

Discrete MRF Inference of Marginal Densities for Non-uniformly Discretized Variable Space

Masaki Saito Takayuki Okatani Koichiro Deguchi
Tohoku University, Japan
{msaito, okatani}@vision.is.tohoku.ac.jp

Abstract

This paper is concerned with the inference of marginal densities based on MRF models. The optimization algorithms for continuous variables are only applicable to a limited number of problems, whereas those for discrete variables are versatile. Thus, it is quite common to convert the continuous variables into discrete ones for the problems that ideally should be solved in the continuous domain, such as stereo matching and optical flow estimation. In this paper, we show a novel formulation for this continuous-discrete conversion. The key idea is to estimate the marginal densities in the continuous domain by approximating them with mixtures of rectangular densities. Based on this formulation, we derive a mean field (MF) algorithm and a belief propagation (BP) algorithm. These algorithms can correctly handle the case where the variable space is discretized in a non-uniform manner. By intentionally using such a non-uniform discretization, a higher balance between computational efficiency and accuracy of marginal density estimates could be achieved. We present a method for actually doing this, which dynamically discretizes the variable space in a coarse-to-fine manner in the course of the computation. Experimental results show the effectiveness of our approach.

1. Introduction

Markov Random Fields (MRFs) have been used to solve various types of problem in computer vision and image processing, such as image restoration [5, 1], super-resolution [10], stereo matching [7, 9], and optical flow estimation [12]. These problems share the same formalization, in which one optimizes the energy function consisting of the data term of each site and the smoothness term of neighboring sites:

$$E(\mathbf{x}) = \sum_i f_i(x_i) + \sum_{(i,j) \in \mathcal{E}} f_{ij}(x_i, x_j), \quad (1)$$

where $\mathbf{x} = [x_1, \dots]^T$, and x_i is the variable of site i .

There are basically two methods for inference using MRF models, MAP (Maximum A Posteriori) inference and MPM (Maximum Posterior Marginal) inference. Both are built upon the Boltzmann distribution $Q(\mathbf{x}) \propto \exp(-E(\mathbf{x}))$. MAP directly obtains the maximizer to $Q(\mathbf{x})$ and uses it as an estimate of \mathbf{x} . MPM first computes the marginal density of each variable x_i ; it then obtains its maximizer and uses it as the estimate of x_i [11, 2, 4].

In this paper, we consider the estimation of marginal densities. Although MAP is in general computationally easier to perform and thus MPM is unlikely to be the first choice when both can be used, there is no other choice when the marginal densities themselves are necessary, e.g., learning the parameters in CRF (Conditional Random Field) models [3, 6]

The computation of the marginal densities is differently formulated depending on whether the variable x_i is continuous or discrete. There are two practical algorithms, the belief propagation (BP) and the mean field (MF) algorithms. Both iteratively estimate the marginal densities by repeatedly exchanging information, or messages, among the neighboring sites. In the case of continuous variables, the marginal densities are represented by some parametric density function and its parameters are iteratively updated at each site. In the case of discrete variables, the marginal densities are represented as discrete densities, and they are iteratively updated at each site.

The former formulation for continuous variables can be used only for a small class of problems, as there are only a few choices for the parametric function representing the marginal densities. In fact, for the BP algorithm, the Gaussian function is practically an only choice. (This limitation comes from the constraint that in the message updating step, the densities before and after the update should be represented by the same parametric function.) For the MF algorithm, this limitation is somewhat relaxed but it is in general difficult to derive an iterative algorithm having good convergence property.

On the other hand, the formulation for discrete variables is free from such a limitation, and it can be used for a wide

range of problem. Thus, it is used not only for the problems originally defined in discrete domain (e.g., multi-label image segmentation) but also for those originally defined in continuous domain. In the latter case, the continuous variables are discretized into discrete ones. For example, in stereo matching and optical flow estimation, the site variable is disparity and a flow vector, respectively, which are both continuous; they are discretized and the energy function is then defined based on the resulting discrete variables.

In this paper, we present a novel formulation for this continuous-discrete conversion (i.e., that the problems that should ideally be dealt with in continuous domain are solved by discretization of the variables). In the conventional formulation, the variables are first discretized and the energy of Eq.(1) is then defined based on those discrete variables. Then, the marginal densities defined in the discrete domain are estimated using the discrete MF or BP algorithm. On the other hand, our formulation starts with the energy defined in the continuous domain. To make its minimization feasible, we “discretize” the marginal density of each site, or more rigorously, approximate the marginal density with a discrete density. We then search for the marginal densities that minimize the energy in the space of the approximating discrete densities.

For the approximating density, we choose a mixture of rectangular densities in this study. The center of each rectangular distribution corresponds to a discrete value in the conventional formulation, and its height is the parameter to be determined in the minimization. Based on this formulation, we derive the MF and BP algorithms.

In our formulation, the rectangular functions in the mixture are allowed to have arbitrary locations and sizes (as long as any two of them are not overlapped in the variable space), which provides a core practical value of our formulation. In fact, when they are placed on a regular grid and have the same size, the new MF and BP algorithms coincide with the conventional ones, whereas otherwise the two are different. To be specific, the updating terms in the new MF and BP algorithms have additional terms as compared with conventional ones; these additional terms are regarded as compensating the non-uniform distribution of rectangular functions. Note that the conventional MF and BP algorithms are independent of how the continuous variables are discretized; as the energy is defined after the discretization, differences in the discretization simply change the meaning of the energy.

This flexibility with our formulation enables the followings:

- One can discretize the variable space in a non-uniform manner (e.g., sampled densely in some region and sparsely in others) to improve the estimation accuracy of the marginal densities without increasing the com-

putational cost.

- One can deal with the case where the variable space is non-Euclidean and is difficult to uniformly discretize, e.g., spherical surface.

The former could be particularly effective for the variable space of two or higher dimensions. For effective non-uniform discretization, some prior knowledge could be used if it is available.

In this paper, taking one step further, we present a method that performs this non-uniform discretization dynamically in the course of the optimization. Our method employs a coarse-to-fine strategy; starting with coarsely divided blocks of the variable space, it recursively divides the block of the largest mixture weight into subblocks. (Each block is the support of a rectangular function in the mixture density.) This block subdivision also requires dividing the current marginal density estimates as well as the messages. We also describe how to do this.

This paper is organized as follows. In Section 2, we derive the new MF and BP algorithms that can deal with non-uniformly discretized variable space. Section 3 presents a method that dynamically discretizes the variable space in a coarse-to-fine manner, which are to be used with the new MF or BP algorithm. Section 4 shows the results of the experiments conducted to examine the effectiveness of our approach. Section 5 concludes this paper.

2. Algorithms for a non-uniformly discretized variable space

In this section we derive the new MF and BP algorithms that can deal with non-uniformly discretized variable space.

2.1. Minimization of the free energy

The conventional MF and BP algorithms for discrete and continuous variables are all derived in the same variational framework for the minimization of the free energy [13]. We briefly summarize it here, as it will be used later to derive the new algorithms

The joint density $Q(\mathbf{x})$ of the variables \mathbf{x} of all the sites of the MRF is given by

$$Q(\mathbf{x}) = \frac{1}{Z} \exp(-E(\mathbf{x})), \quad (2)$$

where Z is a normalizing factor called the partition function. The direct computation of the marginal densities based on Eq.(2) is in general infeasible. Thus, in the variational approach, a density $P(\mathbf{x})$ that approximates $Q(\mathbf{x})$ is introduced. Then, restricting P to a particular class of densities for which the marginal densities can be computed easily, P that the best approximates Q is searched for.

To evaluate the accuracy of the approximation, the following KL divergence between P and Q is used:

$$\mathcal{D}[P \parallel Q] = \sum_{\mathbf{x}} P(\mathbf{x}) \ln \frac{P(\mathbf{x})}{Q(\mathbf{x})}. \quad (3)$$

The substitution of Eq.(2) into Eq.(3) yields

$$\mathcal{D}[P \parallel Q] = \langle E \rangle_P - S[P] + \ln Z, \quad (4)$$

where $\langle E \rangle_P = \int p(\mathbf{x}) E(\mathbf{x}) d\mathbf{x}$, the expectation of the energy with respect to P , and $S[P] = - \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x}$ is the entropy of P . As the third term in Eq.(4) is independent of P , it is sufficient to consider minimization of the rest of the terms called the free energy:

$$F[P] = \langle E \rangle_P - S[P]. \quad (5)$$

In short, P that is the closest to Q in the sense of KL divergence can be found by minimizing $F[P]$.

2.2. Derivation of a new MF algorithm

The central issue of the variational approach is the choice of the class of the approximating densities (P 's). The MF algorithm is derived by choosing the following class of P 's:

$$P(\mathbf{x}) \equiv \prod_i p_i(x_i). \quad (6)$$

This means that the variable of each site is independent of that of any other site. This is in general too restrictive an assumption to accurately approximate the true density, whereas it can significantly simplify computation. Substituting Eq.(6) into Eq.(5), Eq.(5) reduces to

$$\begin{aligned} F[P] = & \sum_i \int p_i(x_i) f_i(x_i) dx_i \\ & + \sum_{(i,j) \in \mathcal{E}} \iint p_i(x_i) p_j(x_j) f_{ij}(x_i, x_j) dx_i dx_j \\ & + \sum_i \int p_i(x_i) \ln p_i(x_i) dx_i, \quad (7) \end{aligned}$$

where \mathcal{E} indicates the set of edges in the graph. Note that the first and second terms correspond to $\langle E \rangle_P$ and the third term to $S[P]$ of Eq.(5), respectively.

We wish to find P that minimizes Eq.(7) under the constraint of $\int p_i(x_i) dx_i = 1$ ($i = 1, \dots, N$). By introducing a Lagrange multiplier for this constraint and solving the Euler-Lagrange equation, we have the following fixed point equation for the unknown $p_i(x_i)$ ($i = 1, \dots, N$):

$$p_i(x_i) \propto \exp \left[- \left(f_i(x_i) + \sum_{j \in \mathcal{N}_i} \int f_{ij}(x_i, x_j) p_j(x_j) dx_j \right) \right], \quad (8)$$

where \mathcal{N}_i is the neighboring site of i -th site. The MF algorithm iteratively updates the estimate of p_i 's by using this equation; the substitution of the current estimates to the right hand side gives an updated estimate on the left hand side. The densities p_i 's after convergence directly give the estimates of the marginal densities of Q .

The above derivation is valid for both cases of continuous and discrete variables, and from here, different formulations are necessary for the two cases.

When x_i is a discrete variable, $p_i(x_i)$ is naturally a discrete density. Letting $[x^1, \dots, x^S]$ be the discrete values that x_i can take, we denote their probabilities by $[p_i^1, \dots, p_i^S]$. Then, Eq.(7) reduces to

$$\begin{aligned} F[\mathbf{p}] = & \sum_i \sum_s p_i^s f_i(x^s) \\ & + \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} p_i^s p_j^t f_{ij}(x^s, x^t) + \sum_i \sum_s p_i^s \ln p_i^s. \quad (9) \end{aligned}$$

The fixed point equation (8) is such that Eq.(7) is minimized under the constraint $\int p_i(x_i) dx_i = 1$. In the discrete case, the constraint becomes $\sum_s p_i^s = 1$ (for any i); under this constraint, Eq.(8) turns to

$$p_i^s \propto \exp \left[- \left(f_i(x^s) + \sum_{j \in \mathcal{N}_i} \sum_{t=1}^S f_{ij}(x^s, x^t) p_j^t \right) \right], \quad (10)$$

which gives the updating rule for the probabilities $[p_i^1, \dots, p_i^S]$ ($i = 1, \dots, N$).

When x_i is a continuous variable, we are to represent $p_i(x_i)$ by some parametric function such as a Gaussian distribution; Eq.(8) will then give an updating equation for the parameters. Note however that this is possible only for parametric functions such that the right hand side of Eq.(8) yields the same parametric function.

Now we present our formulation for discretizing a continuous problem. We wish to make feasible the computation for a problem originally defined in the continuous domain by discretizing the variable space. To do this, sticking to the above continuous formulation, we represent $p_i(x_i)$ by a mixture of S_i rectangular densities as

$$p_i(x_i) \equiv \sum_{s=1}^{S_i} \alpha_i^s h_i^s(x_i) \quad i = 1, \dots, N \quad (11)$$

where α_i^s is the mixing coefficient to be determined in the minimization; h_i^s is a rectangular function fixed during the minimization, which is defined as follows. Let \mathcal{X} be the variable space and d be its dimensionality. Also let \mathcal{X}_i^s be a d -dimensional hyperrectangle (i.e., the Cartesian product of intervals) such that $\mathcal{X}_i^s \cap \mathcal{X}_i^t = \emptyset$. Then, $h_i^s(x_i)$ is defined to be

$$h_i^s(x_i) = \begin{cases} 1/\mathcal{V}_i^s & \text{if } x \in \mathcal{X}_i^s \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

where \mathcal{V}_i^s is the volume of \mathcal{X}_i^s ; thus $\int h_i^s(x_i)dx_i = 1$.

Note that the rectangular functions $h_i^s(x_i)$'s may have non-uniform locations and sizes. Thus, one may distribute $h_i^1(x_i), \dots, h_i^{S_i}(x_i)$ in the variable space \mathcal{X} densely (or sparsely) for particular portions of \mathcal{X} depending on their importance. Note also that their distribution in \mathcal{X} is allowed to be different for each site, so is even S_i . Thus, one may, for example, increase or decrease S_i for particular sites (e.g., an image region) depending on their importance.

Next we derive the updating equation for α_i^s 's similar to Eq.(8). Unlike the earlier cases, it cannot be obtained by directly substituting Eq.(11) into Eq.(8), because of the above generality of our mixtures. (The right hand side of Eq.(8) cannot generally be represented by the mixture of the (i -th) site.

Thus, we trace back to the free energy of Eq.(7). By substituting Eq.(11) into Eq.(7) and introducing new notations, it reduces to

$$F[\alpha] = \sum_i \sum_s \alpha_i^s (f_i^s - B_i^s) + \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} \alpha_i^s \alpha_j^t f_{ij}^{st} + \sum_i \sum_s \alpha_i^s \ln \alpha_i^s, \quad (13)$$

where f_i^s and f_{ij}^{st} are respectively defined by

$$f_i^s = \int f_i(x_i) h_i^s(x_i) dx_i, \quad (14)$$

$$f_{ij}^{st} = \iint f_{ij}(x_i, x_j) h_i^s(x_i) h_j^t(x_j) dx_i dx_j, \quad (15)$$

which are the expectations of the data term and the smoothness term with respect to $h_i^s(x_i)$ and $h_j^t(x_j)$; B_i^s is defined by

$$B_i^s \equiv - \int h_i^s(x_i) \ln h_i^s(x_i) dx_i. \quad (16)$$

An updating equation for α_i^s 's can be directly derived from the similarity between Eq.(13) and Eq.(9). If we equate the pairs $\alpha_i^s \leftrightarrow p_i^s$, $(f_i^s - B_i^s) \leftrightarrow f_i(x^s)$ (not $f_i^s \leftrightarrow f_i(x^s)$), and $f_{ij}^{st} \leftrightarrow f_{ij}(x^s, x^t)$, the two free energies coincide with each other. Moreover, we have the same constraint for α_i^s 's as the one for p_i^s 's under which Eq.(10) is derived from Eq.(9). It is $\sum_s \alpha_i^s = 1$, which is obtained from $\int p_i(x_i) dx_i = 1$ and $\int h_i^s(x_i) dx_i = 1$. Therefore, the fixed point equation for Eq.(9) gives the one for α_i^s as

$$\alpha_i^s \propto \exp \left[- \left((f_i^s - B_i^s) + \sum_{j \in \mathcal{N}_i} \sum_{t=1}^{S_j} f_{ij}^{st} \alpha_j^t \right) \right]. \quad (17)$$

For this derivation, we need also to assume that $S_i = S$ for any i . However, the same equation can be derived for the case where S_i differs for each i ; the details are omitted for the lack of space.

The updating equation (17) has a similar form to Eq.(10). Under the natural correspondences $\alpha_i^s \leftrightarrow p_i^s$, $f_i^s \leftrightarrow f_i(x^s)$, and $f_{ij}^{st} \leftrightarrow f_{ij}(x^s, x^t)$, the only difference is the presence of B_i^s . If $h_i^1(x_i), \dots, h_i^{S_i}(x_i)$ have the same size in \mathcal{X} , then B_i^s becomes constant for any s . If so, it is invalidated in Eq.(17) and the above MF algorithm coincides with the conventional one. Therefore, B_i^s can be regarded as a compensating term for the ‘‘non-uniformity’’ of the discretization of the variable space \mathcal{X} .

2.3. Derivation of a new BP algorithm

For Belief Propagation, the following class of approximating densities P 's is considered.

$$P(\mathbf{x}) = \frac{\prod_{ij} p_{ij}(x_i, x_j)}{\prod_i p_i(x_i)^{z_i-1}}, \quad (18)$$

where z_i is the number of neighboring sites of the i -th site; $p_i(x_i)$ and $p_{ij}(x_i, x_j)$ satisfy

$$\int p_i(x_i) dx_i = 1, \quad \iint p_{ij}(x_i, x_j) dx_i dx_j = 1, \\ \int p_{ij}(x_i, x_j) dx_i = p_j(x_j). \quad (19)$$

This density class has more generality than that for MF (Eq.(6)), and thus the marginal densities estimated by BP tend to be more accurate than MF.

Similarly to the MF algorithm, substituting Eq.(18) into Eq.(5), we have

$$F[P] = \sum_i \int p_i(x_i) f_i(x_i) dx_i + \sum_{(i,j) \in \mathcal{E}} \iint p_{ij}(x_i, x_j) f_{ij}(x_i, x_j) dx_i dx_j - \sum_i (z_i - 1) \int p_i(x_i) \ln p_i(x_i) dx_i + \sum_{(i,j) \in \mathcal{E}} \iint p_{ij}(x_i, x_j) \ln p_{ij}(x_i, x_j) dx_i dx_j. \quad (20)$$

In the conventional discrete formulation, the BP algorithm is derived as follows. We denote the discrete values that x_i takes by $[x^1, \dots, x^S]$ and their probabilities by $[p_i^1, \dots, p_i^S]$ (i.e., $p_i^s \equiv p(x_i = x^s)$). We also define $p_{ij}^{st} \equiv p_{ij}(x_i = x^s, x_j = x^t)$. Rewriting Eq.(20) with the newly defined variables, we have

$$F[\mathbf{p}] = \sum_i \sum_s p_i^s f_i(x^s) + \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} p_{ij}^{st} f_{ij}(x^s, x^t) - \sum_i (z_i - 1) \sum_s p_i^s \ln p_i^s + \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} p_{ij}^{st} \ln p_{ij}^{st}, \quad (21)$$

where \mathbf{p} contain all p_i^s 's and p_{ij}^{st} 's. The constraints on $p_i(x_i)$ and $p_{ij}(x_i, x_j)$ reduce to

$$\sum_s p_i^s = 1, \quad \sum_{s,t} p_{ij}^{st} = 1, \quad \sum_s p_{ij}^{st} = p_j^t. \quad (22)$$

By minimizing $F[\mathbf{p}]$ under these constraints, we have the discrete BP algorithm that iteratively updates the messages m_{ij}^t according to

$$m_{ij}^t \leftarrow \sum_s \phi_i^s \psi_{ij}^{st} \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}^s, \quad (23a)$$

where

$$\phi_i^s = \exp[-f_i(x^s)], \quad (23b)$$

$$\psi_{ij}^{st} = \exp[-f_{ij}(x^s, x^t)]. \quad (23c)$$

In our formulation, we use the same mixture of rectangular densities for representing $p_i(x_i)$ and $p_{ij}(x_i, x_j)$. To be specific, we represent $p_i(x_i)$ and $p_{ij}(x_i, x_j)$ as

$$p_i(x_i) = \sum_{s=1}^{S_i} \alpha_i^s h_i^s(x_i), \quad (24a)$$

$$p_{ij}(x_i, x_j) = \sum_{s=1}^{S_i} \sum_{t=1}^{S_j} \alpha_{ij}^{st} h_i^s(x_i) h_j^t(x_j). \quad (24b)$$

By substituting these into Eqs.(19), from $\int h_i^s(x_i) dx_i = 1$ we have

$$\sum_s \alpha_i^s = 1, \quad \sum_{s,t} \alpha_{ij}^{st} = 1, \quad \sum_s \alpha_{ij}^{st} = \alpha_j^t. \quad (25)$$

These coincide with Eqs.(22).

Substituting Eqs.(24) to Eq.(20) and simplifying the result using B_i^s defined in Eq.(16), Eq.(20) reduces to

$$\begin{aligned} F[\alpha] = & \sum_i \sum_s \alpha_i^s (f_i^s + (z_i - 1)B_i^s) \\ & + \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} \alpha_{ij}^{st} (f_{ij}^{st} - B_i^s - B_j^t) \\ & - \sum_i (z_i - 1) \sum_s \alpha_i^s \ln \alpha_i^s + \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} \alpha_{ij}^{st} \ln \alpha_{ij}^{st}. \quad (26) \end{aligned}$$

Comparing this with Eq.(21), it is seen that the former coincides with the latter if we equate the following four pairs: $\alpha_i^s \leftrightarrow p_i^s$, $\alpha_{ij}^{st} \leftrightarrow p_{ij}^{st}$, $f_i^s + (z_i - 1)B_i^s \leftrightarrow f_i(x^s)$, and $f_{ij}^{st} - B_i^s - B_j^t \leftrightarrow f_{ij}(x^s, x^t)$. Moreover, we have the same constraints for α_i^s and α_{ij}^{st} as those for p_i^s and p_{ij}^{st} given in Eqs.(22), under which the message updating rule of Eqs.(23) are derived from Eq.(21). Therefore, by performing the above four substitution on Eqs.(23), we have the new message updating rule for our formulation, which is given by (the same as Eq.(23a))

$$m_{ij}^t \leftarrow \sum_s \phi_i^s \psi_{ij}^{st} \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}^s, \quad (27a)$$

where ϕ_i^s and ψ_{ij}^{st} are differently calculated as

$$\phi_i^s = \exp[-(f_i^s + (z_i - 1)B_i^s)], \quad (27b)$$

$$\psi_{ij}^{st} = \exp[-(f_{ij}^{st} - B_i^s - B_j^t)]. \quad (27c)$$

From these, the mixture weights are computed as

$$\alpha_i^s \propto \phi_i^s \prod_{k \in \mathcal{N}_i} m_{ki}^s. \quad (28)$$

As mentioned earlier, if $h_i^1(x_i), \dots, h_i^{S_i}(x_i)$ have the same size in \mathcal{X} , then B_i^s becomes constant. If so, all the terms associated with B_i^s are invalidated in the above updating equations and then the above BP algorithm coincides with the conventional one. Therefore, similarly to MF, B_i^s can be regarded as a compensating factor for the non-uniformity of the discretization.

3. Dynamic discretization of the variable space

3.1. Usefulness of non-uniform discretization

The new MF and BP algorithms can deal with non-uniformly discretized variable space. By densely discretizing important portion of the space and sparsely discretizing the rest and then using these algorithms, we will be able to achieve higher balance between computational efficiency and accuracy of marginal density estimates. When the variable space is of two or higher dimensions, this effect will be significant; it is particularly so for the BP algorithms, in which computational cost is mainly determined by the number of labels. (When the dimensionality of the variable space is D and the number of labels is L , the computational cost of BP is proportional to L^{2D} .)

The next question is how to obtain such an effective discretization of the variable space. If we have a prior knowledge about where is more important in the variable space, it will be possible to use it to obtain a good discretization. For the case where no such knowledge is available, we present a method for dynamically discretizing the variable space to have an effective discretization.

3.2. Coarse-to-fine block subdivision

We assume here that behind the estimation of the marginal density, there is a motivation to accurately know its shape around its maximum, e.g., to determine the position of its maximum as accurately as possible. Then, this will be made possible by more densely discretizing the space around the maximum of the marginal density.

As it is in general impossible to know the maximum of the marginal density beforehand, we consider dynamically dividing the variable space, as shown in Fig.1. We start with initial coarse discretization of the variable space, that is, the variable space is divided into a small number of blocks. The

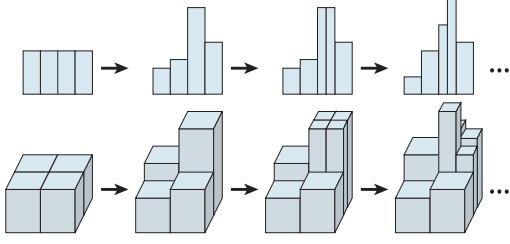


Figure 1. Dynamic discretization of the variable space. Each block indicates the support of a rectangular density composing the mixture approximating the true marginal density. The block having the largest weight is divided into subblocks.

rectangular function whose support is each block composes the mixture density approximating the true marginal density. For this discretization, the MF or BP algorithm is run for a certain iterations. Then, for each site (i), identifying the block (s) whose mixture weight α_i^s is the largest, we divide this block into a number of subblocks. (Multiple blocks with the largest weights may be divided simultaneously.) Then, integrating these new blocks with the blocks that are not divided, we consider a new mixture of rectangular functions whose supports are given by them. We repeatedly perform these three procedures for a desired number of iterations: updating the mixture weights for the current discretization by our MF or BP algorithm, identifying the block(s) with the largest weight(s), and dividing them into subblocks to obtain a new discretization.

3.3. Dividing a rectangular density

The subdivision of a block means dividing the corresponding rectangular density into multiple rectangular densities, as shown in Fig.1. Thus, the mixture of rectangular densities after the subdivision has a different representation from the one before it. Corresponding to this representation change, we need to update α_i^s 's in MF and m_{ij}^s 's in BP. The principle of updating these parameters is that the mixtures before and after the subdivision should be the same density regardless of their difference in representation. Based on this principle, these parameters before the subdivision are processed and transferred to those after the subdivision. Different procedures are necessary for MF and BP.

The procedure for MF is as follows. Suppose that a rectangular density is divided into K rectangular densities of an identical size. Following the above principle, the mixture weights of the new densities are given by the weight of the original density divided by K . This ensures that the mixtures before and after the subdivision have the same shape. Suppose, for example, that the s -th block of the i -th site is divided into two blocks. Denoting S_i pre-division weights by $[\alpha_i^1, \dots, \alpha_i^{s-1}, \alpha_i^s, \alpha_i^{s+1}, \dots, \alpha_i^{S_i}]$, the new weights are of $S_i + 1$ long and is given by

$$[\alpha_i^1, \dots, \alpha_i^{s-1}, \alpha_i^s/2, \alpha_i^s/2, \alpha_i^{s+1}, \dots, \alpha_i^{S_i}].$$

The procedure for BP is as follows. Suppose that we want to perform the subdivision at the i -th site. To do this, we first compute α_i^s 's based on Eq.(28) by using all the messages passed to this site, i.e., $\{m_{ki}^s \mid k \in \mathcal{N}_i\}$. Using these weights, we then perform the block subdivision of the variable space as described above. In the case of MF, the mixture weights $[\alpha_i^1, \dots, \alpha_i^{S_i}]$ are manipulated so as to reflect the subdivision. We apply the same manipulation to $[m_{ki}^1, \dots, m_{ki}^{S_i}]$ for each $k \in \mathcal{N}_i$.

4. Experimental results

To examine the effectiveness of the proposed methods, we conducted experiments.

4.1. Effect of non-uniform discretization on marginal density estimates

To compare the behaviours of the conventional and proposed algorithms when the variable space is non-uniformly discretized, we consider a simple Gaussian MRF for which the exact marginal densities can be analytically obtained. To be specific, we consider a MRF model defined on a 5×5 grid graph that has the following energy:

$$E(\mathbf{x}) = \sum_i x_i^2 + \sum_{(i,j) \in \mathcal{E}} (x_i - x_j)^2. \quad (29)$$

Clearly, its marginal densities are Gaussian distributions having zero mean.

For this MRF model, we divide the variable space in an asymmetric way with respect to the origin $x = 0$ of the space. To be specific, considering only the range of $[-2, 2]$, we discretize its negative part $[-2, 0]$ into 64 blocks and the positive one $[0, 2]$ into 16 blocks. Thus, the continuous MRF is converted into a discrete MRF with 80 labels in total.

Then we apply the conventional MF and BP algorithms and the proposed MF and BP algorithms to this discrete MRF. Figs. 2 and 3 show the results of the MF and BP algorithms, respectively. They show the estimates of the marginal density at the site in the upper-left corner of the 5×5 grid graph. The estimates by the conventional algorithms are shown by red dots; those by the proposed algorithms are shown by blue histograms; the exact marginal densities are shown by a continuous red curve. In the conventional algorithms, a marginal density is represented as a discrete density, i.e., $[p_i^1, \dots, p_i^S]$. In the plots, to enable direct comparison with densities in the continuous domain, its scale (i.e., the heights of the red dots) is appropriately adjusted. In the proposed algorithms, the marginal densities are represented as the mixtures of rectangular densities, which are shown in the plots.

It is seen from Figs.2 and 3 that the estimated marginal densities by the conventional MF and BP algorithms both

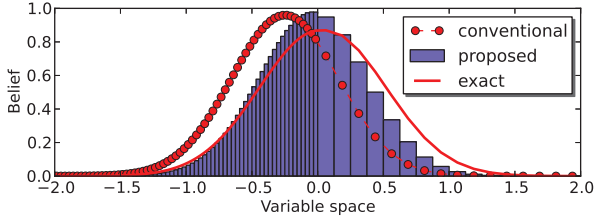


Figure 2. The results of the conventional and proposed MF algorithms. The red dots indicate the marginal density estimate by the conventional MF; the blue histogram indicates those by the proposed MF; the continuous red curve indicates the exact marginal density.

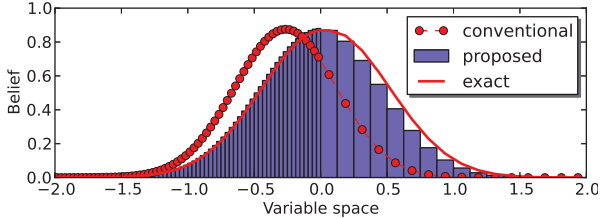


Figure 3. The results of the conventional and proposed BP algorithms. Legends are the same as Fig.2.

have bias; their means deviate from the true mean (i.e., $x = 0$) toward the side of $x < 0$. This is because of the asymmetric discretization; the energy tends to have a lower value when the marginal density estimates are in the side of denser discretization. On the other hand, the proposed MF and BP algorithms both yield more accurate estimates. The result of MF still has a bias but it is much smaller than the conventional one. The result of BP is even more accurate. Although there appears to exist small bias in the variances of the marginal density estimates, this is a fundamental limitation of these algorithms; even in the case of symmetric discretization, the MF and BP algorithms cannot estimate the exact value of the variance.

4.2. Stereo matching

We applied the proposed dynamic discretization method to stereo matching and examined its effectiveness. To generate an energy function, we used the Middlebury MRF library [8]. We set $|L| = 128$, $\lambda = 2$, $\text{smoothmax}=20$, and $\text{truncated} = 2$. We multiply the values of the data and smoothness terms given by the library by $1/10$, as otherwise, the marginal densities will have very sharp peaks, which is not fit for the purpose of this experiment.

The dynamic discretization method is applied to the data as follows. Initially dividing the variable space into eight blocks of an identical width, we iterate the following three steps for eight times: performing the MF or BP algorithm, identifying the block of the largest weight, and dividing the block into two blocks. At each of the eight iterations, the MF or BP update is iterated for 100 times. In the experi-

ment, we set the lower bound of the block size to be 1 for ease of implementation. Thus, if the block with the largest weight has reached this lower bound, we divide the one with the second largest weight. If it has reached the bound, then we divide the next largest one, and so on.

The above recursive subdivision increases the number of blocks from initial eight to sixteen ($= 8 + 8$). Figs.4 and 5 show the initial, intermediate, and final results when the number of blocks is 8, 10, 12, 14, and 16, for the *Aloe* dataset (641×555 pixels). For the sake of comparison, each figure also shows the result obtained with a fixed, uniform discretization; it is obtained by our MF or BP algorithm after 1000 iterations, when the variable space of the range $[0, 128]$ is divided into 16 blocks.

For both results, it is seen that the mixture density depicts the marginal density in a finer way with the increasing number of blocks. Note that the horizontal axes correspond to a portion of the full range $[0, 128]$. (The block sizes in the case of eight and sixteen block divisions are $128/8 = 16$ and $128/16 = 8$, respectively.) As compared with the mixture densities of the fixed discretization, those of the dynamic discretization draw much finer details not only at the same number of blocks (i.e., 16) but even at the smaller number of blocks. As a result, the maxima of the marginal densities can be determined much more accurately, and thus the dynamic discretization yields smoother disparity maps. Clearly, suffering from the insufficient number of divisions, the disparity maps for the fixed discretization are not smooth.

5. Summary

We have described a novel formulation of continuous-discrete conversion for the inference of marginal densities based on MRF models. In the formulation, the marginal densities are estimated in the continuous domain by approximating them with mixtures of rectangular densities. Based on this formulation, we have derived the MF and BP algorithms, which can correctly deal with the non-uniform discretization of variable space. We have also shown the method for dynamically discretizing the variable space in a coarse-to-fine manner in the course of the computation. This enables to improve the accuracy of marginal density estimates without sacrificing computational efficiency. We have shown several experimental results proving the effectiveness of our approach.

Acknowledgement This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas "Shitsukan" (No. 23135501) and JSPS KAKENHI Grant Number 2230057.

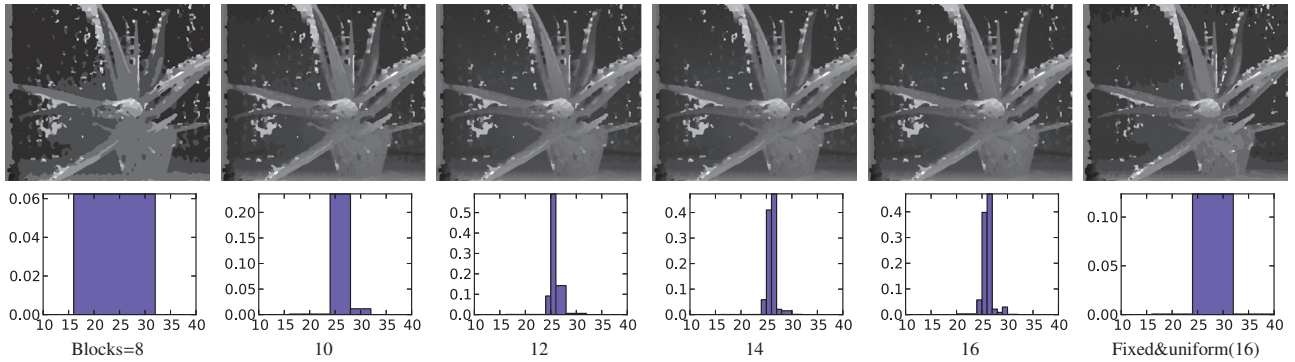


Figure 4. Results of the MF algorithm with the dynamic discretization. Upper row: Disparity maps. Lower row: The marginal density estimates at the site of the image pixel (100, 100).

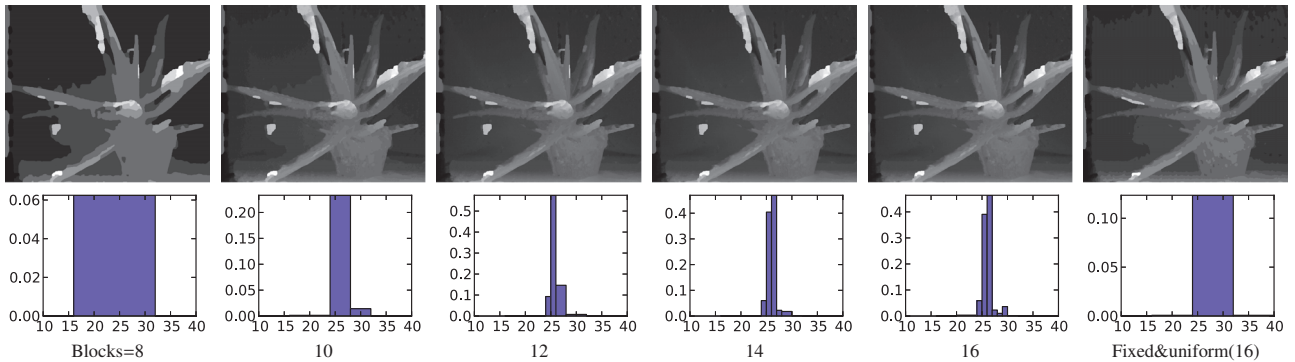


Figure 5. Results of the BP algorithm with the dynamic discretization. Upper row: Disparity maps. Lower row: The marginal density estimates at the site of the image pixel (100, 100).

References

- [1] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI*, 6:721–741, 1984.
- [2] P. Kohli and P. H. S. Torr. Measuring uncertainty in graph cut solutions. *Computer Vision and Image Understanding*, 112(1):30–38, 2008.
- [3] P. Krähenbühl and V. Koltun. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In *NIPS*, pages 109–117. 2011.
- [4] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition, 2009.
- [5] S. Roth and M. J. Black. Fields of Experts. *IJCV*, 82(2):205–229, 2009.
- [6] R. Salakhutdinov and G. E. Hinton. Deep Boltzmann Machines. In *International Conference on Artificial Intelligence and Statistics(AISTATS)*, volume 5, pages 448–455, 2009.
- [7] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo Matching Using Belief Propagation. *PAMI*, 25(7):787–800, 2003.
- [8] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, A. Agarwala, and C. Rother. A comparative study of energy minimization methods for Markov random fields. In *ECCV*, pages 16–29, 2006.
- [9] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *ICCV*, pages 900–907, 2003.
- [10] M. F. Tappen, B. C. Russell, and W. T. Freeman. Exploiting the Sparse Derivative Prior for Super-Resolution and Image Demosaicing. In *IEEE Workshop on Statistical and Computational Theories of Vision at ICCV*, pages 900–907, 2003.
- [11] D. Tarlow and R. P. Adams. Revisiting Uncertainty in Graph Cut Solutions. In *CVPR*, pages 2440–2447, 2012.
- [12] L. Xu, J. Jia, and Y. Matsushita. Motion Detail Preserving Optical Flow Estimation. In *CVPR*, pages 1744–1757, 2010.
- [13] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Understanding Belief Propagation and Its Generalizations. In *Exploring artificial intelligence in the new millennium*, chapter 8, pages 239–269. Morgan Kaufmann, 2003.