Application of the Mean Field Methods to MRF Optimization in Computer Vision

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

Abstract

The mean field (MF) methods are an energy optimization method for Markov random fields (MRFs). These methods, which have their root in solid state physics, estimate the marginal density of each site of an MRF graph by iterative computation, similarly to loopy belief propagation (LBP). It appears that, being shadowed by LBP, the MF methods have not been seriously considered in the computer vision community. This study investigates whether these methods are useful for practical problems, particularly MPM (Maximum Posterior Marginal) inference, in computer vision. To be specific, we apply the naive MF equations and the TAP (Thouless-Anderson-Palmer) equations to interactive segmentation and stereo matching. In this paper, firstly, we show implementation of these methods for computer vision problems. Next, we discuss advantages of the MF methods to LBP. Finally, we present experimental results that the MF methods are well comparable to LBP in terms of accuracy and global convergence; furthermore, the 3rd-order TAP equation often outperforms LBP in terms of accuracy.

1. Introduction

Markov Random Fields have been applied to solve many problems in the fields of computer vision and image processing. The applications include image restoration [20, 4], super-resolution [24], stereo matching [19, 23], and optical flow estimation [26]. 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:

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

where $\boldsymbol{x} = [x_1, \ldots]^\top$, and x_i is the variable of site i.

Estimation of $\boldsymbol{x}$ is formulated as a statistical inference problem under the assumption that the Boltzmann distribution specified by $E(\boldsymbol{x})$ gives the probability density of $\boldsymbol{x}$. There are two methods for solving this inference problem. One is the MAP (Maximum A Posteriori) inference, which seeks the maximum likelihood estimate; the other is the MPM (Maximum Posterior Marginal) inference, which calculates the marginal density of each variable x_i (and uses, say, its maximum likelihood value, as the estimate of x_i).

In this paper, we study only the MPM inference. Although it is in general considered to be more difficult to perform than the MAP inference, it is necessary or desirable in some applications to estimate the marginal density of each x_i, not x_i itself; moreover, even when we are only interested in the estimation of $\boldsymbol{x}$, the MPM inference often outperforms the MAP inference in terms of estimation accuracy [8].

For the MAP inference, there exist several powerful methods, such as graph cuts [23] and Tree-reweighted BP [10], which have been widely used in the computer vision community. On the other hand, for the MPM inference, there is in practice no effective method other than the original LBP (Loopy Belief Propagation) and at best its modestly efficient implementation [3] which performs the computation in a coarse-to-fine manner. (There is also the method that can compute the rigorous solutions of the marginal densities based on the graph theory [8], whose computational complexity is so large that it cannot be used in practical applications.)

However, there is another approach that calculates marginal densities in a similar way to LBP: the mean field (MF) methods. These methods have their root in the field of solid state physics, where they were developed to model and simulate phase transitions and other critical phenomena of ferromagnetic materials. These methods were once applied to several computer vision/image processing problems around early 90's [11, 27], but it did not become a mainstream method. As far as we are aware of, the methods have not been reconsidered in the computer vision community since then. One of a few exceptions is [25], where BP and a naive MF equation are compared, but the comparison is made only for tree-structured MRFs. Another is [6], where it is pointed out that the convolutional networks have a connection to a MF method, but the recent developments of the MF methods that will be studied in this paper are not considered.

Meanwhile, in other fields such as neural network and information theory, significant developments have been made on the MF methods [14, 7, 2, 22]. One is a variational

framework of reinterpreting them from the viewpoint of the minimization of free energy. In this framework, it can be understood that LBP and MF both search for an approximate solution to the true minimizer of a free energy using different approximation methods. Furthermore, the method of the TAP (Thouless-Anderson-Palmer) equations [16, 14], which were developed to improve the accuracy of the standard MF equations, have been given a new theoretical basis.

In this paper, we propose to use these MF methods, i.e., the naive MF equations and the TAP equations, for the problems of computer vision. Their real and potential advantages to LBP are:

- More flexibility in the choice of MRF models. The probability densities that can be dealt with are not limited to the Gaussian distribution or discrete representations.
- Faster computation. They are faster than at least the naive implementation of LBP. They are also suitable for parallel computing.
- Improvement of estimation accuracy. Although a naive MF equation is less accurate then LBP, high-order TAP equations could outperform LBP.

Although the first item is the most attractive among those listed above, its investigation is left for a subsequent paper. This paper focuses on examining whether the MF methods are really useful for practical computer vision problems. Toward this ends, we consider the problems of discrete label classification, for which LBP is mainly used.

Our contribution is summarized as follows:

- We show detailed implementations of the MF methods including the TAP equations for practical computer vision problems.
- We present experimental results that the MF methods show performance comparable to LBP and sometimes even better.

This paper is organized as follows. Section 2 shows the derivations of several MF methods including the MF equation for multi-label classification. Section 3 discusses the advantages of the MF methods to LBP. Section 4 presents experimental results for selected computer vision problems. Section 5 concludes this study.

2. Derivation of MF and TAP equations

This section summarizes the derivation of MF methods, i.e., the naive MF equation and the TAP equations.

2.1. Minimization of the free energy

The naive MF equation is classically derived by operations on the expectations of the variables [11]. A modern approach is to formulate the problem as minimization of free energy based on the variational principle [14]. This approach can naturally be extended to the cases of multiple states (labels) and continuous variables, and also makes it possible to compare the MF methods and LBP within the same framework. In this section, we summarize the literature [14, 15] in order to make clear their comparisons described later.

To begin with, we represent the joint density (i.e., the Boltzmann distribution) of all sites of the MRF associated with the energy of Eq.(1) as

$$Q(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-\frac{1}{T} E(\boldsymbol{x})\right), \tag{2}$$

where Z is the normalizing factor called a partition function and T is temperature. Introducing an arbitrary density $P(\boldsymbol{x})$ having some desired property, we wish to find P that best approximates Q. Toward this goal, we consider the KL distance between them:

$$D_{\mathrm{KL}}(P \,\|\, Q) = \sum_{\boldsymbol{x}} P(\boldsymbol{x}) \ln\left(\frac{P(\boldsymbol{x})}{Q(\boldsymbol{x})}\right). \tag{3}$$

The substitution of Eq.(2) into the above yields

$$D_{\mathrm{KL}}(P \,\|\, Q) = \frac{1}{T} \langle E \rangle_P - S_P + \ln Z, \tag{4}$$

where $\langle E \rangle_P$ and S_P is the expectation of the energy and the entropy of P, respectively. We can neglect the third term on the right hand side, since T and Z are nonnegative constants independent of P. The sum of the remaining terms

$$\begin{aligned} F(P) &\equiv \frac{1}{T} \langle E \rangle_P - S_P \\ &= \frac{1}{T} \sum_x P(\boldsymbol{x}) E(\boldsymbol{x}) + \sum_x P(\boldsymbol{x}) \ln P(\boldsymbol{x}) \end{aligned} \tag{5}$$

is called the free energy. Thus, minimizing the KL distance is equivalent to minimizing the free energy $F(P)$.

2.2. Naive MF equations

Following the above, the naive MF equations are derived as follows. The fundamental assumption here is the independence of x_i of each site. We then define an approximate joint density $P(\boldsymbol{x})$ by

$$P(\boldsymbol{x}) = \prod_i p_i(x_i). \tag{6}$$

Although this assumption is generally not true, the strategy is to select P that best approximates Q from the above class of P. By substituting Eq.(6) into Eq.(5), $F(p)$ can be written as

$$\begin{aligned} F(P) = &\frac{1}{T} \sum_i \sum_{x_i} p_i(x_i) f_i(x_i) \\ &+ \frac{1}{T} \sum_{(i,j) \in \mathcal{E}} \sum_{x_i} \sum_{x_j} p_i(x_i) p_j(x_j) f_{ij}(x_i, x_j) \\ &+ \sum_i \sum_{x_i} p_i(x_i) \ln p_i(x_i). \end{aligned} \tag{7}$$

2.2.1 Binary classification

An MF equation for binary classification problems is derived as follows. Assuming each site to have a binary state (label) $x_i \in \{-1, +1\}$, we define the energy function $E(\boldsymbol{x})$ as

$$E(\boldsymbol{x}) = -\sum_{(i,j)\in\mathcal{E}} x_i x_j J_{ij} - \sum_i h_i x_i, \tag{8}$$

where $J_{ij} > 0$ and $h_i \in \mathbb{R}$ are constants corresponding to the smoothness term and the data term, respectively.

In this case, the marginal density $p_i(x_i)$ can be represented by a single continuous variable m_i which has a value within $[-1, 1]$ as

$$p_i(x_i = +1) \equiv \frac{1 + m_i}{2}, \quad p_i(x_i = -1) \equiv \frac{1 - m_i}{2}. \tag{9}$$

This variable m_i corresponds to the expectation of x_i, i.e., $m_i = p_i(x_i = +1) - p_i(x_i = -1)$. Substituting Eq.(8) and Eq.(9) into Eq.(7), we can rewrite $F(P)$ as a function only of $\boldsymbol{m} = [m_1, \cdots]$. Since we want $\boldsymbol{m}$ minimizing F, we differentiate this function with respect to $\boldsymbol{m}$ to obtain

$$m_i = \tanh\left[\frac{1}{T}\left(h_i + \sum_{j\in\mathcal{N}_i} J_{ij} m_j\right)\right], \tag{10}$$

where $\mathcal{N}_i$ is the set of the neighboring sites of site i. The minimizing solution $\boldsymbol{m}$ should satisfy the above equation for all i's. A simple approach to obtain such $\boldsymbol{m}$ is iteratively updating $\boldsymbol{m}$ according to the above equation; the current estimates are set to m_j's on the right hand side and the new estimate is given by m_i on the left hand side. After convergence, p_i is computed from m_i by Eq.(9).

Algorithm 1 Algorithm for binary classification

1: **for all** i **do**
2: $\quad m_i^0 \leftarrow \tanh(h_i/T)$
3: **end for**
4: **repeat**
5: $\quad$ **for all** i **do**
6: $\quad\quad m_i^{t+1} \leftarrow \tanh[(h_i + \sum_{j\in\mathcal{N}_i} J_{ij} m_j^t)/T]$
7: $\quad$ **end for**
8: **until** convergence

2.2.2 Multi-label classification

An MF equation for multi-label classification problems is derived as follows. Let n be the number of labels and μ represent a label, i.e., $\mu \in \{1, 2, \ldots, n\}$. For simplicity, we denote the probability that site i has label μ by $p_\mu^i \equiv p_i(\mu)$; we also denote $f_i(x_i)$ of Eq.(5) when site i has label μ by $\epsilon_\mu^i \equiv f_i(\mu)$, and $f_{ij}(x_i, x_j)$ when the pair $(i, j) \in \mathcal{E}$ of sites has the labels (μ, ν) by $\epsilon_{\mu\nu}^{ij} \equiv f_{ij}(\mu, \nu)$. Then, the free energy is written as

$$F(P) = \frac{1}{T}\sum_i \sum_\mu p_\mu^i \epsilon_\mu^i + \frac{1}{T}\sum_{(i,j)\in\mathcal{E}} \sum_\mu \sum_\nu p_\mu^i p_\nu^j \epsilon_{\mu\nu}^{ij} + \sum_i \sum_\nu p_\mu^i \ln p_\mu^i.$$

To minimize $F(P)$ subject to a constraint $\sum_\mu p_\mu^i = 1$, we define $J \equiv F(P) + \sum_i \lambda_i(\sum_\mu p_\mu^i - 1)$, where λ_i is a Lagrange multiplier. Differentiating J with respect to p_μ^i and equating the resulting expression to zero, we have

$$p_\mu^i = \frac{1}{Z_i}\exp\left[-\frac{1}{T}\left(\epsilon_\mu^i + \sum_{j\in\mathcal{N}_i}\sum_\nu p_\nu^i \epsilon_{\mu\nu}^{ij}\right)\right], \tag{11}$$

where Z_i contains λ_i and can be regarded as a normalizing factor; Thus, we have

$$Z_i = \sum_\mu \exp\left[-\frac{1}{T}\left(\epsilon_\mu^i + \sum_{j\in\mathcal{N}_i}\sum_\nu p_\nu^i \epsilon_{\mu\nu}^{ij}\right)\right]. \tag{12}$$

Eq.(11) is the MF equation for multi-label classification. This can be used to iteratively update p_μ^i as in the case of binary classification. The algorithm is given as follows.

Algorithm 2 Algorithm for multi-label classification

1: **for all** i and μ **do**
2: $\quad p_\mu^i \leftarrow \exp(-\epsilon_\mu^i/T)/\sum_\mu \exp(-\epsilon_\mu^i/T)$
3: **end for**
4: **repeat**
5: $\quad$ **for all** i and μ **do**
6: $\quad\quad$ Update p_μ^i according to Eqs.(11) and (12)
7: $\quad$ **end for**
8: **until** convergence

2.3. The TAP (Thouless-Anderson-Palmer) equations

As shown above, the naive MF equations are derived by assuming the independence of the sites and representing $P(\boldsymbol{x})$ by Eq.(6). However, the resulting estimate will usually have errors due to this assumption. The TAP equations, which are similar updating equations to naive MF equations, are derived without assuming the independence of the sites [14].

The basic idea of their derivation is as follows. The free energy F is represented as a function of $P(\boldsymbol{x})$ as in Eq.(5). Since our interest is in estimating the marginal density $p_i(x_i)$ of each site, we consider representing F as a function of $p_i(x_i)$. Specifying $p_i(x_i)$ of each site, we consider a class of joint densities $P(\boldsymbol{x})$'s that are consistent with the specified $p(x_i)$'s; note that $P(\boldsymbol{x})$ is not unique for the specified $p(x_i)$'s. We then search for $P(\boldsymbol{x})$ that minimizes F in this class. The problem is formulated as a constrained minimization problem, which is solvable in some cases.

The case of binary-label classification is as follows. In this case, the marginal density $p_i(x_i)$ of site i is specified by the expectation m_i of x_i, as shown above. Thus, when all the marginal densities, i.e., the expectation $\boldsymbol{m} = [m_1, \ldots]^\top$, are specified, a consistent $P(\boldsymbol{x})$ should satisfy a constraint that the expectation $\langle \boldsymbol{x} \rangle_P$ of $\boldsymbol{x}$ with respect to P is equal to $\boldsymbol{m}$, and vice versa. Therefore, introducing a Lagrange

multiplier $\boldsymbol{\lambda} = [\lambda_1, \ldots]^\top$, the problem is rewritten as the following constrained minimization:

$$\min_{P,\boldsymbol{\lambda}} F(P) - \sum_i \lambda_i(\langle x \rangle_i - m_i), \tag{13}$$

The minimizing P is calculated as

$$P(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\lambda})} \exp\left(-\beta E(\boldsymbol{x}) + \sum_i \lambda_i x_i\right), \tag{14}$$

where $Z(\boldsymbol{\lambda})$ is a partition function, i.e., the normalizing factor, and $\beta \equiv 1/T$. By using the relation of a partition function to the minimum free energy , namely $F = -\ln Z$, F can be rewritten as

$$F(\boldsymbol{m}, \boldsymbol{\lambda}) = \sum_i \lambda_i m_i - \ln \sum_{\boldsymbol{x}} \exp\left(-\beta E(\boldsymbol{x}) + \sum_i \lambda_i x_i\right) \tag{15}$$

Now, we minimize F with respect to $\boldsymbol{m}$ and $\boldsymbol{\lambda}$. We use here the fact that $E(\boldsymbol{x})$ vanishes when $\beta = 0$. When $\beta = 0$, the minimizing $\boldsymbol{\lambda}$ is simply given by the relation

$$m_i = \tanh(\lambda_i). \tag{16}$$

We consider a Taylor series expansion of Eq.(15) with respect to β around $\beta = 0$, i.e.,

$$F = F_0 + \beta F_1 + \frac{\beta^2}{2} F_2 + \frac{\beta^3}{6} F_3 + \cdots, \tag{17}$$

where $F_n = \partial^n F / \partial \beta^n|_{\beta=0}$. Combining this expansion with Eq.(16), we can express F as a function only of $\boldsymbol{m}$, i.e., $F = F(\boldsymbol{m})$

By differentiating $F(\boldsymbol{m})$ with $\boldsymbol{m}$, we have as many nonlinear equations as the number of sites. The equation for site i is given as

$$\begin{aligned} m_i = \tanh\Big[&\beta h_i + \beta \sum_{j \in \mathcal{N}_i} J_{ij} m_j - \beta^2 \sum_{j \in \mathcal{N}_i} J_{ij}^2 m_i (1 - m_j^2) \\ &+ \frac{2\beta^3}{3} \sum_{j \in \mathcal{N}_i} J_{ij}^3 (1 - 3m_i^2) m_j (1 - m_j^2) \\ &- 2\beta^3 \sum_{(j,k) \in \mathcal{N}_i^2} J_{ij} J_{jk} J_{ki} m_i (1 - m_j^2)(1 - m_k^2) + O(\beta^4)\Big], \end{aligned} \tag{18}$$

where $(j,k) \in \mathcal{N}_i^2$ represents all possible pairs of neighboring sites of site i. The above is called the TAP equation. Since it has the form of self-consistency equations, updating $\boldsymbol{m}$ according to this equation constitutes an iterative algorithm similar to naive MF.

The order of the terms on the right hand side of Eq.(18) directly corresponds to the Taylor series expansion of F, which theoretically implies that more accurate solution will be obtained by using higher-order terms. It is an interesting coincidence that the naive MF equation is equivalent to the first-order TAP equation.

Since the above method is based on the Taylor series expansion with small β, it will be theoretically effective only when β is small (i.e., T is large). However, β appears only in the form of βE in F, and thus it does not make sense to discuss the choice of β independently of the design of E. Therefore, the effects of the approximation with small β can only be investigated through experiments.

The overall algorithm is the same as Algorithm 1 except that the variable m_i is updated using Eq.(18) instead of Eq.(10).

3. Advantages of MF methods

This section discusses several real and potential advantages of the MF methods to LBP. LBP iterates message passing between neighboring sites until convergence; at each iteration, the following two steps are performed alternately and independently at each site:

(a) the addition of messages from the neighboring sites and the data term of the site

(b) the computation of messages to be sent to the neighboring sites.

3.1. More flexible choice of MRF models

The MF methods are more flexible than LBP in the choice of MRF models, especially of the representation of the marginal density $p_i(x_i)$. LBP can handle only the Gaussian distribution in the continuous domain, whereas the density functions that the MF methods can deal with are not limited to the Gaussian distribution. The limitation of LBP stems from the fact that in step (b), LBP marginalizes over the variables of the neighboring sites; it is required that the marginalized density should be represented by the same function. Although there are methods based on particle filtering to overcome this limitation [18, 5], there will emerge other issues such as large computational complexity and difficulty with maintaining accuracy. There is no such requirement in the MF methods, and they could deal with all sorts of continuous parametric density functions besides discrete representations. However, it is necessary to derive a different algorithm for each assumed parametric function; moreover, it is another issue whether or not the derived algorithm will be convergent.

3.2. Faster computation

Even in the discrete domain, as far as the computational complexity per iteration is concerned, the MF methods are faster than at least a naive implementation of LBP. In each iteration, the MF methods merely updates the state of the site by referring to the states of its neighboring sites. Its computational complexity is roughly comparable to the computation of a single message in step (b) of LBP. Thus, LBP will be several times (i.e., the number of edges per site) slower than MF. Moreover, LBP needs to access all the neighboring sites to compute a single message, and thus the number of total memory accesses is by the same factor larger than MF. When implementing on parallel systems

such as GPU, the gap could become larger, since overall speed tends to be constrained by the number of memory accesses in these systems.

Of course, smaller computational complexity per iteration does not mean smaller overall computational cost. The other equally important factor is the number of iterations needed until convergence. This basically depends on each problem and datum, and can be investigated only by experiments. According to our experiments, the MF methods are basically comparable to LBP in this repsect.

It should be noted that there are variants of LBP algorithms that perform step (b) efficiently based on distance transform [3, 1]. To be specific, the naive implementation of step (b) has computational complexity of $O(n^2)$ where n is the number of labels, while the efficient algorithms perform this in $O(n)$, although they can be used for a particular class of energy. A similar efficient method is not known for MF. Thus, the efficient LBP algorithms will be faster than our current implementations of the MF methods, especially when n is large. Note, however, that this will not be a problem if n is small.

3.3. Accuracy

As is described above, MF computes the marginal density at each site in an approximate sense, so does LBP. Thus, our concern is with the accuracy of the approximations. As mentioned above, MF is based on the assumption of the independence of each site, i.e., Eq.(6). LBP is based on a supposedly more accurate assumption such that the marginal density of a site is represented by

$$P(\boldsymbol{x}) = \frac{\prod_{(i,j)\in\mathcal{E}} p_{ij}(x_i, x_j)}{\prod_i p_i(x_i)^{q_i-1}} \tag{19}$$

Thus, LBP is considered to be more accurate than the naive MF equations to the extent of the difference between Eqs.(6) and (19).

Although its derivation is considerably different, the method of the TAP equation can be regarded as improving the accuracy of the naive MF equation. Thus, which prevails between the above difference of naive MF from LBP and the improvement by the TAP equation. This generally has to be investigated by experiments. According to our experiments shown in Sec. 4, the 3rd-order TAP method generally yields more accurate results than LBP.

4. Experimental results

We conducted experiments to compare the performances of the MF methods and LBP. As example problems, we consider interactive segmentation, which is a binary label classification problem, and stereo matching, which is a multi-label classification problem. For the former, we tested the naive MF, 2nd-order TAP and 3rd-order TAP methods, and for the latter, we tested the multi-label naive MF method. For LBP, we used the naive implementation of the sum-product algorithm.

In the experiments, Intel Core i7 2.67GHz CPU and nVidia GeForce GTX480 GPU were used. MSVC2010 is used for implementations on the CPU; /fp:fast option is specified to maximize the performance of floating point arithmetic and further the code is parallelized using OpenMP. CUDA is used for implementations on the GPU.

4.1. Interactive segmentation

Following the same procedure as GrabCut [17], we use brushes to roughly specify the foreground and the background pixels of an image, as shown in Fig.1, from which their color models are learned. We used the functions `calcNWeights()` and `constructGCGraph()` from OpenCV2.3 to generate the energy function, where the parameters were set as $\gamma = 50$ and $\lambda = 450$. Defining the variable x_i to indicate whether the pixel i is foreground or background, each method estimates the marginal density of $Q(\boldsymbol{x})$ at each pixel. The parameter T was empirically chosen as $T = 80$. For target images, we used an image from [13] and four images from the dataset of [12].

In the original GrabCut, optimization is iteratively performed a few times, while the Gaussian mixture models (GMMs) of the foreground and background pixels are updated at each iteration. When the same procedure is carried out in our case, **the MF methods and LBP yielded almost identical results; they are too close to find a significant difference in terms of accuracy**. (The results are also almost the same as those of GC, when each pixel is classified as foreground and background by thresholding with $p = 0.5$.) Therefore, we carry out the optimization only once while fixing the energy initially determined by manual brushes. For the purpose of evaluating the accuracy of the estimation of the marginal densities, we also estimate them by Gibbs sampling [9] and use the estimates as the ground truths. In this computation, a sufficient number (= 20000) of samples are generated and used per pixel.

Figure 1 shows the results for the image *Bird* from [13]. The size of the images is 640×480 pixels. The brightness of each pixel represents the probability that the pixel belongs to the foreground; white is 1.0 and black is 0.0. Comparing the results of the three MF methods with the ground truth obtained by Gibbs sampling, it is observed that the errors tend to decrease in the order of MF, 2nd-order TAP, and 3rd-order TAP. This is more clearly seen in Fig.2 which shows how the errors decrease with the number of iterations. This improvement in accuracy is considered to be due to the effect of the higher-order terms of the TAP equations. Moreover, it is observed from Fig.2 that LBP has smaller errors than 2nd-order TAP, but has larger errors than 3rd-order TAP. This can be visually confirmed in Fig.1.

Table 1 shows the computational time of 100 iterations for each method. It is seen from this table that as compared with LBP, the three MF methods are 3-5 times faster on CPU and 6-10 times faster on GPU. This increase in speed is due to the fact that at each iteration, LBP needs to compute messages in eight directions, one of which is computationally comparable to a single iteration of the MF methods.

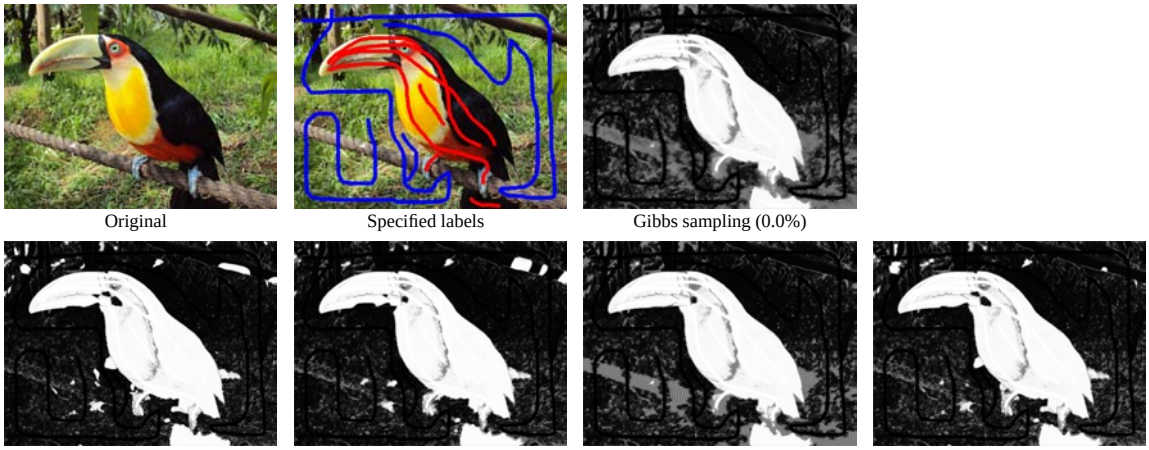

Figure 1. Results of interactive segmentation for an image *Bird*. The numbers in the parenthesis are the residual errors after convergence of each method. The methods are ordered from inaccurate to accurate: MF, 2nd-order TAP, LBP, and 3rd-order TAP.

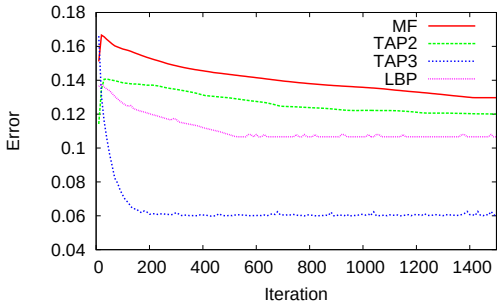

Figure 2. Errors per pixel for interactive segmentation vs. the number of iterations.

Table 1. Computational time for 100 iterations of each method.

	CPU[ms]	GPU[ms]	CPU/GPU
MF	730.0	16.8	43.5
TAP(2nd)	782.12	17.6	44.4
TAP(3rd)	1344.45	26.3	51.1
LBP	4253.01	163.61	26.0

It should also be noted that comparing CPU and GPU implementations, the speed ratios are 40-50 for the MF methods, whereas that for LBP is only 26.0. This is attributable to the fact that LBP requires more memory accesses than the MF methods.

Figures 3 and 4 show the results for *Flower*, *Horse*, *Starfish*, and *Tiger* from the dataset of [12]. Figure 5 presents the residual errors after convergence for each method. It can be seen that the same observation as above holds true for these images; the result is more accurate in the order of MF, 2nd-order TAP, LBP, and 3rd-order TAP.

4.2. Stereo matching

We also performed experiments of stereo matching. We used Middlebury MRF energy minimization library [21] to generate the energy function with the parameters ($|L| = 16$, $\lambda = 20$, and truncated $= 2$). The naive MF algorithm for multi-label classification (Algorithm 2) and LBP were applied to the *Tsukuba* data. As in the above, we used Gibbs sampling with a sufficient number (=100000) of samples to obtain the ground truth of the marginal density of each pixel. We empirically set the temperature as $T = 20$.

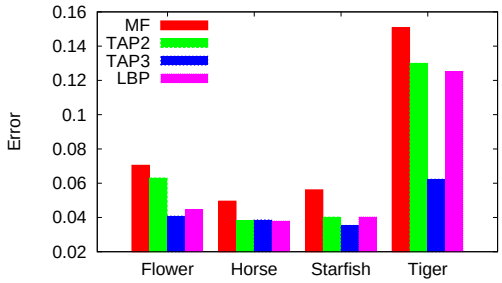

Figure 5. Residual errors per pixel for the four images

Figure 6 shows the results. The residual errors for MF and LBP are 8.9% and 8.0%, respectively. Thus, the error of LBP is smaller reflecting the approximation accuracies of the two methods. However, the estimation results are visually comparable; MF does not appear to be much inferior to LBP. Figure 7 shows the error of the estimate obtained by the MF method versus the number of iterations. It is observed that the error decreases at a reasonable speed as compared with the standard results of LBP reported in the literature (e.g., [3]). Computational time for 100 iterations are 5.0 sec. for MF and 13.5 sec. for LBP. Thus, MF is about three times faster. This matches our earlier analysis; the MRF graph of this problem has four edges at each pixel.

5. Summary and discussions

In this paper, we have investigated whether the MF methods are useful for practical problems of computer vision. To be specific, we applied the naive MF equations and the TAP equations to interactive segmentation and stereo matching.

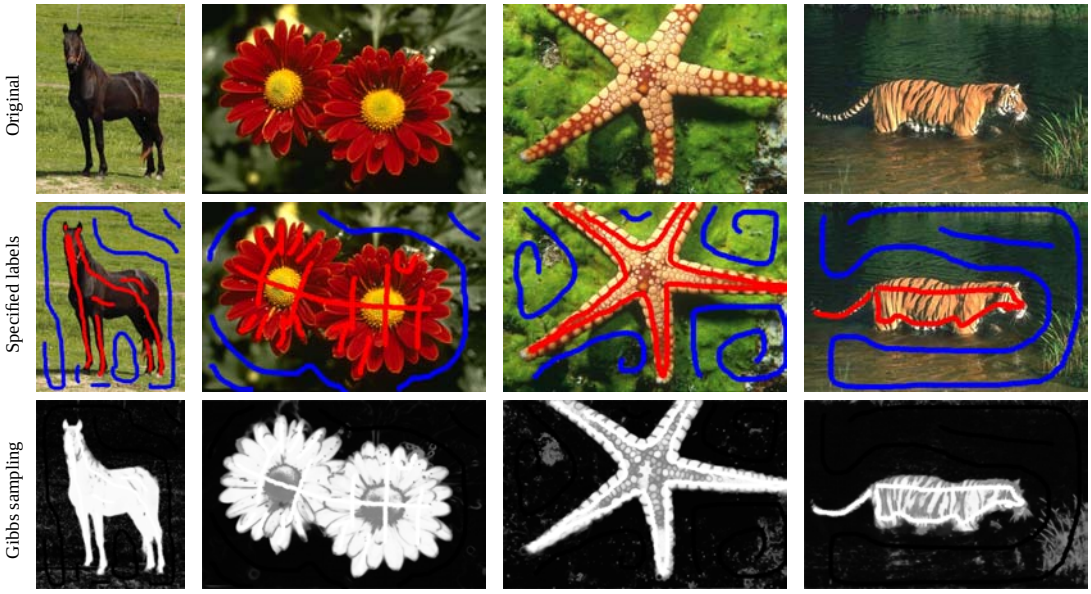

Figure 3. Input images and specified labels for interactive segmentation (*Horse, Flower, Starfish* and *Tiger*).

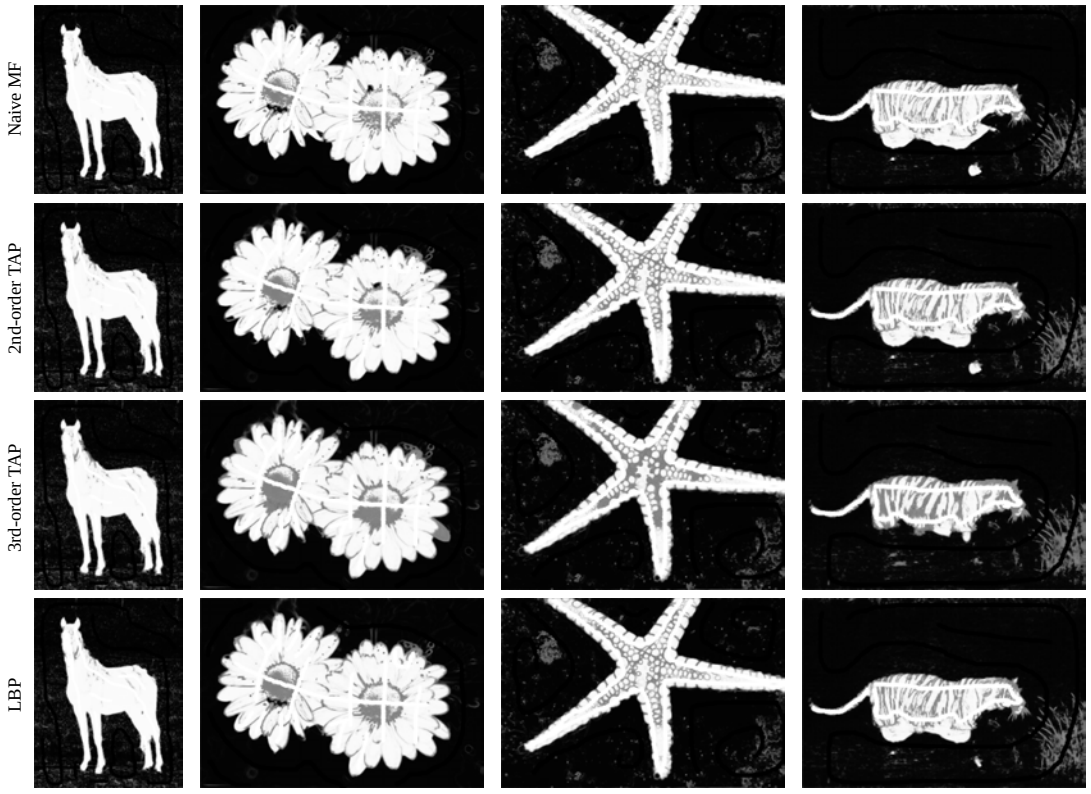

Figure 4. Results for the four images (*Horse, Flower, Starfish* and *Tiger*).

We have presented experimental results to compare the MF methods and LBP. They show that the MF methods show performances at least comparable to LBP. The naive MF equation is less accurate than LBP, which is consistent with theoretical analysis. However, the 3rd-order TAP equation yields more accurate estimates than LBP. As compared with the naive implementation of LBP, the MF methods are faster in terms of a single iteration than LBP. The gap becomes even larger for their implementations on GPU.

We think that being shadowed by LBP, the MF methods have not been correctly evaluated in our community. We suspect that a possible reason to this is in the choice of the

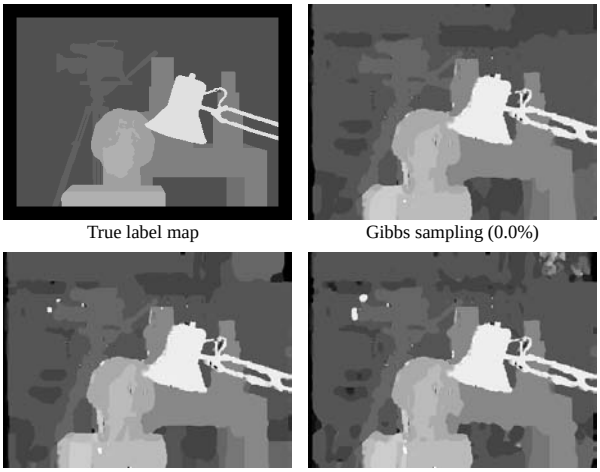

Figure 6. Results of stereo matching. The numbers in the parentheses are residual errors after convergence for each method.

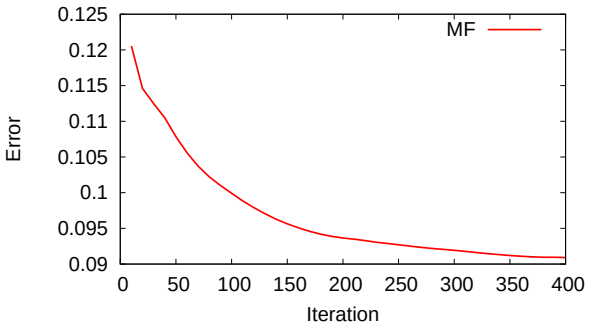

Figure 7. Error of naive MF for stereo matching vs. the number of iterations.

parameter T (or equivalently, $\beta = 1/T$). This parameter can be merely regarded as a scaling parameter of the energy, as in Eq.(2), and therefore it is often simply neglected. This is particularly the case with computer vision problems, in which the energy functions are usually designed by hand for each problem and thus are not based on physics. However, in our experiences, assuming a fixed energy function, T controls the trade-off between global convergence and convergence speed, and its optimal value tends to be different for LBP and the MF methods. To be specific, larger values need to be set to T for the MF methods. Therefore, as far as global convergence is concerned, it can occur that the MF methods do not converge well for the same T (and the energy) as the ones for which LBP works fine. This might have led to a (wrong) observation that the MF methods performs poorly as compared with LBP. To further investigate the true implication of this to the performance of the MF methods, more experiments need to be conducted, which will be a future work.

As has been discussed, the MF methods have several attractive properties such as being able to handle more flexible MRF models. In addition to the one mentioned above, the future work includes the derivation of the TAP equations for multi-label classification and the MF/TAP equations for various continuous parametric density functions other than the Gaussian distribution.

Acknowledgement This work was supported by KAKENHI 22300057.

References

[1] S. Alchatzidis, A. Sotiras, and N. Paragios. Efficient parallel message computation for map inference. In *Proc. ICCV*, 2011.

[2] S. Amari, S. Ikeda, and H. Shimokawa. Information geometry of α-projection in mean field approximation. In *Advanced Mean Field Methods: Theory and Practice*, chapter 16. The MIT Press, 2001.

[3] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70(1):41–54, 2006.

[4] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. PAMI*, (6):721–741, 1984.

[5] M. Isard. Pampas: Real-valued graphical models for computer vision. In *Proc. CVPR*, 2003.

[6] V. Jain and H. S. Seung. Natural image denoising with convolutional networks. In *Proc. NIPS*, pages 769–776, 2008.

[7] H. J. Kappen and W. J. Wiegerinck. Mean field theory for graphical models. In *Advanced Mean Field Methods: Theory and Practice*, chapter 4. The MIT Press, 2001.

[8] P. Kohli and P. H. S. Torr. Measuring uncertainty in graph cut solutions. *Comput. Vis. Image Underst.*, 112(1):30–38, Oct. 2008.

[9] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[10] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. PAMI*, 28(10):1568–1583, 2006.

[11] S. Z. Li. *Markov random field modeling in image analysis*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 3rd. edition, 2009.

[12] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. ICCV*, volume 2, pages 416–423, July 2001.

[13] A. Noma, A. B. V. Graciano, R. M. Cesar-Jr., L. A. Consularo, and I. Bloch. Interactive image segmentation by matching attributed relational graphs. *Pattern Recognition*, 2012.

[14] M. Opper and D. Saad. From naive mean field theory to the tap equations. In *Advanced Mean Field Methods: Theory and Practice*, chapter 2. The MIT Press, 2001.

[15] M. Opper and O. Winther. Tractable approximations for probabilistic models: The adaptive thouless-anderson-palmer mean field approach. *Phys. Rev. Lett.*, 86:3695–3699, Apr 2001.

[16] T. Plefka. Convergence condition of the tap equations for the infinite-ranged ising spin glass model. *Journal of Physics A: Mathematical and General*, page 15, 1982.

[17] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph*, (23):309–314, 2004.

[18] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric belief propagation. In *Proc. CVPR*, 2003.

[19] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo matching using belief propagation. *IEEE Trans. PAMI*, 7(7):787–800, 2003.

[20] J. Sun, N.-N. Zheng, and H.-Y. Shum. Fields of experts. *IJCV*, 2(82):205–229, 2009.

[21] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields. In *Proc. ECCV*, volume 2, pages 19–26, 2006.

[22] T. Tanaka. Information geometry of mean-field approximation. In *Advanced Mean Field Methods: Theory and Practice*, chapter 17. The MIT Press, 2001.

[23] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo using identical MRF parameters. In *Proc. ICCV*, 2003.

[24] 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*, 2003.

[25] Y. Weiss. Comparing the mean field method and belief propagation for approximate inference in mrfs. In *Advanced Mean Field Methods: Theory and Practice*, chapter 15. The MIT Press, 2001.

[26] L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. In *Proc. CVPR*, 2010.

[27] A. Yuille. Generalized deformable models, statistical physics and matching problems. *Neural Computation*, 2:1–24, 1990.