \equiv \sqrt{\beta(t)})$ in the EM algorithm. The open circle and the solid circle are corresponding to the results in the EM algorithm for the loopy belief propagation and the exact calculation, respectively.

propagation by the generalized belief propagation in the scheme of the present section are also shown in Table 1 and Figure 2. The results are more close to the exact results than the loopy belief propagation. The details have been given in Ref.[27].

5 Concluding Remarks

In the present tutorial lecture note, we have briefly explained the formulation of the probabilistic image processing based on the Markov random fields, the Bayesian networks and the loopy belief propagation. As an examples of the applications of the probabilistic image processing, we have reviewed the noise reduction by means of the Gaussian graphical model as Markov random fields.

The accuracy of the loopy belief propagation has been investigated for the Gaussian graphical model. As is one of important results in the investigation of the loopy belief propagation, it is well known that the loopy belief propagation can give us the exact results for the average in the Gaussian graphical model, though the variance and the co-variance might be approximate ones[7]. This fact is valid not only in the loopy belief propagation but also the generalized belief propagation for the Gaussian graphical model[27].

References

[1] D. Geman, "Random Fields and Inverse Problems in Imaging," Lecture Notes in Mathematics, no.1427, pp.113-193, Springer-Verlag, 1990.

[2] R. Chellappa and A. Jain (eds), *Markov Random Fields*: *Theory and Applications*, Academic Press, New York, 1993.

[3] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, Tokyo, 1995.

[4] A. S. Willsky, "Multiresolution Markov Models for Signal and Image Processing," Proceedings of IEEE, vol.90, no.8, pp.1396-1458, 2002.

[5] D. M. Chickering, D. Heckermana and C. Meek, "Large-Sample Learning of Bayesian Networks is NP-Hard," Journal of Machine Learning Research vol.5, pp.1287-1330, 2004.

[6] J. Pearl: *Probabilistic Reasoning in Intelligent Systems*: *Networks of Plausible Inference*, Morgan Kaufmann, 1988.

[7] 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.

[8] F. R. Kschischang, B. J. Frey and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," IEEE Trans. Inform. Theory, vol.47, no.2, pp.498-519, February 2001.

[9] Y. Kabashima and D. Saad: "Belief propagation *vs.* TAP for decoding corrupted messages," Europhysics Letters, vol.44, no.5, pp.668-674, 1998.

[10] M. Opper and D. Saad (eds): *Advanced Mean Field Methods —Theory and Practice—*, MIT Press, 2001.

[11] J. S. Yedidia, W. T. Freeman and Y. Weiss: "Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms," IEEE Transactions on Information Theory, vol.51, no.7, pp.2282-2312, 2005.

[12] S. Ikeda, T. Tanaka and S. Amari: "Stochastic reasoning, free energy, and information geometry," Neural Computation, vol.16, no.9, pp.1779-1810, 2004.

[13] S. Ikeda, T. Tanaka and S. Amari: "Information geometry of turbo and low-density parity-check codes," IEEE Transactions on Information Theory, vol.50, no.6, pp.1097-1114, 2004.

[14] W. T. Freeman, E. C. Pasztor, O. T. Carmichael, "Learning Low-Level Vision," International Journal of Computer Vision, vol.40, no.1, pp.25-47, 2000.

[15] W. T. Freeman, T. R. Jones and E. C. Pasztor, "Example-based super-resolution," IEEE Computer Graphics and Applications, vol.22, no.2, pp.56-65, 2002.

[16] J. Sun, N.-N. Zheng and H.-Y. Shum, "Stereo matching using belief propagation," IEEE Transactions on Patten Analysis and Machine Intelligence, vol.25, no.7, pp.787-800, 2003.

[17] B. J. Frey and N. Jojic: "A comparison of algorithms for inference and learning in probabilistic graphical models," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.27, no.9, pp.1392-1416, 2005

[18] K. Tanaka, "Statistical-mechanical approach to image processing (Topical Review)," Journal of Physics A: Mathematical and General, vol.35, no.37, pp.R81-R150, 2002.

[19] K. Tanaka, H. Shouno, M. Okada and D. M. Titterington, "Accuracy of the Bethe approximation for hyperparameter estimation in probabilistic image processing," Journal of Physics A: Mathematical and General, vol.37, no.36, pp.8675-8696, 2004.

[20] T. Heskes, "On the uniqueness of loopy belief propagation fixed points," Neural Computation, vol.16, pp.2379-2413, 2004.

[21] D. J. MacKay, "Bayesian interpolation," Neural Computation, vol.4, pp.415-447, 1992.

[22] K. Tanaka and J. Inoue, "Maximum likelihood hyperparameter estimation for solvable Markov random field model in image restoration," IEICE Transactions on Information and Systems, vol.E85-D, no.3, pp.546-557, March 2002.

[23] K. Tanaka, "Statistical-mechanical Iterative Algorithm by means of cluster variation method in compound Gauss-Markov random field model, Transactions of Japanese Society for Artificial Intelligence, vol.16, no.2, pp.259-267, 2001.

[24] K. Tanaka, J. Inoue and D. M. Titterington, "Probabilistic image processing by means of Bethe approximation for Q-Ising model," Journal of Physics A: Mathematical and General, vol.36, no.43, pp.11023-11036, 2003.

[25] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford University Press, Oxford, 2001.

[26] J. Inoue and K. Tanaka, "Dynamics of the maximum likelihood hyper-parameter estimation in image restoration: Gradient descent versus expectation and maximization algorithm," Physical Review E, vol.65, no.1, Article No. 016125, pp.1-11, 2002.

[27] K. Tanaka, "Generalized belief propagation formula in probabilistic information processing based on Gaussian graphical Model, IEICE Transactions (D-II), vol.J88-D-II, 2005, in press.