Randomness and Computation
Joint Workshop "New Horizons in Computing" and "Statistical Mechanical Approach to Probabilistic Information Processing" (18-21 July, 2005, Sendai, Japan)
Lecture Note in Tutorial Sessions

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 paper, we briefly explain the formulation of the probabilistic image processing and the theoretical structure of the belief propagation. As two examples of the applications of the probabilistic image processing, we introduce a noise reduction and a segmentation, and extend the segmentation to a motion detection.

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.

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

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], 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]. The belief propagation has been applied to many practical problems in the computer vision[13]-[22]. Theoretical study of the application of the belief propagation to image processing has been done in the statistical point of view[23, 24].

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. Some applications are described in Sections 3 and 4. In Section 3, we explain the probabilistic image restorations and noise reductions by using the Gaussian graphical model. In Section 4, we demonstrate the image segmentation which is constructed as a Bayesian network by using the Gauss mixture model and the Markov random field. We show a simple demonstration for the object recognition in the motion images by extending the scheme of the image segmentation by the Bayesian network. 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*[25, 26].

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 $P(\vec{a})$:

$$P(\vec{a})=\frac{\prod_{ij\in\mathcal{N}}\Phi_{ij}(a_i,a_j)}{\sum_{\vec{z}}\prod_{ij\in\mathcal{N}}\Phi_{ij}(z_i,z_j)}. \tag{5}$$

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.(5), 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:

$$P_i(a_i) \equiv \sum_{\vec{z}} \delta_{z_i,a_i} P(\vec{a}) \simeq \frac{\prod_{k\in\mathcal{N}_i} \mathcal{M}_{k\to i}(a_i)}{\sum_{z_i} \prod_{k\in\mathcal{N}_i} \mathcal{M}_{k\to i}(z_i)}, \tag{6}$$

$$P_{ij}(a_i,a_j) \equiv \sum_{\vec{z}} \delta_{z_i,a_i}\delta_{z_j,a_j} P(\vec{a}) \simeq \frac{\big(\prod_{k\in\mathcal{N}_i} \mathcal{M}_{k\to i}(a_i)\big)\Phi_{ij}(a_i,a_j)\big(\prod_{l\in\mathcal{N}_j} \mathcal{M}_{l\to j}(a_j)\big)}{\sum_{z_i}\sum_{z_j}\big(\prod_{k\in\mathcal{N}_i} \mathcal{M}_{k\to i}(z_i)\big)\Phi_{ij}(z_i,z_j)\big(\prod_{l\in\mathcal{N}_j} \mathcal{M}_{l\to j}(z_j)\big)}. \tag{7}$$

The quantity $\mathcal{M}_{j\to i}(a_i)$ in Eqs.(6) and (7) 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:

$$\mathcal{M}_{j\to i}(a_i) = \frac{\sum_{z_j} \Phi_{ij}(z_i,a_j) \prod_{k\in\mathcal{N}_j\backslash\{i\}} \mathcal{M}_{k\to j}(z_j)}{\sum_{z_i}\sum_{z_j} \Phi_{ij}(z_i,z_j) \prod_{k\in\mathcal{N}_j\backslash\{i\}} \mathcal{M}_{k\to j}(z_j)}, \quad \mathcal{M}_{i\to j}(a_j) = \frac{\sum_{z_i} \Phi_{ij}(a_i,z_j) \prod_{k\in\mathcal{N}_i\backslash\{j\}} \mathcal{M}_{k\to i}(z_i)}{\sum_{z_i}\sum_{z_j} \Phi_{ij}(z_i,z_j) \prod_{k\in\mathcal{N}_i\backslash\{j\}} \mathcal{M}_{k\to i}(z_i)}. \tag{8}$$

The graphical representations of Eqs.(6)-(8) are shown in Figure 1.

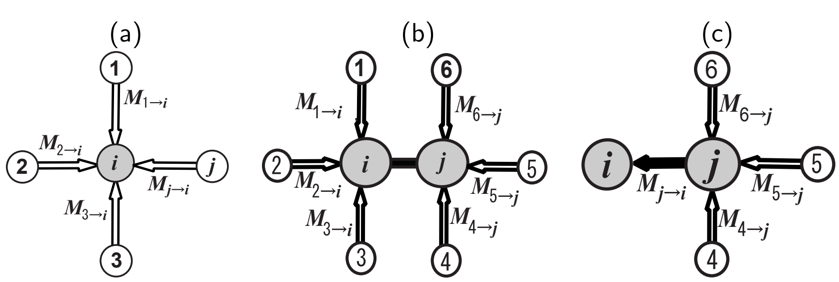

Figure 1: Graphical representations of the belief propagation. (a) Eq.(6). (b) Eq.(7). (c) Eq.(8).

The simultaneous fixed point equations (8) are derived by substituting Eqs.(6) and (7) to the reducibility conditions $P_i(a_i) = \sum_{z_j} P_{ij}(a_i,z_j)$ and $P_j(a_j) = \sum_{z_i} P_{ij}(z_i,a_j)$. We can calculate the numerical solutions of the simultaneous fixed point equations (8), numerically, by using the iteration method.

By applying the numerical solutions to Eqs.(6) and (7), we obtain the marginal probability $P_i(a_i)$ and $P_{ij}(a_i, a_j)$.

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:

$$\mathcal{P}(\vec{D}=\vec{d}|\vec{A}=\vec{a},\beta=\sigma^{-2}) = \prod_{i\in\Omega}\frac{1}{\sqrt{2\pi}\sigma}\exp\Big(-\frac{1}{2\sigma^2}(a_i-d_i)^2\Big), \tag{9}$$

$$\mathcal{P}(\vec{A}=\vec{a}|\alpha) = \frac{\displaystyle\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(a_i-a_j)^2\Big)}{\displaystyle\int\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(a_i-a_j)^2\Big)d\vec{z}} \quad (\alpha>0). \tag{10}$$

By substituting Eqs.(9) and (10) to the Bayes formula, the *a posteriori* probability is expressed as follows:

$$\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{a},\alpha,\beta=\sigma^{-2}) = \frac{\Big(\displaystyle\prod_{i\in\Omega}\exp\big(-\tfrac{1}{2\sigma^2}(a_i-d_i)^2\big)\Big)\Big(\displaystyle\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(a_i-a_j)^2\Big)\Big)}{\displaystyle\int\Big(\prod_{i\in\Omega}\exp\big(-\tfrac{1}{2\sigma^2}(z_i-d_i)^2\big)\Big)\Big(\prod_{ij\in\mathcal{N}}\exp\Big(-\tfrac{1}{2}\alpha(z_i-z_j)^2\Big)\Big)d\vec{z}}. \tag{11}$$

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 $\mathcal{P}(A_i=a_i|\vec{D}=\vec{d},\alpha,\beta=\sigma^{-2})$[24]. Because both probabilities in Eqs.(10) and (11) 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[26]. 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.

5 Image Segmentation

In this section, we explain the image segmentation in the natural image by means of the Gauss mixture model and the Markov random field model. In this case, the random fields of the segmented image and of the observed image are denoted by $\vec{A}$ and $\vec{D}$, respectively. In the segmented image, suppose that each pixel should classify into K classes and then takes one of the classes $\{1,2,\cdots,K\}$.

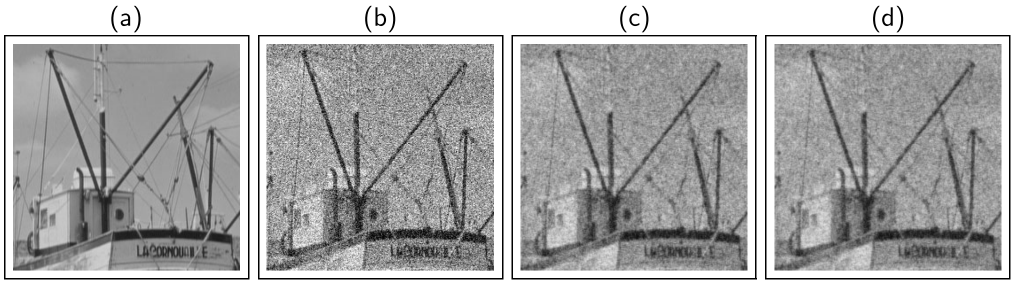

Figure 2: Noise reduction by using the Gaussian graphical model and the belief propagation. (a) Original image. (b) Degraded image by the additive white Gaussian noise with the mean 0 and the variance 40^2. (c) Restored image in the belief propagation. (d) Restored image in the exact calculation by means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

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, $\Delta_{\rm SNR} \equiv 10\log_{10}\big(\frac{||\vec{a}-\vec{d}||^2}{||\vec{a}-\widehat{\vec{a}}||^2}\big)$ (dB), in the noise reductions by means of the Gaussian graphical model given in Figure 2.

	$\widehat{\alpha}$	$\widehat{\sigma}$	$\mathrm{MSE}(\vec{a}, \widehat{\vec{a}})$	$\Delta_{\rm SNR}$ (dB)
Figure 2(c)	0.000565	33.950	260	7.686
Figure 2(d)	0.000640	34.915	237	8.103

The generating process of an observed image when a segmented image $\vec{a}$ is given is expressed in terms of the following conditional probability:

$$\mathcal{P}(\vec{D} = \vec{d} | \vec{A} = \vec{a}, \vec{\mu}, \vec{\sigma}) = \prod_{i\in\Omega} \frac{1}{\sqrt{2\pi}\sigma(a_i)} \exp\Big(- \frac{1}{2\sigma(a_i)^2} \big(d_i - \mu(a_i)\big)^2 \Big), \tag{12}$$

where $\vec{\mu} = (\mu(1), \mu(2), \cdots, \mu(K))$ and $\vec{\sigma} = (\sigma(1), \sigma(2), \cdots, \sigma(K))$ are the hyperparameters. Moreover, the *a priori* probability distribution that the segmented image is $\vec{a}$ is assumed to be

$$\mathcal{P}(\vec{A} = \vec{a} | \alpha, \vec{\gamma}) = \frac{\big(\prod_{i\in\Omega} \gamma(a_i)\big)\big(\prod_{ij\in\mathcal{N}} \exp(\alpha\delta_{a_i,a_j})\big)}{\sum_{\vec{z}} \big(\prod_{i\in\Omega} \gamma(z_i)\big)\big(\prod_{ij\in\mathcal{N}} \exp(\alpha\delta_{z_i,z_j})\big)}, \tag{13}$$

where α and $\vec{\gamma} = (\gamma(1), \gamma(2), \cdots, \gamma(K))$ are the hyperparameters.

By substituting Eqs.(12) and (13) to the Bayes formula, the *a posteriori* probability is derived as follows:

$$\mathcal{P}(\vec{A} = \vec{a} | \vec{D} = \vec{d}, \alpha, \vec{\gamma}, \vec{\mu}, \vec{\sigma}) = \frac{\Bigg(\prod_{i\in\Omega} \Big(\frac{\gamma(a_i)}{\sigma(a_i)}\Big) \exp\Big(- \frac{1}{2\sigma(a_i)^2} \big(d_i - \mu(a_i)\big)^2 \Big)\Bigg) \Bigg(\prod_{ij\in\mathcal{N}} \exp(\alpha\delta_{a_i,a_j})\Bigg)}{\sum_{\vec{z}} \Bigg(\prod_{i\in\Omega} \Big(\frac{\gamma(z_i)}{\sigma(z_i)}\Big) \exp\Big(- \frac{1}{2\sigma(z_i)^2} \big(d_i - \mu(z_i)\big)^2 \Big)\Bigg) \Bigg(\prod_{ij\in\mathcal{N}} \exp(\alpha\delta_{z_i,z_j})\Bigg)}, \tag{14}$$

The segmented image is obtained by maximizing the posterior marginal probability

$$\mathcal{P}(A_i = a_i | \vec{D} = \vec{d}, \alpha, \vec{\gamma}, \vec{\mu}, \vec{\sigma}) \equiv \sum_{\vec{z}} \mathcal{P}(\vec{A} = \vec{a} | \vec{D} = \vec{d}, \alpha, \vec{\gamma}, \vec{\mu}, \vec{\sigma}), \tag{15}$$

at each pixel i. In the present case, we first set $\alpha = 0$ and determine the other hyperparameters $(\vec{\gamma}, \vec{\mu}, \vec{\sigma})$ so as to maximize the marginal likelihood:

$$\mathcal{P}(\vec{D} = \vec{d}|\alpha, \vec{\gamma}, \vec{\mu}, \vec{\sigma}) = \sum_{\vec{a}} \mathcal{P}(\vec{D} = \vec{d}|\vec{A} = \vec{a}, \vec{\mu}, \vec{\sigma}) \mathcal{P}(\vec{A} = \vec{a}|\alpha, \vec{\gamma}), \tag{16}$$

such that

$$(\widehat{\vec{\gamma}}, \widehat{\vec{\mu}}, \widehat{\vec{\sigma}}) = \arg \max_{(\vec{\gamma}, \vec{\mu}, \vec{\sigma})} \mathcal{P}(\vec{D} = \vec{d}|\alpha = 0, \vec{\gamma}, \vec{\mu}, \vec{\sigma}). \tag{17}$$

The above procedure to determine the estimates $(\widehat{\vec{\gamma}}, \widehat{\vec{\mu}}, \widehat{\vec{\sigma}})$ can be achieved exactly and we do not have to use the belief propagation. After that, we set α to a positive value and calculate the posterior marginal probability $\mathcal{P}(A_i = a_i|\vec{D} = \vec{d}, \alpha, \vec{\gamma} = \widehat{\vec{\gamma}}, \vec{\mu} = \widehat{\vec{\mu}}, \vec{\sigma} = \widehat{\vec{\sigma}})$ by using the belief propagation. In this scheme, we obtain the estimate of segmented image $\widehat{\vec{a}}$.

Numerical experiments for the image segmentation by combining the Gauss mixture model with the Markov random field model and the belief propagation are given in Figure 3. The result in Figure 3(c) is obtained by setting $\alpha = 0$ when we do not only the maximization of the marginal likelihood but also the calculation of the posterior marginal probability. In the result in Figure 3(d), we set $\alpha = 2$ and employ the belief propagation in the calculation of the posterior marginal probability. Details of calculations and

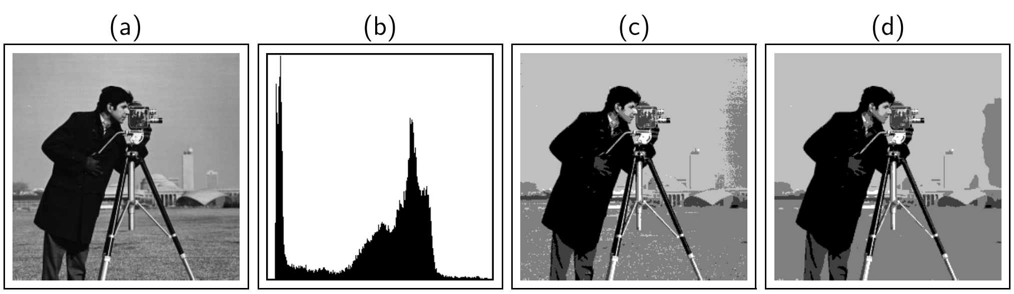

Figure 3: Image segmentation from a natural image by combining the Gaussian mixture model with the Markov random field model and the belief propagation. (a) Observed image. (b) Histogram of the observed image. (c) Segmented image for $\alpha = 0$. (d) Segmented image for $\alpha = 2$.

the other developments have been given in Ref.[27].

Moreover, we can extend the above segmentation scheme to an object detection in motion images. Figure 4 gives the object detection from three successive pictures. A man walks from left side to right side in the pictures. In the detected image in Figure 4, only the man has been detected.

6 Conclusion

In the present paper, we review the theoretical frameworks and some elementary applications of the probabilistic image processing by means of the Markov random field and the belief propagation. As elementary applications, we show some numerical experiments of the noises reduction and the segmentations.

Figure 4: Motion detection from three successive images (a), (b) and (c) by using the Gaussian mixture model, the Markov random field model and the belief propagation. The result of motion detection is given in (d).

Although we explained the intensity field as the Markov random field in the present paper, the compound Gauss-Markov random field models with both line field and intensity field are usually employed in the Bayesian image analysis. The line field is assigned to all the neraest-neighbour pairs of pixels and the random variable expresses whether an edge exists at each pair of pixels. The belief propagation can be applicable to the compound Gauss-Markov random field models. The compound Gauss-Markov random field model is powerful for the noise reduction and the edge detection. Moreover, by combing the Bayesian network with the belief propagation, it is expect that new and powerful vision algorithm can be provided[28]. As the examples, we can mention the super-resolution[13, 14, 15], the stereo vision[16, 17], the face recognition[18], the shape-time photography[19], the visual hand tracking[20], object recognition[21, 22] and so on.

On the other hand, it is important to clarify the detailed theoretical framework of the Bayesian image processing in the mathematical point of view. Particularly, the estimation of hyperparameters from the given data by means of the framework in the statistics is interesting problem, and the author investigated it in the framework for the maximization of the marginal likelihood in some elementary probabilistic models[29, 30, 31].

As one of the other important problems, we have the estimation of the statistical performance in the probabilistic inference systems. Suppose that the estimated image by means of the *a posteriori* probability $\mathcal{P}(\vec{A}=\vec{a}|\vec{D}=\vec{d},\vec{\alpha},\vec{\beta})$ is denoted by the notation $\widehat{\vec{a}}(\vec{d},\vec{\alpha},\vec{\beta})$. Let us introduce the following quantity as a measure of performance of the Bayesian network for image processing given in Section 2:

$$\sum_{\vec{a}}\Big(\sum_{\vec{d}}||\widehat{\vec{a}}(\vec{d},\vec{\alpha},\vec{\beta})-\vec{a}||^2\mathcal{P}(\vec{D}=\vec{d}|\vec{A}=\vec{a},\vec{\beta})\Big)\mathcal{P}(\vec{A}=\vec{a}|\vec{\alpha}). \tag{18}$$

This quantity can be calculated exactly in some special probabilistic models. One of them is the Gaussian graphical model and the other one is a probabilistic model given on the complete graph. The statistical performance for Gaussian graphical model can be calculated by using the multi-dimensional Gauss integral formula[23, 26] and the one for a probabilistic model given on the complete graph is obtained by using the replica method[32, 33, 34]. Although a part of these problems have come to settle, most of problems have still remained. These are the future problems.

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, October 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: "Generalized belief propagation," Advances in Neural Information Processing Systems, vol.13, pp.689-695, 2001 (Cambridge, MA: MIT Press).

[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] 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.

[14] 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.

[15] M. F. Tappen, B. C. Russell and W. T. Freeman, "Efficient graphical models for processing images," Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (27 June-2 July 2004, Washington, DC, USA), vol.2, pp.673-680, 2004.

[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] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," Proceedings of Ninth IEEE International Conference on Computer Vision (Nice, France, 13-16 October), vol.2, pp.900-906, 2003.

[18] E. B. Sudderth, A. T. Ihler, W. T. Freeman and A. S. Willsky, "Nonparametric belief propagation," Proceedings. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (18-20 June, 2003, Madison, WI, USA), vol.1, pp.605-612, 2003.

[19] W. T. Freeman and H. Zhang, "Shape-time photography," Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (18-20 June, 2003, Madison, WI, USA), vol.2, pp.151-157, 2003.

[20] E. B. Sudderth, M. I. Mandel, W. T. Freeman and A. S. Willsky, "Visual hand tracking using nonparametric belief propagation," Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (June 27 - July 02, 2004, Washington, DC, USA), pp.189-189.

[21] K. Murphy, A. Torralba, and W. Freeman. “Using the forest to see the trees: a graphical model relating features, objects and scenes,” In Advances in Neural Information Processing Systems, 2003.

[22] A. Torralba, K. P. Murphy, W. T. Freeman “Contextual models for object detection using boosted random fields,” In Advances in Neural Information Processing Systems, 2004.

[23] 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.

[24] 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.

[25] D. J. MacKay, “Bayesian interpolation,” Neural Computation, vol.4, pp.415-447, 1992.

[26] 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.

[27] F. Chen, K. Tanaka and T. Horiguchi, “Image segmentation based on Bethe approximation for Gaussian mixture model,” Interdisciplinary Information Sciences, vol.11, no.1, pp.17-29, 2005.

[28] K. Tanaka, “Staistical-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.

[29] 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.

[30] K. Tanaka, J. Inoue and D. M. Titterington, “Loopy belief propagation and probabilistic image processing,” Neural Networks for Signal Processing XIII —Proceedings of the 2003 IEEE Signal Processing Society Workshop (17-19 September, 2003, Toulouse, France), pp.329-338, IEEE Computer Society Press, 2003.

[31] K. Tanaka and D. M. Titterington, “Probabilistic image processing based on the Q-Ising model by means of the mean-field method and loopy belief propagation,” Proceedings of 17th International Conference on Pattern Recognition, vol.2 (23-26 August, 2004, Combridge, UK), pp.40-43, IEEE Computer Society Press, 2004.

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

[33] J. Inoue and K. Tanaka, “Dynamical properties of image restoration and hyper-parameter estimation,” Neural Networks for Signal Processing XI —Proceedings of the 2001 IEEE Signal Processing Society Workshop (10-14 September, 2001, Boston, USA), pp.383-392, 2001.

[34] 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.