

LOOPY BELIEF PROPAGATION AND PROBABILISTIC IMAGE PROCESSING

Kazuyuki Tanaka

Graduate School of Information Sciences, Tohoku University,
Aramaki-Aza-Aoba 09, Aoba-ku, Sendai 980-8579, Japan

E-mail: kazu@statp.is.tohoku.ac.jp

URL: <http://www.statp.is.tohoku.ac.jp/~kazu/>

Jun-ichi Inoue

Complex Systems Engineering, Graduate School of Engineering,
Hokkaido University, N13-W8, Kita-ku, Sapporo 060-8628, Japan

E-mail: j_inoue@complex.eng.hokudai.ac.jp

D. M. Titterton

Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK

E-mail: mike@stats.gla.ac.uk

Abstract. Estimation of hyperparameters by maximization of the marginal likelihood in probabilistic image processing is investigated by using the cluster variation method. The algorithms are substantially equivalent to generalized loopy belief propagation.

INTRODUCTION

Combination of the Bayesian approach with Markov random field models provides a powerful method for probabilistic image processing[1]. It is also known that advanced mean-field methods and other statistical-mechanical methods are applicable to Bayesian image processing[2, 3].

Advanced mean-field methods have been widely applied to many problems in computer science[4]. The methods can be formulated as variational methods based on the minimization of an approximate Kullback-Leibler divergence or an approximate free energy. The cluster variation method, sometimes referred to as the Kikuchi method[5, 6], is one of the familiar advanced mean-field methods. Probabilistic inference algorithms in artificial intelligence are constructed by applying loopy belief propagation to probabilistic models with graphical representations[8]. It has been suggested that, for probabilistic models, the extremum conditions for approximate free energies associated with some approximations generated by the cluster variation method are equivalent to simultaneous fixed-point equations of loopy belief propagation,

and ordinary loopy belief propagation can be extended to provide generalized belief propagation by using the cluster variation method[13, 14].

Tanaka and Morita[7] constructed probabilistic image processing algorithms by the Bayesian approach with Markov random field models by using the cluster variation method, and Weiss[9] investigated approximate inference for Markov random fields by means of loopy belief propagation. The loopy belief propagation has been applied to practical computer vision problems[10]. In the probabilistic image processing with Markov random field models, hyperparameters should be estimated from the observed data. As one of the familiar frameworks to estimate hyperparameters, we have a maximum likelihood estimation[11]. Tanaka[12] applied a pair approximation version of the cluster variation method to the estimation of hyperparameters by maximizing a marginal likelihood and concluded that use of a pair approximation can improve the quality of the restored image compared to what is possible with the mean-field approximation.

In the present paper, we investigate hyperparameter estimation by maximization of the marginal likelihood in probabilistic image processing by using the cluster variation method. The present framework is applicable for probabilistic models on any *factor graphs* and is a generalization of [12]. We adopt a square-cactus approximation as well as a pair approximation in the cluster variation method. The algorithms are essentially the same as those corresponding to generalized loopy belief propagation.

BAYESIAN IMAGE ANALYSIS USING GRAPHICAL MODELS

We consider an image on a square lattice $\Omega \equiv \{i\}$. Each pixel takes one of the Q gray-levels $\mathbf{Q} = \{0, 1, 2, \dots, Q - 1\}$. Each gray-level expresses the intensity of light at a pixel in computer vision and 0 and $Q - 1$ correspond respectively to black and white. Random variables representing intensity at pixel i in the original image and the degraded image are denoted by F_i and G_i , respectively, and the random fields of intensities in the original image and the degraded image are represented by $\mathbf{F} \equiv \{F_i | i \in \Omega\}$ and $\mathbf{G} \equiv \{G_i | i \in \Omega\}$, respectively. The actual original image and the degraded image are denoted by $\mathbf{f} = \{f_i\}$ and $\mathbf{g} = \{g_i\}$, respectively.

In the present paper, it is assumed that the degraded image \mathbf{g} is generated from the original image \mathbf{f} by changing the intensity of each pixel to another intensity with the same probability p , independently of the other pixels. The conditional probability distribution of the degradation process when the original image is \mathbf{f} is given as follows:

$$\Pr\{\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, p\} = \prod_{i \in \Omega} \left(p(1 - \delta_{f_i, g_i}) + (1 - (Q - 1)p)\delta_{f_i, g_i} \right), \quad (1)$$

where $\delta_{a,b}$ is the Kronecker delta. Moreover, the *a priori* probability that the

original image is \mathbf{f} is assumed to be

$$\Pr\{\mathbf{F} = \mathbf{f}|\alpha\} = \frac{\prod_{ij \in N} \exp\left(-\frac{1}{2}\alpha(f_i - f_j)^2\right)}{\sum_{\mathbf{z}} \prod_{ij \in N} \exp\left(-\frac{1}{2}\alpha(z_i - z_j)^2\right)}, \quad (2)$$

where N is the set of all the nearest-neighbor pairs of pixels on the square lattice Ω and the summation $\sum_{\mathbf{z}}$ is taken over all possible configurations of an image $\mathbf{z} = \{z_i|i \in \Omega\}$. By substituting Eqs.(1) and (2) into the Bayes formula, we have

$$\Pr\{\mathbf{F} = \mathbf{f}|\mathbf{G} = \mathbf{g}, p, \alpha\} = \frac{\Pr\{\mathbf{G} = \mathbf{g}|\mathbf{F} = \mathbf{f}, p\}\Pr\{\mathbf{F} = \mathbf{f}|\alpha\}}{\sum_{\mathbf{z}} \Pr\{\mathbf{G} = \mathbf{g}|\mathbf{F} = \mathbf{z}, p\}\Pr\{\mathbf{F} = \mathbf{z}|\alpha\}}, \quad (3)$$

thereby creating the *a posteriori* probability distribution $\Pr\{\mathbf{F} = \mathbf{f}|\mathbf{G} = \mathbf{g}, p, \alpha\}$.

In the maximum marginal likelihood estimation method, the hyperparameters α and p are determined so as to maximize the marginal likelihood $\Pr\{\mathbf{G} = \mathbf{g}|\alpha, p\}$:

$$\Pr\{\mathbf{G} = \mathbf{g}|\alpha, p\} \equiv \sum_{\mathbf{z}} \Pr\{\mathbf{G} = \mathbf{g}|\mathbf{F} = \mathbf{z}, p\}\Pr\{\mathbf{F} = \mathbf{z}|\alpha\}. \quad (4)$$

The maximizers of $\Pr\{\mathbf{G} = \mathbf{g}|\alpha, p\}$ are denoted by $\hat{\alpha}$ and \hat{p} , such that

$$(\hat{\alpha}, \hat{p}) = \arg \max_{(\alpha, p)} \Pr\{\mathbf{G} = \mathbf{g}|\alpha, p\}. \quad (5)$$

The conditions for an extremum of $\Pr\{\mathbf{G} = \mathbf{g}|\alpha, p\}$ at $\alpha = \hat{\alpha}$ and $p = \hat{p}$ can be reduced to the following simultaneous equations:

$$\begin{aligned} & \sum_{ij \in N} \sum_{z_i \in \mathbf{Q}} \sum_{z_j \in \mathbf{Q}} (z_i - z_j)^2 \Pr\{F_i = z_i, F_j = z_j|\mathbf{G} = \mathbf{g}, \hat{\alpha}, \hat{p}\} \\ & = \sum_{ij \in N} \sum_{z_i \in \mathbf{Q}} \sum_{z_j \in \mathbf{Q}} (z_i - z_j)^2 \Pr\{F_i = z_i, F_j = z_j|\hat{\alpha}\}, \end{aligned} \quad (6)$$

$$\sum_{\{i|i \in \Omega\}} \sum_{z_i \in \mathbf{Q}} \delta_{z_i, g_i} \Pr\{F_i = z_i|\mathbf{G} = \mathbf{g}, \hat{\alpha}, \hat{p}\} = 1 - (Q - 1)p. \quad (7)$$

Here $\Pr\{F_i = f_i|\mathbf{G} = \mathbf{g}, \alpha, p\}$, $\Pr\{F_i = f_i, F_j = f_j|\mathbf{G} = \mathbf{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j|\alpha\}$ are marginal probabilities defined by

$$\Pr\{F_i = f_i|\mathbf{G} = \mathbf{g}, \alpha, p\} \equiv \sum_{\mathbf{z}} \delta_{f_i, z_i} \Pr\{\mathbf{F} = \mathbf{z}|\mathbf{G} = \mathbf{g}, \alpha, p\}, \quad (8)$$

$$\Pr\{F_i = f_i, F_j = f_j|\mathbf{G} = \mathbf{g}, \alpha, p\} \equiv \sum_{\mathbf{z}} \delta_{f_i, z_i} \delta_{f_j, z_j} \Pr\{\mathbf{F} = \mathbf{z}|\mathbf{G} = \mathbf{g}, \alpha, p\}, \quad (9)$$

$$\Pr\{F_i = f_i, F_j = f_j|\alpha\} \equiv \sum_{\mathbf{z}} \delta_{f_i, z_i} \delta_{f_j, z_j} \Pr\{\mathbf{F} = \mathbf{z}|\alpha\}. \quad (10)$$

Given $\hat{\alpha}$ and \hat{p} , the restored image $\hat{\mathbf{f}} = \{\hat{f}_i | i \in \Omega\}$ is determined by

$$\hat{f}_i = \arg \max_{z_i \in \mathcal{Q}} \Pr\{F_i = z_i | \mathbf{G} = \mathbf{g}, \hat{\alpha}, \hat{p}\}. \quad (11)$$

This approach to image restoration is called maximum posterior marginal estimation.

LOOPY BELIEF PROPAGATION AND THE CLUSTER VARIATION METHOD

In the above framework, we have to calculate the marginal probability distributions $\Pr\{F_i = f_i | \mathbf{G} = \mathbf{g}, \alpha, p\}$ ($i \in \Omega$), $\Pr\{F_i = f_i, F_j = f_j | \mathbf{G} = \mathbf{g}, \alpha, p\}$ ($ij \in N$) and $\Pr\{F_i = f_i, F_j = f_j | \alpha\}$ ($ij \in N$). Since it is hard to calculate these marginal probability distributions exactly, we apply the cluster variation method to the models given by $\Pr\{\mathbf{F} = \mathbf{f} | \mathbf{G} = \mathbf{g}, \alpha, p\}$ and $\Pr\{\mathbf{F} = \mathbf{f} | \alpha\}$.

In order to explain the framework of the cluster variation method, we introduce some notation for clusters. A cluster is a set of nodes. When a node i belongs to cluster γ , we call i an element of γ and we write $i \in \gamma$. When all the nodes in a cluster γ' belong to a cluster γ , we call γ' a subcluster of γ , and we write $\gamma' \leq \gamma$. We write $\gamma' < \gamma$ when a cluster γ' is a proper subcluster of γ . The set of nodes which belong to cluster γ , but not to cluster γ' , is denoted by $\gamma \setminus \gamma'$. The notation $\Omega \setminus \gamma$ represents the set of all nodes not belonging to the cluster γ .

First, we have to specify a set B of basic clusters. Every basic cluster must not be a subcluster of another basic cluster. We consider a set C of clusters such that a cluster is in C if and only if it is a member of B or is the cluster of the common nodes of two or more members of B , excluding the empty cluster $\mathbf{0}$. The set of all clusters which belonging to C , but not to B , is denoted by $C \setminus B$. A set of random variables f_i associated with nodes i belonging to a cluster γ is denoted by $\mathbf{f}_\gamma \equiv \{f_i | i \in \gamma\}$.

We consider a probability distribution given by

$$P(\mathbf{f}) \equiv \frac{1}{Z} \prod_{\{\gamma | \gamma \in C\}} W_\gamma(\mathbf{f}_\gamma)^{-\mu(\gamma)}, \quad (12)$$

where Z is a normalizing constant and $\mu(\gamma)$ ($\gamma \in C$) is a so-called Möbius function, defined by

$$\mu(\gamma) \equiv -1 - \sum_{\{\gamma' | \gamma' > \gamma, \gamma' \in C\}} \mu(\gamma'). \quad (13)$$

We wish to compute the marginal probability distribution defined by

$$P_\gamma(\mathbf{f}_\gamma) \equiv \sum_{\mathbf{z}} \left(\prod_{\{i | i \in \gamma\}} \delta_{f_i, z_i} \right) P(\mathbf{z}). \quad (14)$$

It is difficult to obtain the exact values of the marginal probability distribution and we have to employ some approximations.

We introduce the Kullback-Leibler divergence between the probability distributions $P(\mathbf{f})$ and $Q(\mathbf{f})$, defined by

$$\mathcal{D}[Q||P] \equiv \sum_{\mathbf{z}} Q(\mathbf{z}) \ln \left(\frac{Q(\mathbf{z})}{P(\mathbf{z})} \right). \quad (15)$$

where $P(\mathbf{f})$ is given in Eq.(12) and $Q(\mathbf{f})$ is a trial. In the cluster variation method, the probability distribution $Q(\mathbf{f})$ is approximately restricted to a distribution of the form

$$Q(\mathbf{f}) = \prod_{\{\gamma|\gamma \in C\}} Q_{\gamma}(\mathbf{f}_{\gamma})^{-\mu(\gamma)}, \quad (16)$$

where $Q_{\gamma}(\mathbf{f}_{\gamma})$ is the marginal probability distribution obtained from the probability distribution $Q(\mathbf{f})$ and is defined by

$$Q_{\gamma}(\mathbf{f}_{\gamma}) \equiv \sum_{\mathbf{z}} \left(\prod_{\{i|i \in \gamma\}} \delta_{f_i, z_i} \right) Q(\mathbf{z}). \quad (17)$$

The definition (17) can be rewritten in the form of the following relationships among the marginal probability distributions $Q_{\gamma}(\mathbf{f}_{\gamma})$:

$$Q_{\gamma}(\mathbf{f}_{\gamma}) = \sum_{\mathbf{z}_{\gamma'}} \left(\prod_{\{i|i \in \gamma\}} \delta_{f_i, z_i} \right) Q_{\gamma'}(\mathbf{z}_{\gamma'}) \quad (\gamma < \gamma', \gamma \in C \setminus B, \gamma' \in C). \quad (18)$$

In the cluster variation method, the above relationships are called *reducibilities*.

Use of Eqs.(16) and (17) reduces the Kullback-Leibler divergence $\mathcal{D}[Q||P]$ to

$$\mathcal{D}[Q||P] = \mathcal{F}[\{Q_{\gamma}|\gamma \in C\}] + \ln(Z), \quad (19)$$

where

$$\mathcal{F}[\{Q_{\gamma}|\gamma \in C\}] \equiv - \sum_{\{\gamma|\gamma \in C\}} \mu(\gamma) \mathcal{D}[Q_{\gamma}||W_{\gamma}]. \quad (20)$$

If we choose functions $W_{\gamma}(\mathbf{f}_{\gamma})$ so as to satisfy the normalization conditions, $\mathcal{D}[Q_{\gamma}||W_{\gamma}]$ can be also regarded as the Kullback-Leibler divergence between $Q_{\gamma}(\mathbf{f}_{\gamma})$ and $W_{\gamma}(\mathbf{f}_{\gamma})$. In the cluster variation method, the approximate marginal probability distributions $P_{\gamma}(\mathbf{f}_{\gamma})$ are determined so as to minimize the right-hand side of Eq.(19), the approximate form of the Kullback-Leibler divergence in the cluster variation method, subject to constraints imposed by the normalizations and the reducibilities required of the marginal probability distributions. Thus

$$\{P_{\gamma}|\gamma \in C\} \simeq \arg \min_{\{Q_{\gamma}|\gamma \in C\}} \left\{ \mathcal{F}[\{Q_{\gamma}|\gamma \in C\}] \middle| \sum_{\mathbf{z}_{\gamma}} Q_{\gamma}(\mathbf{z}_{\gamma}) = 1 \quad (\gamma \in C), \right. \\ \left. Q_{\gamma}(\mathbf{f}_{\gamma}) = \sum_{\mathbf{z}_{\gamma'}} \left(\prod_{\{i|i \in \gamma\}} \delta_{f_i, z_i} \right) Q_{\gamma'}(\mathbf{z}_{\gamma'}) \quad (\gamma < \gamma' \in C) \right\}. \quad (21)$$

We introduce Lagrange multipliers associated with the normalizations and the reducibilities and take the first variation of $\mathcal{F}[\{Q(\gamma)|\gamma \in C\}]$ with respect to the marginal probability distributions $Q_\gamma(\mathbf{f}_\gamma)$. This leads to the following approximations for the marginal probability distributions $Q_\gamma(\mathbf{f}_\gamma)$:

$$Q_\gamma(\mathbf{f}_\gamma) = \frac{1}{Z_\gamma} W_\gamma(\mathbf{f}_\gamma) \prod_{\{\gamma'|\gamma' \leq \gamma, \gamma' \in C \setminus B\}} \exp(\lambda_{\gamma', \gamma}(\mathbf{f}_{\gamma'})), \quad (22)$$

where

$$\lambda_{\gamma, \gamma'}(\mathbf{f}_\gamma) = - \sum_{\{\gamma'|\gamma' > \gamma, \gamma' \in C\}} \frac{\mu(\gamma')}{\mu(\gamma)} \lambda_{\gamma, \gamma'}(\mathbf{f}_\gamma). \quad (23)$$

Here $\lambda_{\gamma, \gamma'}$ are Lagrange multipliers for the reducibilities (18) and are determined so as to satisfy Eqs.(18) and (23) with (22).

We choose the sets B and C so as to satisfy $C \setminus B = \Omega \cap C$ and consider the cluster variation method for the probabilistic model

$$P(\mathbf{f}) = \frac{1}{Z} \left(\prod_{\{i|i \in \Omega\}} W_i(f_i) \right) \left(\prod_{\{\gamma|\gamma \in B\}} W_\gamma(\mathbf{f}_\gamma) \left(\prod_{\{j|j \in \gamma\}} W_j(f_j)^{-1} \right) \right). \quad (24)$$

Here $W_\gamma(\mathbf{f}_\gamma)$ and $W_i(f_i)$ are always positive for any values of \mathbf{f}_γ and f_i . We introduce a set $\mathbf{c}_i \equiv \{\gamma|\gamma > i, \gamma \in C\}$ for each pixel i . By replacing $\lambda_{i, \gamma}(f_i)$ by $M_{\gamma \rightarrow i}(f_i)$, defined by

$$M_{\gamma \rightarrow i}(f_i) = \prod_{\gamma' \in \mathbf{c}_i \setminus \gamma} \exp(\lambda_{i, \gamma'}(f_i)), \quad (25)$$

we obtain the following equations for the sets of marginal probabilities $\{P_i(\xi)|i \in \Omega\}$ and $\{P_\gamma(\mathbf{f}_\gamma)|\gamma \in B\}$:

$$P_i(f_i) \simeq \frac{W_i(f_i) \prod_{\{\gamma'|\gamma' \in \mathbf{c}_i\}} M_{\gamma' \rightarrow i}(f_i)}{\sum_{z_i} W_i(z_i) \prod_{\{\gamma'|\gamma' \in \mathbf{c}_i\}} M_{\gamma' \rightarrow i}(z_i)}, \quad (26)$$

$$P_\gamma(\mathbf{f}_\gamma) \simeq \frac{W_\gamma(\mathbf{f}_\gamma) \prod_{\{i|i \in \gamma\}} \prod_{\{\gamma'|\gamma' \in \mathbf{c}_i \setminus \gamma\}} M_{\gamma' \rightarrow i}(f_i)}{\sum_{\mathbf{z}_\gamma} W_\gamma(\mathbf{z}_\gamma) \prod_{\{i|i \in \gamma\}} \prod_{\{\gamma'|\gamma' \in \mathbf{c}_i \setminus \gamma\}} M_{\gamma' \rightarrow i}(z_i)}, \quad (27)$$

$$M_{\gamma \rightarrow i}(f_i) = \frac{\sum_{\mathbf{z}_\gamma} \delta_{f_i, z_i} \left(\frac{W_\gamma(\mathbf{z}_\gamma)}{W_i(z_i)} \right) \prod_{\{j|j < \gamma\}} \prod_{\{\gamma'|\gamma' \in \mathbf{c}_j \setminus \gamma\}} M_{\gamma' \rightarrow j}(z_j)}{\sum_{\mathbf{z}_\gamma} \left(\frac{W_\gamma(\mathbf{z}_\gamma)}{W_i(z_i)} \right) \prod_{\{j|j < \gamma\}} \prod_{\{\gamma'|\gamma' \in \mathbf{c}_j \setminus \gamma\}} M_{\gamma' \rightarrow j}(z_j)}. \quad (28)$$

The marginal probability distribution $P_{ij}(f_i, f_j)$ ($ij \leq \gamma$, $\gamma \in B$) is

$$P_{ij}(f_i, f_j) = \sum_{\mathbf{z}_\gamma} \delta_{z_i, f_i} \delta_{z_j, f_j} P_\gamma(\mathbf{z}_\gamma). \quad (29)$$

In the pair approximation and the cactus-square approximation versions of the cluster variation method, the sets B are chosen as shown in Figs.1(a) and 1(b), respectively. If we set

$$W_i(f_i) = p(1 - \delta_{f_i, g_i}) + (1 - (Q - 1)p) \delta_{f_i, g_i} \quad (i \in \Omega), \quad (30)$$

$$W_\gamma(\mathbf{f}_\gamma) = \left(\prod_{\{i|i \in \Omega\}} W_i(f_i) \right) \exp\left(-\frac{1}{2} \alpha \sum_{\{ij|ij \leq \gamma, ij \in N\}} (f_i - f_j)^2 \right) \quad (\gamma \in B), \quad (31)$$

the marginal probabilities $\Pr\{F_i = f_i | \mathbf{G} = \mathbf{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j | \mathbf{G} = \mathbf{g}, \alpha, p\}$ are obtained as $P_i(f_i)$ and $P_{ij}(f_i, f_j)$, respectively. By setting

$$W_i(f_i) = 1 \quad (i \in \Omega), \quad (32)$$

$$W_\gamma(\mathbf{f}_\gamma) = \exp\left(-\frac{1}{2} \alpha \sum_{\{ij|ij \leq \gamma, ij \in N\}} (f_i - f_j)^2 \right) \quad (\gamma \in B), \quad (33)$$

we obtain the marginal probabilities $\Pr\{F_i = f_i, F_j = f_j | \alpha\}$ as $P_{ij}(f_i, f_j)$. Though these forms may be not so familiar to some physicists, $\ln(M_{i \rightarrow j}(\xi))$ corresponds to the effective field in the conventional Bethe approximation. In probabilistic inference, the quantity $M_{i \rightarrow j}(\xi)$ is called a message propagated from i to j .

Eqs.(28) have the form of fixed-point equations for the messages $M_{i \rightarrow j}(\xi)$. In practical numerical calculations, we solve the simultaneous equations (28) by numerical search in an iteration method with an annealing procedure[3]. For various values of the hyperparameters α and p , we obtain the marginal probability distributions $\Pr\{F_i = f_i | \mathbf{G} = \mathbf{g}, \alpha, p\}$, $\Pr\{F_i = f_i, F_j = f_j | \mathbf{G} = \mathbf{g}, \alpha, p\}$ and $\Pr\{F_i = f_i, F_j = f_j | \alpha\}$ and search for the optimal set of values, $(\hat{\alpha}, \hat{p})$, that satisfy Eqs.(6) and (7) numerically.

NUMERICAL EXPERIMENTS

In this section, we report about some numerical experiments. The optimal values for the hyperparameters, $(\hat{\alpha}, \hat{p})$, are determined by means of maximum marginal likelihood estimation together with the cluster variation method. We adopt a pair approximation and a square-cactus approximation when applying the cluster variation method. The sets B of basic clusters in the pair approximation and in the square-cactus approximation are shown in Fig.1.

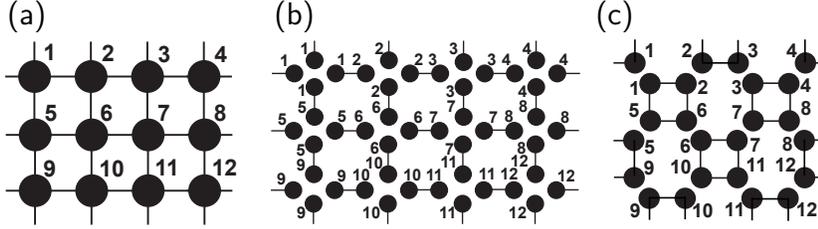


Figure 1: The square lattice and the sets of basic clusters for the pair approximation and the square-cactus approximation as used in the cluster variation method. (a) The square lattice. (b) The set B of basic clusters in the pair approximation. (c) The set B of basic clusters in the square-cactus approximation.

We performed numerical experiments for artificial binary images generated from the 256-valued standard image “Mandrill” by thresholding. The restored images obtained by means of iterative algorithms based on the mean-field approximation, the pair approximation and the square-cactus approximation are shown in Fig. 2. In table 1 we give the estimates of the hyperparameters, \hat{p} and $\hat{\alpha}$, and the values of the improvement of the signal-to-noise ratio, Δ_{SNR} (dB), defined by

$$\Delta_{\text{SNR}} \equiv 10 \log_{10} \left(\frac{\|\mathbf{f} - \mathbf{g}\|^2}{\|\mathbf{f} - \hat{\mathbf{f}}\|^2} \right) \text{ (dB)}. \quad (34)$$

As shown in Fig. 1 and Table 1, although use of the pair approximation and

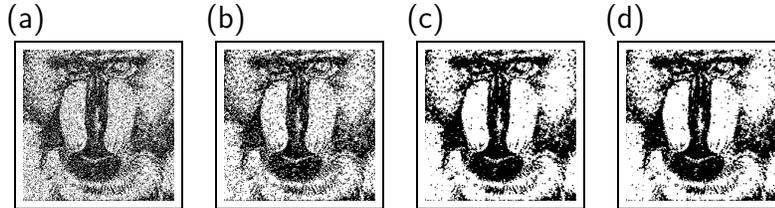


Figure 2: Image restorations for a binary image \mathbf{f} generated from the 256-valued standard image “Mandrill” by thresholding. (a) The degraded image \mathbf{g} ($p = 0.2$). (b) The restored image $\hat{\mathbf{f}}$ based on the mean-field approximation. (c) The restored image $\hat{\mathbf{f}}$ based on the pair approximation. (d) The restored image $\hat{\mathbf{f}}$ based on the square-cactus approximation.

the square-cactus approximation in the cluster variation method can improve the quality of the restored image relative to the mean-field approximation, the result obtained from the square-cactus approximation is very similar to that from the pair approximation. Moreover, we have analysed twenty binary images generated by Monte Carlo simulation from the *a priori* probability distribution (2) for $\alpha = 2.15^{-1}$ and obtained results similar to those in Fig. 2 and Table 1.

TABLE 1: THE ESTIMATES OF THE HYPERPARAMETERS, \hat{p} AND $\hat{\alpha}$, AND THE VALUES OF Δ_{SNR} OBTAINED FOR THE DEGRADED IMAGE \mathbf{g} GIVEN IN FIGURE 2(A). THE HYPERPARAMETERS ARE ESTIMATED BY APPLYING THE MEAN-FIELD APPROXIMATION, THE PAIR APPROXIMATION AND THE SQUARE-CACTUS APPROXIMATION, RESPECTIVELY, IN THE CONTEXT OF MAXIMUM MARGINAL LIKELIHOOD ESTIMATION. HERE THE TRUE VALUE OF p IS 0.2.

Approx.	\hat{p}	$\hat{\alpha}$	Δ_{SNR} (dB)
MFA	0.100666	0.250971	0.979681
PA	0.179189	0.396343	2.304090
SCA	0.179701	0.393583	2.297306

CONCLUDING REMARKS

We have summarized the cluster variation method for probabilistic image processing based on Bayesian analysis and maximization of the marginal likelihood. In particular, we developed a generalized belief propagation algorithm by restricting the set $C \setminus B$ to that consisting only of single pixels, so that $C \setminus B = \Omega \cap C$, which is called *factor graph*[8]. Of course, we expect that the results obtained would improve if we adopted larger basic clusters. However, the computational complexity would also grow. A most important point is how large the basic clusters have to be for us to obtain sufficiently good results. The results in the present paper suggest that the pair approximation can give us good results for the probabilistic model based on interactions between the nearest-neighbor pairs of pixels.

Fig.3(c) shows a restoration of a 4-valued image obtained by means of the pair approximation. This result is obtained by assuming Eq.(2) as the *a priori* probability distribution. The probabilistic model given in Eq.(2) is referred to as the Q -state Ising model in statistical mechanics. The Q -state Potts model is the other familiar probabilistic model in statistical mechanics[12]. The Q -state Potts model reflects spatial flatness in images, while the Q -state Ising model reflects spatial smoothness. We are currently investigating the estimation of hyperparameters in probabilistic image restoration when we adopt the Q -state Potts model or the Q -state Ising model as *a priori* probability distribution. One of the results obtained by using the Q -state Potts model as *a priori* probability distribution is shown in Fig.3(d). We will report this investigation in detail elsewhere[15].

Acknowledgements

The authors are grateful to Professor T. Horiguchi of the Graduate School of Information Science, Tohoku University, for valuable discussions. This work was partly supported by the Grants-In-Aid (No.13680384 and No.14084203) for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

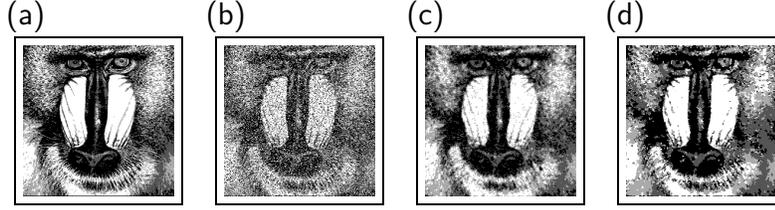


Figure 3: Image restorations for an artificial 4-valued image f ($Q = 4$). (a) The original image f . (b) The degraded image g ($(Q - 1)p = 0.2$). (c) The restored image \hat{f} based on the pair approximation for the *a priori* probability distribution (2). (d) The restored image \hat{f} based on the pair approximation for the *a priori* probability distribution given as a Q -state Potts model.

REFERENCES

- [1] S. Z. Li, *Markov Random Field Modeling in Computer Vision* (Springer-Verlag, 1995).
- [2] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, 2001)
- [3] 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.
- [4] M. Opper and D. Saad (edited), *Advanced Mean Field Methods — Theory and Practice* (MIT Press, 2001).
- [5] R. Kikuchi, "A theory of cooperative phenomena," *Phys. Rev.* vol.81, no.6, pp.988-1003, March 1951.
- [6] T. Morita, "Cluster variation method for non-uniform Ising and Heisenberg models and spin-pair correlation function," *Prog. Theor. Phys.*, vol.85, no.2, pp.243-255, Feb. 1991.
- [7] K. Tanaka and T. Morita, "Cluster variation method and image restoration problem," *Physics Letters A*, vol.203, no.2-3, pp.122-128, 1995.
- [8] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, 1988).
- [9] Y. Weiss, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Computation*, vol.13, pp.2173-2200, 2001.
- [10] W. T. Freeman, T. R. Jones and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications* (March/April), pp.56-65, 2002.
- [11] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. Signal Processing*, 40, pp. 2570-2583, 1992.
- [12] K. Tanaka, "Theoretical study of hyperparameter estimation by maximization of marginal likelihood in image restoration by means of cluster variation method," *Electronics and Communications in Japan, Part 3*, vol.85, no.7, pp.50-62, 2002.
- [13] J. S. Yedidia, W. T. Freeman and Y. Weiss, "Generalized belief propagation," *Advances in Neural Information Processing Systems*, vol.13, T. Leen, T. Dietterich, V. Tresp (eds), pp.689-695 (Cambridge, MA: MIT Press, 2001).
- [14] H. J. Kappen and W. Wiegner, "Novel iteration schemes for the cluster variation method," *Advances in Neural Information Processing System*, vol.14, T. G. Dietterich, S. Becker, and Z. Ghahramani (eds), pp.415-422 (Cambridge, MA: MIT Press, 2002).
- [15] K. Tanaka, J. Inoue and D. M. Titterton, "Probabilistic image processing by means of Bethe approximation for multi-state classical spin systems," to be submitted to *J. Phys. A*, 2003.