修士学位論文

コンピュータビジョンにおける
マルコフ確率場最適化問題への
平均場近似及び
その拡張手法の応用に関する研究

平成24年度

(平成25年1月31日提出)

東北大学大学院情報科学研究科

システム情報科学専攻

齋藤 真樹

A Study of the Mean Field Approximation and its Extension Methods for Markov Random Field Optimization in Computer Vision

Masaki Saito

Abstract

This thesis considers the problem of optimization using the probabilistic model called Markov Random Fields (MRFs). The MRF model is one of the fundamental probabilistic models in the field of computer vision. It is used for inference of unknown parameters in a wide range of problems, such as image restoration, super-resolution, stereo matching, optical flow estimation, denoising, image segmentation, 3D reconstruction, and object recognition.

We first consider application of the Thouless-Anderson-Palmer (TAP) equation to MRF optimization problems in computer vision. The TAP equation, which has its roots in solid state physics, was proposed to extend the mean field approximation. Although there exist applications of the TAP equation to several problems in the field of solid state physics and information theory, there are practically no application in the field of computer vision. The TAP equation has a form of a Taylor series expansion around zero of the parameter called an inverse temperature. Although this has a certain physical meaning, it is unclear whether this will be a valid assumption for the problems of computer vision. Furthermore, the original TAP equation is developed only for binary-label classification problems and not for multi-label classification problems, which need to be considered in computer vision. In this study, we first investigate whether the TAP equation is useful for practical problems in computer vision. To be specific, we apply the TAP equation for binary classification to the problem of interactive segmentation. We then derive the TAP equation for multi-label classification and apply it to the problem of stereo matching. The experimental results show that the higher-order TAP equation often outperforms Loopy Belief Propagation (LBP), which is recognized as one of the best methods for solving these problems.

We also propose a new method for discretization of continuous MRF models. Existing optimization algorithms for continuous MRF models are only applicable to a limited number of

problems, whereas those for discrete MRF models 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 study, we propose 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 an algorithm using a discrete mean field approximation and one using a belief propagation. These methods 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 can be achieved. We present a method for actually doing this, which dynamically discretizes the variable space in the course of the computation. Experimental results show the effectiveness of our approach.

This thesis is organized as follows. In Chapter 1, we explain the background and motivation of this study. In Chapter 2, we describe the detail of Markov Random Fields and existing optimization methods for them such as a Mean Field Approximation and a Belief Propagation. Chapter 3 describes the detail of the TAP equations and its application to several MRF inference problems in computer vision. Chapter 4 derives the new algorithms based on mean field approximation and belief propagation that can deal with non-uniformly discretized variable space and also presents a method that dynamically discretizes the variable space. Chapter 5 concludes this thesis.

目次

第1章　はじめに

1.1　背景

本研究ではマルコフ確率場 (Markov Random Field, MRF) と呼ばれる確率モデルの最適化問題について考える．マルコフ確率場はコンピュータビジョンの分野において特に重要な位置を占める確率モデルである．1984 年に Geman ら [14] がマルコフ確率場を画像復元の問題に対して応用したことをそのきっかけとして，マルコフ確率場はコンピュータビジョンの多くの問題を解くための確率モデルとして積極的に用いられてきた歴史を持つ．その応用範囲は幅広く，例として画像復元 [14, 48]，超解像 [47, 57]，ステレオマッチング [53, 56]，オプティカルフロー推定 [65, 62]，ノイズ除去 [36, 2]，画像のセグメンテーション [11, 49]，物体の 3 次元復元 [10]，一般物体認識 [33] などが挙げられる．

マルコフ確率場がこのような多くの問題に対して応用される理由は大別して 3 つある．一つは，コンピュータビジョンの問題が持つ統計的性質の多くが，マルコフ確率場の統計的性質と合致している点にある．例えば，ノイズが多少含まれている画像からノイズが含まれない画像を生成する，画像のノイズ除去という推定問題を考える．このようなある入力を元に，望ましい出力を推定する問題を解くための方法としては次の手法を取ることが多い．まず，自然画像全体の背後に隠れている統計的性質 (隣接画素同士は同じ色を取りやすい，隣接エッジ同士の勾配はなめらかに繋がる場合が多い等) を元に，入力画像が与えられた際の，出力画像の条件付き確率を適当な確率モデルを用いてモデリングする．次に，その条件付き確率を元に，尤もらしい出力画像を適当な最適化アルゴリズムを用いて推定する．マルコフ確率場は，この隣り合う画素同士が持つ関連性を記述できる確率モデルである．この性質は画像のノイズ除去だけでなく，多くのビジョンの問題が持つ自然な性質である．例えば，画像を人物と景観のような，前景と背景の 2 つのラベルに分割する二値ラベルセグメンテーションの問題では，隣接画素同士のラベルは同じラベルである確率が高い．また，画像にどのような物体が映っているのかを判定する一般物体認識の問題では，隣接画素同士は同じカテゴリ (車，人間，草木など) に属する可能性が高い．マ

ルコフ確率場はこのような隣り合う画素同士の関連性を自然に記述できるため，汎用性が高い．

もちろん，このような統計的性質をマルコフ確率場で記述することなく，推定結果を直接的に求めるアルゴリズムも複数存在する．前述のノイズ除去の例では，バイラテラルフィルタ [59] などの適当なフィルタを用いて推定結果を直接求めることもできる．マルコフ確率場がコンピュータビジョンの分野で頻繁に用いられる第二の理由としては複雑な相互作用を考慮した推定によって，推定結果を直接的に求めるアルゴリズムと比較して大幅な推定精度の改善が期待できる点にある．ここで，複雑な相互作用の具体的な効果について，前述の二値ラベルセグメンテーションの問題を例に説明する．仮にある画素のラベルが前景である確率が非常に高いものとすると，セグメンテーションがもつ統計的性質から，その周囲の画素はおそらく前景のラベルをとる確率が高い．そして，その周囲の画素が前景である確率が高いのであれば，周囲の画素に隣接する画素が前景のラベルを取る確率は同様に高い値をとる．マルコフ確率場はこのような確率的な伝搬構造を自然に記述できる．すなわち，マルコフ確率場の下では隣接画素同士は互いに影響を受け，その影響は隣の画素に対して間接的に伝搬していく構造をとる．これは，マルコフ確率場がすべての画素がすべての画素に対して間接的に影響を与える，複雑な相互作用を考慮した確率モデルであることを表す．結果的に，マルコフ確率場の下で得られた推定結果は，多くの場合に従来手法を上回る精度を達成できる．

しかしながら，このような複雑な相互作用を考慮したモデルの場合，その推定の厳密解を求めるための計算コストは相互作用を考慮しないモデルと比較して飛躍的に増大してしまう．マルコフ確率場が多くのコンピュータビジョンの問題に応用される第三の理由は，このような複雑なモデルを効率よく推定するための計算アルゴリズムが数多く提案されている点にある．

これはマルコフ確率場が持つ歴史の長さに由来する．マルコフ確率場は元々，物性物理学における多体系のモデルを単純化したイジング模型と呼ばれるモデルがそのルーツであり，1930 年代に Ising によって提案された．しかし，単純化したとはいえイジング模型の式の中には相互作用を表す項–これはマルコフ確率場の相互作用項と等価である–が含まれているため，このモデルのふるまいを解析に求めることは多くの場合不可能であった．そのため，Ising は同時にイジング模型のふるまいを高速に計算するための，平均場近似と呼ばれる近似的な推定手法を提案した．この推定手法は精度を犠牲にして高速なマルコ

フ確率場の推定が行える手法であり，現在ではマルコフ確率場だけでなく，他の確率モデルに対しても変分ベイズ法 [5, 27] などの別名で応用されている．また，コンピュータビジョンの分野では殆ど知られていないが，物性物理学の分野では TAP 方程式と呼ばれる平均場近似を拡張した方程式についても知られている．TAP 方程式についての詳細は §1.2 ならびに §3 で述べる．

一方，情報科学の分野からもマルコフ確率場の推定を高速に行うための手法が複数提案されている．その中で最も有名である手法は，Pearl が 1982 年に提案した確率伝搬法 [45] である．この手法は元々対象のマルコフ確率場の構造が木構造と呼ばれる特殊な構造の場合における系の推定解を厳密に求めるための手法であったが，現在は木構造でない任意のマルコフ確率場の構造においても，系の推定解を近似的に求められることが知られている [63].

オペレーションズ・リサーチの方面からも，マルコフ確率場の問題を解くための多くの研究がさかんに行われている．最も有名な成果はマルコフ確率場の推定問題を等価なネットワーク最大流の問題に変換し，最小切断アルゴリズムを用いることで系の大域最適解を求めるグラフカットの手法である．グラフカットの初出は定かでないが，少なくともオペレーションズ・リサーチの方面では 40 年前から用いられており [17]，コンピュータビジョンの分野では 80 年代後半に，2 値画像のノイズ除去の問題に対してグラフカットの手法が用いられている [4, 16]．このような，計算量が多い問題を効率的に解くための最適化手法の多さは，マルコフ確率場が持つ魅力の一つである．

1.2　本研究の目的

以上の理由から，マルコフ確率場はコンピュータビジョンの多くの問題を解くための確率モデルとして幅広く用いられてきた．そのため，本研究ではコンピュータビジョンの可能性を広げる目的として，マルコフ確率場に対する最適化手法–特に物性物理の分野で用いられる平均場近似–について，主に 2 つの研究を行った．

一つは TAP 方程式と呼ばれる物理方程式の，コンピュータビジョンに対する応用である．TAP 方程式は物性物理の分野にその起源を持ち，平均場近似を拡張する目的で提案された方程式である．物性物理や情報理論の分野ではマルコフ確率場の問題に対する TAP 方程式の応用がいくつかなされているものの，コンピュータビジョンの分野に関してはわ

れわれの知る限りほとんど存在しない．理論上，TAP 方程式の精度は平均場近似を上回る．しかしながら，TAP 方程式の導出は『逆温度』と呼ばれるパラメータが 0 付近であるという仮定 (詳細は §3 で述べる) を元にするため，この仮定が実際のコンピュータビジョンの領域でどのような影響を及ぼすのかについては定かでなく，実験的に確かめるほかない．さらに，元々の物性物理学分野で提案された TAP 方程式は，対象のマルコフ確率場の取りうる状態が 2 通りしかない二値ラベル分類問題に対してのみ導出されており，コンピュータビジョンの問題 (物性物理学の分野では対応する問題が存在しない) で頻繁に用いられる，取りうる状態が複数存在する多値ラベル分類問題に対しては導出されていない．

本研究では二値ラベル分類問題についての TAP 方程式を特定のコンピュータビジョンの問題に対して応用し，その精度を確認した．さらに，多値ラベル分類問題に対しての TAP 方程式を新しく導出し，同様にコンピュータビジョンの問題に対して実験を行った．われわれの実験では，高次の TAP 方程式は精度と計算速度の両面で，コンピュータビジョンの領域でマルコフ確率場の問題を解く目的で主に用いられる確率伝搬法の精度を上回ることが分かった．

もう一つは連続的なマルコフ確率場の離散化問題についての新しい提案である．多くのコンピュータビジョンの問題では，対象のマルコフ確率場の取りうる値は離散的ではなく，連続的な変数として扱ったほうが自然である場合が多い．例えば，与えられた左右 2 枚の画像から各画素毎の画像の深度を推定するステレオマッチングの問題では，各画素の値は深さを表す連続値として扱われる．また，時間的に連続する 2 枚の画像から各画素がどの画素の位置に移動したのかを推定するオプティカルフローの問題では，各画素の値は 2 次元の連続的なベクトルとして扱われる．すなわち，これらの問題をマルコフ確率場のフレームワークを用いて推定する場合，本来の解法としては連続的な値を取るマルコフ確率場の問題を考え，各画素の連続値を推定することが望ましい．

対象の確率モデルが非常に単純なモデルであれば，その連続的な推定値を求められる場合もあるが，多くのコンピュータビジョンの問題で扱うより複雑なモデルの場合，連続的な推定解の導出には事実上無限大の計算コストを要するため不可能である．一方，離散的なマルコフ確率場の推定解を平均場近似や確率伝搬法の手法を用いて近似的に求めようとする場合はそのような問題は起こりえず，現実的な時間内に計算を行える．そのため，コンピュータビジョンの問題の多くはたとえ対象の問題が連続的な場合であっても，連続的なマルコフ確率場の問題を離散的なマルコフ確率場の問題へ離散化し，その上で離散的

なマルコフ確率場の最適化問題を解くことで推定解を求めていた．

本研究ではこの連続的なマルコフ確率場の問題を，離散的なマルコフ確率場の問題に離散化するための方法を新しく提案した．具体的には，提案手法ではマルコフ確率場それ自体は連続的な分布としてとり扱い，離散化は行わない．その代わり，提案手法では後述する周辺分布それ自体を『離散化』することで最適化を行う (詳細は §4.2 で述べる)．このような定式化の下で，われわれは新しい平均場近似と確率伝搬法のアルゴリズムを新しく導出した．これらのアルゴリズムは離散的な平均場近似と確率伝搬法のアルゴリズムと類似しているため，従来手法と同じ計算コストで推定解を求められるが，補正項と呼ばれる項が一部に含まれている点が異なる．この補正項の存在によって，従来では正当に扱うことができなかった非一様な変数空間の離散化に対しても，正しい推定解を求められる．加えて，提案手法では重要でない箇所は疎に離散化し，逆に重要な箇所は密に離散化することで，計算速度を保ちつつ精度の高い周辺分布を求められる．さらに，非一様な離散化を効率的に行うため，本研究では動的離散化という，疎から密に変数空間を離散化するための手法を新しく提案した．最後に複数の実験を通して，本研究の提案手法が従来手法と比較して効率的であることを示した．

1.3　本論文の構成

本論文の構成は以下の通りである．§1 では，研究の背景と目的について述べた．§2 では，前述した 2 種類の研究に共通する，マルコフ確率場の確率モデルならびにその推定手法の代表例である平均場近似と確率伝搬法についてその詳細と導出手法について述べる．§3 では一つの研究である，TAP 方程式を用いた平均場近似の拡張とその応用について述べる．具体的には，TAP 方程式の概念とその導出手法について述べた後に，TAP 方程式の利点についてその詳細を示す．また，実験を通して，TAP 方程式の手法が実用上有効であることを示す．§4 ではもう一つの研究である，非一様に離散化された変数空間上でのマルコフ確率場の推定手法について述べる．具体的には，連続的なマルコフ確率場を離散的な問題へ変換するための従来のアプローチについて説明した後に，新しい平均場近似と確率伝搬法の離散化の方法について述べる．さらに，提案手法である動的離散化の方法について説明した後に，複数の実験を通して，提案手法が効果的であることを示す．§5 は本論文のまとめである．

第2章　マルコフ確率場とその最適化手法

本研究はマルコフ確率場と呼ばれる確率モデルに基づく．また，マルコフ確率場を用いて推定値を求めるための代表的なアルゴリズムとして，平均場近似と確率伝搬法が知られている．本章では，マルコフ確率場の確率モデルについて述べた後に，具体的な推論手法である平均場近似と確率伝搬法について説明する．

2.1　ボルツマン分布

マルコフ確率場 (Markov Random Field, MRF)[30] は確率モデルの一つであり，場のマルコフ性と呼ばれる性質をもつ確率変数の集合である．この性質によって，マルコフ確率場は確率変数をノード，変数間の依存関係をエッジとする無向グラフで表現できる．このような確率変数間の依存関係をグラフで表現した確率モデルを，一般にグラフィカルモデル (graphical model) という [28]．グラフィカルモデルは有向グラフで表現される場合と無向グラフで表現される場合の 2 種類があり，前者はベイジアンネットワーク [29]，後者はマルコフ確率場と呼ばれる．以降は無向グラフのグラフィカルモデル (マルコフ確率場) について説明する．

以下にマルコフ確率場の定義を示す．まず，N 個のサイトから成る無向グラフを $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ と定義する．ここで，$\mathcal{V}$ は $\mathcal{V} = \{1, 2, \cdots, N\}$ で表される頂点集合であり，$\mathcal{E}$ は $(i, j) \in \mathcal{E}$ で表されるエッジ集合である．次に，$\mathcal{G}$ によって結び付けられた確率変数のベクトルを $\mathbf{X} = (X_1, \cdots, X_N)^\top$ と定義する．最後に，$\mathbf{X}$ がある値 $\mathbf{x} = (x_1, \cdots, x_N)^\top$，をとった際の結合確率分布を $Q(\mathbf{x})$ とし，$\mathbf{X}$ が取りうる値の集合を $\mathcal{X}$ と定義する．

このとき，マルコフ確率場とは以下の 2 つの性質をもつ確率モデルを表す [9]．

1. 任意の $\mathbf{x} \in \mathcal{X}$ について，$Q(\mathbf{x}) > 0$ が成り立つ．

2. 任意のサイト i について，次の関係が成り立つ．

$$\Pr(X_i | X_{\mathcal{V} \backslash \mathrm{i}}) = \Pr(X_i | X_{\mathcal{N}_i}) \tag{2.1}$$

ここで，$\mathcal{V}\backslash i$ は $\mathcal{V}$ からサイト i を除いた集合，$\mathcal{N}_i$ はサイト i に隣接するサイトの集合を表す．つまり，場のマルコフ性の下では，サイト i 以外の確率変数を定めた際の X_i の条件付き確率は，サイト i 近傍の確率変数だけを定めた際の，X_i の条件付き確率に等しい．

このような性質を利用することで，マルコフ確率場の振る舞いは以下の簡単な式で表せられる．Hammersley-Clliford 定理によると，マルコフ確率場の結合確率分布 $Q(\mathbf{x})$ は次式で表現できる [3, 13].

$$Q(\mathbf{x}) = \prod_{c \in C_{\mathcal{G}}} \tilde{\phi}_c(\mathbf{x}_c) \tag{2.2}$$

ここで，$\tilde{\phi}_c(\mathbf{x}_c)$ は正の値をとる関数，$C_{\mathcal{G}}$ をグラフ $\mathcal{G}$ 上のクリーク集合，$\mathbf{x}_c$ をクリーク c 上のサイトの値を表すベクトルとする．つまり，マルコフ確率場の下では，結合確率分布 $Q(\mathbf{x})$ はクリーク c 上のサイトの値によって定まる関数の積で表現できる．

ここで，マルコフ確率場の一種であるペアワイズマルコフ確率場 (pairwise Markov Random Field) と呼ばれる単純な確率モデルを考える[1]．(2.2) 式では，すべてのクリーク c に対して関数 $\tilde{\phi}_c$ を定義する形を取っているが，ペアワイズマルコフ確率場では，クリークの大きさが 2 の $\tilde{\phi}_c$ だけが $Q(\mathbf{x})$ を定めているものと定義する．つまり，ペアワイズマルコフ確率場の下では，$Q(\mathbf{x})$ はグラフ上のエッジ (i, j) だけで定められる $\tilde{\phi}_c$ によって決定される，このとき，適当な正の関数 $\psi_{ij}(x_i, x_j)$, $\phi_i(x_i)$ を定めることで，(2.2) 式は次の形で書ける [63].

$$Q(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^{N} \phi_i(x_i) \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) \tag{2.3}$$

ここで Z は正規化定数であり，次の関係を満たす．

$$Z = \int \prod_{i=1}^{N} \phi_i(x_i) \prod_{(i,j) \in \mathcal{E}} \psi_{ij}(x_i, x_j) d\mathbf{x} \tag{2.4}$$

コンピュータビジョンの領域では，(2.3) 式の ϕ_i, ψ_{ij} をそのまま定義することで確率分布を決定する場合もあるが，多くは利便性のためにエネルギー分布と呼ばれる関数を定義することで，間接的に確率分布を決定する．まず，次の関係を用いて関数 $f_i(x_i)$, $f_{ij}(x_i, x_j)$

[1] コンピュータビジョンの領域においては，ペアワイズマルコフ確率場それ自体をマルコフ確率場と呼ぶ場合が多い．一方，ペアワイズでないマルコフ確率場については高次のマルコフ確率場 (higher-order Markov Random Field) と新しく呼ぶことで区別している．

を定義する．

$$f_i(x_i) \equiv -T \ln \phi_i(x_i) \tag{2.5a}$$

$$f_{ij}(x_i, x_j) \equiv -T \ln \psi_{ij}(x_i, x_j) \tag{2.5b}$$

ここで，T は温度 (temperature) と呼ばれる正の定数である．f_i, f_{ij} をそれぞれデータ項 (data term)，平滑化項 (smoothness term) と呼ぶ．(2.5a) 式, (2.5b) 式を用いることで，(2.3) 式は次のように再定義できる．

$$Q(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{1}{T} E(\mathbf{x})\right) \tag{2.6}$$

ここで，$E(\mathbf{x})$ はエネルギー分布 (energy distribution) と呼ばれる関数であり，次の形で定義される．

$$E(\mathbf{x}) = \sum_{i=1}^{N} f_i(x_i) + \sum_{(i,j)\in\mathcal{E}} f_{ij}(x_i, x_j) \tag{2.7}$$

このとき，正規化定数 Z は次の関係をもつ．

$$Z = \int \exp\left(-\frac{1}{T} E(\mathbf{x})\right) d\mathbf{x} \tag{2.8}$$

(2.6) 式を一般にボルツマン分布 (Boltzmann distribution)，もしくはギブス分布 (Gibbs distribution) と呼ぶ．また，(2.8) 式の Z は分配関数 (partition function) と呼ばれることもある．このように，マルコフ確率場の $Q(\mathbf{x})$ は，適当なエネルギー分布 $E(\mathbf{x})$ を定めることでも定義できる．

エネルギー分布によって確率分布を決定することの利点は，コンピュータビジョンの各問題に対するマルコフ確率場のモデリングが行いやすいことにある．コンピュータビジョンの多くの問題は，ϕ_i, ϕ_{ij} を指数分布として設定する場合が多い．その際に (2.5a) 式, (2.5b) 式の定義を用いると，全体のエネルギー分布は L1, L2 ノルムの和のような単純な形へ書き直せる．これは全体のマルコフ確率場の設計が行いやすいことに繋がる．さらに，エネルギー分布を用いた表記によって，後述するマルコフ確率場の MAP 推定問題はエネルギー分布の最小化問題へ帰着される．これによって，最小化問題のための各種アルゴリズムをマルコフ確率場の問題に適用できる利点もある．

2.2　MAP 推定，MPM 推定

前述のとおり，コンピュータビジョンの多くの問題は (2.7) 式 で定義された f_i, f_{ij} を各問題毎に設定することで結合確率分布 $Q(\mathbf{x})$ を作り，その上で各サイトの状態を推定する．サイト毎の状態を推定するための手法としては，MAP 推定, MPM 推定と呼ばれる 2 つの推定手法が知られている．

まず，MAP(Maximum A Posteriori) 推定は，結合確率分布 $Q(\mathbf{x})$ が最大となる状態 $\mathbf{x}^{\mathrm{MAP}}$ をサイトの推定値とする手法である．すなわち，

$$\mathbf{x}^{\mathrm{MAP}} = \arg\max_{\mathbf{x}} Q(\mathbf{x}) = \arg\min_{\mathbf{x}} E(\mathbf{x})$$

である．(2.6) 式の定義から，$Q(\mathbf{x})$ の最大化はエネルギー分布 $E(\mathbf{x})$ の最小化と等価である．

一方，MPM(Maximum Posterior Marginal) 推定はまず $Q(\mathbf{x})$ の周辺化によって x_i についての周辺分布 (marginal distribution) $q_i(x_i)$ を求める．$q_i(x_i)$ はビリーフ (belief) とも呼ばれる．次に，その周辺分布を最大化する値 x_i^{MPM} を推定値とする．すなわち

$$x_i^{\mathrm{MPM}} = \arg\max_{x_i} q_i(x_i)$$

を推定値とする．ここで,

$$q_i(x_i) = \int dx_1 \cdots \int dx_{i-1} \int dx_{i+1} \cdots \int dx_N Q(\mathbf{x}) \tag{2.9}$$

とした．以降はこちらの MPM 推定について説明する．

2.2.1　MPM 推定の利点

一般に，MPM 推定と MAP 推定で得られる推定解の精度は大差ないことが知られている．MAP 推定に関しては，マルコフ確率場の問題のいくつかは MAP 推定専用のアルゴリズム，例えばグラフカット [57, 6]，双対分解 [32, 31] などを用いて効率的に解けることが知られている．そのような問題に対して，MAP 推定のアルゴリズムで解ける問題を MPM 推定のアルゴリズムで行う理由はあまりない．

その一方で，MPM 推定には MAP 推定より優れる場合が大別して 3 つある．

1. **周辺分布それ自体が必要である場合**: 多くのコンピュータビジョンの問題は入力画像が与えられた際に適切な形の解を返す形をとる．この場合に対する問題設定の一つ

として，入力画像が与えられた際の解の条件付き確率を考え，その確率がマルコフ確率場の構造を持つ場合が挙げられる．このような，条件付き確率がマルコフ確率場の構造を持つ確率モデルを一般に条件付き確率場 (Conditional Random Field, CRF) という．ここで，CRF モデルの構造を定めるパラメータを学習する場合を考える．このとき，パラメータの更新の際に条件付き確率場の周辺分布が必要となる [25] が，MPM 推定のアルゴリズムを使うことで，このような問題に対しても正しく CRF モデルの学習パラメータを更新できる利点がある．

2. **MPM 推定を用いたほうが効率的に推定値を求められる場合**: 問題によっては，MPM 推定のアルゴリズムを用いたほうが MAP 推定よりも高速に問題を解ける場合がある．例えば，Krahenbuhl らが提案した一般物体認識の手法 [33] は，すべてのサイトがすべてのサイトに結合した fully-connected MRF と呼ばれるグラフィカルモデルを採用している．このようなモデルで MAP 推定を行うことは多くの計算時間を要するため現実的でないが，MPM 推定では後述する平均場近似のアルゴリズムを用いることで，このような問題設定でも効率的に推定を行える．

3. **MPM 推定のアルゴリズムを MAP 推定に応用する場合**: MPM 推定は MAP 推定を含む．具体的には，MAP 推定解が一意である場合，$T \to 0$ の極限での MPM 推定解は MAP 推定解と等しい．これは，MPM 推定のアルゴリズムを MAP 推定に関しても応用できることを意味する．なお，この利点に関しての詳しい議論は §2.3.3 で行う．

2.3 自由エネルギー最小化による近似

2.3.1 周辺分布を推定するための 2 種類の推定手法

$Q(\mathbf{x})$ の MPM 推定を行うためには周辺分布 $q_i(x_i)$ を求める必要があるが，(2.9) 式を使って厳密解を直接計算することは困難である．なぜなら，仮にすべての x_i が 2 つの状態しか取らない場合でも，(2.9) 式に従って周辺分布 $q_i(x_i)$ を計算するためには 2^N 通りの状態を考慮する必要があるため，計算量が爆発するからである．

一般に，周辺分布を計算するためのアルゴリズムとして 2 種類の方法が知られている．一つはマルコフ連鎖モンテカルロ (Markov Chain Monte Carlo, MCMC) 法によって $Q(\mathbf{x})$

に従ったサンプルを大量に生成することで，周辺分布を近似的に求める方法 [41] である．MCMC は主にメトロポリス・ヘイスティングス (Metropolis-Hastings) 法 [38] やギブスサンプリング (Gibbs sampling) 法 [12] に代表される，数多くの手法がこれまでに提案されている [28] が，マルコフ確率場の最適化では主にギブスサンプリング法が用いられる．MCMC で推定した周辺分布は，サンプル数を十分に多く取ることで本来の周辺分布に漸近的に一致することが知られているが，十分な数のサンプルを得るためには多くの計算時間を必要とする．

もう一つは，後述する自由エネルギー最小の変分原理に基づいて，周辺分布の近似解を高速に計算する方法である．この方法で得られる解は一般に真の周辺分布とは一致しない近似的なものだが，MCMC 法と比較して高速に Q の近似周辺分布を求められる．周辺分布を計算するための近似手法としては主に平均場近似 [24] と確率伝搬法 [63] が知られており，後述する自由エネルギーと呼ばれる汎関数を最小化することで近似的な周辺分布の推定を行う．以降は後者の手法について説明する．

2.3.2 自由エネルギーの導入

前述のとおり，Q の周辺分布を直接求めることはほとんどの場合不可能である．そのため，平均場近似と確率伝搬法は，まず Q を近似する新しい分布 P を導入する．そして，P には周辺分布を高速に計算できる性質をもつ分布を選ぶ．ここで導入する仮定は平均場近似と確率伝搬法でそれぞれ異なるため，導出されるアルゴリズムもそれぞれ異なる．次に，P を Q に関して近づけることで近似分布を得る．この操作は後述する自由エネルギー $F[P]$ の最小化に対応する．最後に，P の周辺分布を求め，これを Q の近似周辺分布とする．

以下に具体的な導出を示す．まず，P, Q 間の近さは，次の Kullback-Leibler(KL) と呼ばれる情報量 [5] を用いて図る．

$$\mathcal{D}[P\|Q] = \int P(\mathbf{x}) \ln \frac{P(\mathbf{x})}{Q(\mathbf{x})} d\mathbf{x} \tag{2.10}$$

KL 情報量は $P = Q$ のときに最小値 0 をとる非負の関数であるため，KL 距離とも呼ばれる．ただし，反対称性 $\mathcal{D}[P\|Q] \neq \mathcal{D}[Q|P]$ から，KL 距離は厳密な意味での距離ではない．このような，関数を引数にとり，その関数全体の形によって値が定まる関数のことを，一般に汎関数 (functional) と呼ぶ [5].

次に，P, Q 間の KL 距離を P に関して最小化させることで，Q の近似分布 P^* を得る．

$$P^*(\mathbf{x}) = \arg\min_P \mathcal{D}[P\|Q] \tag{2.11}$$

このような，汎関数の最大値，最小値，停留値を求める手法全般のことを，一般に変分法 (variational method) と呼ぶ [5]．なお，$\mathcal{D}[P\|Q]$ における P の最適化は，一般に多くの局所解をもつ [24]．これは，平均場近似，確率伝搬法の双方とも，限られた条件の元でしか P の推定解を一意に求められないことを意味する．

(2.6) 式を (2.10) 式に代入することで次式を得る．

$$\begin{aligned}\mathcal{D}[P\|Q] &= \frac{1}{T}\int P(\mathbf{x})E(\mathbf{x})d\mathbf{x} - \left(-\int P(\mathbf{x})\ln P(\mathbf{x})d\mathbf{x}\right) + \ln Z \\ &= \frac{1}{T}\langle E\rangle_P - S[P] + \ln Z \end{aligned} \tag{2.12}$$

ここで，$\langle E\rangle_P$ はエネルギー分布 $E(\mathbf{x})$ の P に対する期待値，$S[P]$ は分布 P のエントロピーを表す．$\langle E\rangle_P$ は分布 P についての平均エネルギー (mean energy) とも呼ばれる．このとき，$\ln Z$ は分布 P によって変化しないため，(2.11) 式は次式の関係と等価である．

$$P^* = \arg\min_P F[P] \tag{2.13}$$

ここで，$F[P]$ は自由エネルギー (free energy) と呼ばれる汎関数であり，次式の関係で定義される[2]．

$$F[P] = \beta\langle E\rangle_P - S[P] \tag{2.14}$$

ここで，$\beta = 1/T$ である．平均場近似，確率伝搬法の双方とも，付与された制約条件の元で $F[P]$ を最小化することで，Q の周辺分布を近似的に求めている．

ここで，KL 距離が最小となる P は $P = Q$ のときであり，そのときの自由エネルギー $F[P_{\min}]$ は次式の関係を満たす．

$$F[P_{\min}] = -\ln Z \tag{2.15}$$

この関係は後述する TAP 方程式の導出の際に用いる．

[2]統計物理の分野では $F[P]$ に T を乗じた $F'[P] = \langle E\rangle_P - TS[P]$ を自由エネルギーと定義しているが，どちらの汎関数を利用して最適化したとしても分布 P の形に違いはない．

2.3.3　MPM 推定解と MAP 推定解の一致

前述の通り，MAP 推定解が一意である場合において，MPM 推定解は $T \to 0$ の極限で MAP 推定解に一致する．これは (2.14) 式によって説明できる．$T \to 0$ の極限では $\beta \to \infty$ となるため，自由エネルギーは第二項である $S[P]$ の影響を受けない．つまり $F[P] \to \langle E \rangle_P$ となる．このような $F[P]$ を最小にする分布 P^* は，エネルギー分布が最小となる状態だけをとる分布である．つまり，

$$P^*(\mathbf{x}) \to \prod_{i=1}^{N} \delta(x_i - x_i^{\mathrm{MAP}}) \tag{2.16}$$

となる．ここで $\delta(x)$ は Dirac のデルタ関数，x_i^{MAP} は MAP 推定解 $\mathbf{x}^{\mathrm{MAP}}$ の第 i 成分を表す．(2.16) 式の MPM 推定解は明らかに MAP 推定解に等しい．以上から，$T \to 0$ の極限で MPM 推定と MAP 推定は一致する．

2.4　平均場近似

本節では，最も単純な近似手法の一つである平均場近似 (mean field approximation) について説明する．平均場近似は，元々物性物理学にそのルーツを持つ [8]．物性物理学では，強磁性体の相転移現象をモデル化する用途としてイジング模型 (Ising model) と呼ばれる単純な多体系のモデルを使用するが，相互作用項と呼ばれるサイト間の相互作用を記述する項–これはマルコフ確率場の領域では平滑化項と等価である–の存在によって，磁性体のふるまいを解析的な計算で求めることは非常に困難であった．元々の平均場近似はこの相互作用項を期待値の平均に置き換えることで計算を単純化し，磁性体のふるまいを近似的に求める手法である．一方，情報科学の分野では，平均場近似の手法はしばしば自由エネルギー最小の変分原理に基づいた，近似手法の一種とみなされる．このアプローチを用いる利点としては，物理的な概念 (例; 熱力学的極限) を使用せずに，確率論の範囲だけで平均場近似の手法を説明できることが挙げられる．ゆえに，本節では後者の手法を用いて，平均場近似の方程式を導出する．

まず，平均場近似では，分布 P に対し各サイト i がそれぞれ独立であるという制約条件の元で，P を Q に近づける．つまり，

$$P_{\mathrm{MF}}(\mathbf{x}) = \prod_{i=1}^{N} p_i(x_i) \tag{2.17}$$

という確率モデルの下で，$F[P]$ を最小化する．この仮定は一般のマルコフ確率場では正しくない確率モデルであるものの，高速に近似分布を計算できる利点を持つ．(2.17) 式から，P_{MF} のサイト i における周辺分布は p_i と等価であることが分かる．

(2.17) 式を $F[P]$ に代入することで次式を得る．

$$F[P_{\mathrm{MF}}] = \beta \sum_{i=1}^{N} \int p_i(x_i) f_i(x_i) dx_i + \beta \sum_{(i,j)\in\mathcal{E}} \iint p_i(x_i) p_j(x_j) f_{ij}(x_i, x_j) dx_i dx_j + \sum_{i=1}^{N} \int p_i(x_i) \ln p_i(x_i) dx_i \quad (2.18)$$

(2.18) 式の第一項，第二項が $F[P]$ の第一項である $\langle E \rangle_P$(平均エネルギー)，(2.18) 式の第三項が $F[P]$ の第二項である $S[P]$ (分布 P のエントロピー) にそれぞれ対応している．

次に，導出された (2.18) 式を用いて，$F[P_{\mathrm{MF}}]$ を P_{MF} について最小化する．ただし，(2.17) 式から p_i は確率密度関数でなければならないため，$\int p_i(x_i) = 1$ という制約条件の元で (2.18) 式の最小化を行う．まず，Lagrange 乗数 γ_i を付与した次の Lagrange 関数 J_{MF} を導入する [24].

$$J_{\mathrm{MF}} \equiv F[P_{\mathrm{MF}}] + \sum_{i=1}^{N} \gamma_i \left(1 - \int p_i(x_i) dx_i \right) \quad (2.19)$$

J_{MF} の導入により，制約条件下での $F[P_{\mathrm{MF}}]$ の最適化問題は，制約条件が存在しない J_{MF} の最適化問題へと変形できる．次に，汎関数 J_{MF} の停留点を Euler-Lagrange 方程式を利用して求める [5]．ただし，今回の場合汎関数は $\partial P_{\mathrm{MF}}/\partial \mathrm{x}$ を引数に持っていないため，Euler-Lagrange 方程式による停留点は通常の微分のように，J_{MF} を $p_i(x_i)$ に関して微分することで求められる．ゆえに，J_{MF} を $p_i(x_i)$ に関して微分し，$\int p_i(x_i) = 1$ の関係から γ_i を消去することで次の固定点方程式を得る．

$$p_i(x_i) \propto \exp \left[-\beta \left(f_i(x_i) + \sum_{j\in\mathcal{N}_i} \int p_j(x_j) f_{ij}(x_i, x_j) dx_j \right) \right] \quad (2.20)$$

ここで，$\mathcal{N}_i$ はサイト i の隣接サイトの集合を表す．以上から，$F[P]$ の局所解は，(2.20) 式を満たす $p_i(x_i)$ を求めることで求められる．

2.4.1 反復法による近似周辺分布の推定

(2.20) 式より，$F[P]$ の停留点は，(2.20) 式の関係を満たす非線形の固定点方程式として，$P_{\mathrm{MF}} = G[P_{\mathrm{MF}}]$ のような形で書ける．平均場近似では，このような固定点方程式の解を求

めるための手法として一般に反復法を用いる．

具体的な方法を以下に示す．はじめに，周辺分布を適当な P_{MF}^0 によって初期化する．次に，p_i を $P_{\mathrm{MF}}^{t+1} = G[P_{\mathrm{MF}}^t]$ に従って逐次更新していく．反復法では，このような手続きを T 回繰り返す．平均場近似では，このような手順で得られる P_{MF}^T は発散，振動せずに必ず自由エネルギー $F[P]$ の停留点へ収束することが知られている [24]．具体的な平均場近似のアルゴリズムを Alg.1 に示す．ここで，Z_i を正規化定数とした．

Algorithm 1 連続分布における平均場近似

1: **for all** i **do**
2: $\quad p_i^0(x_i) \leftarrow \exp\left[-\beta f_i(x_i)\right]/Z_i$
3: **end for**
4: **for** $t = 0$ to $T-1$ **do**
5: $\quad$ **for all** i **do**
6: $\quad\quad p_i^{t+1}(x_i) \leftarrow \exp\left[-\beta\left(f_i(x_i) + \sum_{j\in\mathcal{N}_i}\int p_j^t(x_j) f_{ij}(x_i, x_j)dx_j\right)\right]/Z_i$
7: $\quad$ **end for**
8: **end for**

2.4.2 二値ラベル分類問題における平均場近似の導出

前節ではすべての x_i が連続的な値をとるものと仮定していたが，コンピュータビジョンの問題の中には，すべての x_i が二つの状態だけを取る問題もある (例えば，対象の画像の各画素を前景と背景の 2 つのラベルに分割するセグメンテーションの問題)．本節では，対象の問題が二値ラベル分類問題である場合における，平均場近似の固定点方程式を導出する．

すべての x_i が $\{-1, +1\}$ の値を取るものとすると，二値ラベル分類問題におけるエネルギー分布の形は次のように書ける．

$$E(\mathbf{x}) = -\sum_i h_i x_i - \sum_{i,j} J_{ij} x_i x_j \tag{2.21}$$

ここで，$J_{ij}, h_i \in \mathbb{R}$ は定数であり，それぞれ平滑化項，データ項のエネルギー分布を定めるパラメータを表す (例えば，J_{ij} が正であれば，隣接サイト x_i, x_j は同じ状態を取りやすくなり，一方 J_{ij} が負であれば，x_i, x_j は異なる状態を取りやすくなる)．

二値ラベル分類問題も前節と同様の手順を用いることで，固定点方程式を導出できる．ここで，サイト x_i の期待値を $m_i \equiv p_i(+1) - p_i(-1), -1 \leq m_i \leq 1$ と定義すると，周辺分布 p_i は m_i によっても表現できる．すなわち

$$p_i(x_i = +1) = \frac{1 + m_i}{2} \tag{2.22a}$$

$$p_i(x_i = -1) = \frac{1 - m_i}{2} \tag{2.22b}$$

である．これらの関係を $F[P]$ に代入し，整理することで次の自由エネルギーを得る．

$$F[\mathbf{m}] = \sum_i \left[\frac{1 + m_i}{2} \ln \left(\frac{1 + m_i}{2} \right) + \frac{1 - m_i}{2} \ln \left(\frac{1 - m_i}{2} \right) \right] - \sum_i h_i m_i - \sum_{i,j} J_{ij} m_i m_j \tag{2.23}$$

(2.23) 式を m_i について微分し，m_i についての停留点を求めることで次の固定点方程式を得る．

$$m_i = \tanh \left[\beta \left(h_i + \sum_{j \in \mathcal{N}_i} J_{ij} m_j \right) \right] \tag{2.24}$$

なお，実際の周辺分布の推定は Alg.1 と同様である．

2.4.3　多値ラベル分類問題における平均場近似の導出

二値ラベル分類問題と同様に，すべての x_i が $\{1, \cdots, S\}$ の S 個のラベルを取る多値ラベル問題に対しても，平均場近似のアルゴリズムを導出できる．

まず，データ項，平滑化項のエネルギー分布と周辺分布を，次の $f_i^s, f_{ij}^{st}, p_i^s$ を用いて表す．

$$f_i^s \equiv f_i(x_i = s) \tag{2.25a}$$

$$f_{ij}^{st} \equiv f_{ij}(x_i = s, x_j = t) \tag{2.25b}$$

$$p_i^s \equiv p_i(x_i = s) \tag{2.25c}$$

これらの関係を $F[P]$ に代入し，整理することで次の自由エネルギーを得る．

$$F[\mathbf{p}] = \beta \sum_{i=1}^{N} \sum_{s=1}^{S} p_i^s f_i^s + \beta \sum_{(i,j) \in \mathcal{E}} \sum_{s,t} p_i^s p_j^t f_{ij}^{st} + \sum_{i=1}^{N} \sum_{s=1}^{S} p_i^s \ln p_i^s \tag{2.26}$$

§2.4 と同様の手順で $F[\mathbf{p}]$ を $\sum_s p_i^s = 1$ の制約条件の下で最小化することで，次の固定点方程式を得る．

$$p_i^s \propto \exp\left[-\beta\left(f_i^\alpha + \sum_{j\in\mathcal{N}_i}\sum_{t=1}^{S} p_j^t f_{ij}^{st}\right)\right] \tag{2.27}$$

なお，実際の周辺分布の推定は Alg.1 と同様である．

2.5 確率伝搬法

本節では，もう一つの計算手法である確率伝搬法について述べる．確率伝搬法は，情報科学と物性物理学の両方にそのルーツを持つ．情報科学の分野では，グラフィカルモデルが木構造である場合の周辺分布を効率的に求める目的で，Pearl が 1982 年に提案した確率伝搬法がそのルーツ [45] であるが，一方で物性物理学の分野では，イジング模型の磁性体の振る舞いを調べる目的で提案された，Bethe 近似と呼ばれる近似手法 [63] がルーツである．近年になって，この両者の手法が等価な手法であり，後述する近似分布 P の自由エネルギーの停留点を求めるアルゴリズムであることが知られている [63]．ゆえに，今回は自由エネルギー最小の変分原理に従って，確率伝搬法のアルゴリズムを導出する．

なお，確率伝搬法には 2 種類のアルゴリズムが存在し，それぞれ sum-product, max-product(もしくは min-sum) アルゴリズムと呼ばれている [23, 26]．sum-product は MPM 推定を求めるためのアルゴリズムであり，自由エネルギー最小の変分原理によって求められる．一方，max-product は MAP 推定を求めるためのアルゴリズムであり，sum-product の $T \to 0$ の極限下で導出できる [28]．今回は sum-product アルゴリズムによる確率伝搬法について説明する．

前述のとおり，確率伝搬法に関しても平均場近似と同様に，近似分布 P がある確率分布のクラスを取るという仮定の下で自由エネルギーを最小化する．ただし，その近似分布の形は平均場近似のそれとは異なる．具体的には，確率伝搬法は P を次の確率分布のモデルとする．

$$P_{\mathrm{BP}}(\mathbf{x}) = \frac{\prod_{ij} p_{ij}(x_i, x_j)}{\prod_i p_i(x_i)^{z_i - 1}} \tag{2.28}$$

ここで，z_i はサイト i における隣接サイトの数を表す (すなわち $z_i = |\mathcal{N}_i|$)．また，(2.28)

式の $p_i(x_i), p_{ij}(x_i, x_j)$ は，それぞれ次の 3 つの関係を満たすものとする．

$$\int p_i(x_i)dx_i = 1 \tag{2.29a}$$

$$\iint p_{ij}(x_i, x_j)dx_i dx_j = 1 \tag{2.29b}$$

$$\int p_{ij}(x_i, x_j)dx_j = p_i(x_i) \tag{2.29c}$$

このような分布 P_{BP} の下で $F[P]$ を最小化する．平均場近似と同様に，(2.14) 式に (2.28) 式を代入し，整理することによって次式を得る．

$$\begin{aligned} F[P_{\mathrm{BP}}] = \beta \sum_i \int p_i(x_i) f_i(x_i) dx_i + \beta \sum_{(i,j)\in\mathcal{E}} \iint p_{ij}(x_i, x_j) f_{ij}(x_i, x_j) dx_i dx_j \\ - \sum_i (z_i - 1) \int p_i(x_i) \ln p_i(x_i) dx_i + \sum_{(i,j)\in\mathcal{E}} \iint p_{ij}(x_i, x_j) \ln p_{ij}(x_i, x_j) dx_i dx_j \end{aligned} \tag{2.30}$$

(2.30) 式の第一項，第二項が $F[P]$ の第一項である $\beta \langle E \rangle_P$(平均エネルギー)，第三項，第四項が $F[P]$ の第二項である $S[P]$(分布 P のエントロピー) にそれぞれ対応する．(2.30) 式の $F[P_{\mathrm{BP}}]$ は一般にベーテ自由エネルギー (Bethe free energy) と呼ばれる．

ここで，(2.30) 式を簡略化する目的で，局所エネルギー (local energy) と呼ばれる次の分布 $\hat{f}_{ij}, \hat{f}_i$ を導入する．

$$\hat{f}_{ij}(x_i, x_j) = f_{ij}(x_i, x_j) + f_i(x_i) + f_j(x_j) \tag{2.31a}$$

$$\hat{f}_i(x_i) = f_i(x_i) \tag{2.31b}$$

(2.31a) 式, (2.31b) 式を用いることで，(2.30) 式は次式で書き直せる．

$$\begin{aligned} F[P_{\mathrm{BP}}] = - \sum_i (z_i - 1) \int p_i(x_i) \left(\beta \hat{f}_i(x_i) + \ln p_i(x_i) \right) dx_i \\ + \sum_{i,j} \iint p_{ij}(x_i, x_j) \left(\beta \hat{f}_{ij}(x_i, x_j) + \ln p_{ij}(x_i, x_j) \right) dx_i \end{aligned} \tag{2.32}$$

次に，導出された (2.32) 式を用いて，$F[P_{\mathrm{BP}}]$ を P_{BP} について最小化する．ただし，制約条件が 1 種類しかない平均場近似とは異なり，今回の制約条件は (2.29a) 式, (2.29b) 式, (2.29c) 式と 3 種類ある．ゆえに，3 種類の Lagrange 乗数 γ_i, γ_{ij}, λ_{ij} を付与した次の Lagrange 関

数 J_{BP} を導入する．

$$
\begin{aligned}
J_{\mathrm{BP}} = F[P_{\mathrm{BP}}] + \sum_i \gamma_i \left(1 - \int p_i(x_i) dx_i\right) + \sum_{(i,j)\in\mathcal{E}} \gamma_{ij} \left(1 - \iint p_{ij}(x_i, x_j) dx_i dx_j\right) \\
+ \sum_{(i,j)\in\mathcal{E}} \int \lambda_{ij}(x_j) \left(p_j(x_j) - \int p_{ij}(x_i, x_j) dx_i\right) \\
+ \sum_{(i,j)\in\mathcal{E}} \int \lambda_{ji}(x_i) \left(p_i(x_i) - \int p_{ij}(x_i, x_j) dx_j\right)
\end{aligned}
\tag{2.33}
$$

J_{BP} の導入によって，制約条件下での $F[P_{\mathrm{BP}}]$ の最適化問題は，制約条件が存在しない J_{BP} の最適化問題へ変形できる．

次に，Euler-Lagrange 方程式を適用し，$p_{ij}(x_i, x_j), p_i(x_i)$ についての停留点を求める．J_{BP} を $p_{ij}(x_i, x_j), p_i(x_i)$ に関してそれぞれ微分することで，次の関係を得る．

$$
\ln p_{ij}(x_i, x_j) = -\beta \hat{f}_{ij}(x_i, x_j) + \lambda_{ij}(x_j) + \lambda_{ji}(x_i) + \gamma_{ij} - 1 \tag{2.34a}
$$

$$
(z_i - 1)(\ln p_i(x_i) + 1) = -\beta(z_i - 1)\hat{f}_i(x_i) + \sum_{i\in\mathcal{N}_i} \lambda_{ji}(x_i) + \gamma_i \tag{2.34b}
$$

ここで，次式で定義されるメッセージ (message) と呼ばれる関数 m_{kj} を用いて，Lagrange 乗数 λ_{ji} を再定義する．

$$
\lambda_{ij}(x_j) = \ln \prod_{k\in\mathcal{N}_j\setminus i} m_{kj}(x_j) \tag{2.35}
$$

(2.35) 式を (2.34a) 式, (2.34b) 式 に代入し，(2.29a) 式, (2.29b) 式の関係を用いて γ_{ij}, γ_i を消去することで，次式を得る．

$$
p_{ij}(x_i, x_j) \propto \psi_{ij}(x_i, x_j)\phi_i(x_i)\phi_j(x_j) \prod_{k\in\mathcal{N}_i\setminus j} m_{ki}(x_i) \prod_{l\in\mathcal{N}_j\setminus i} m_{lj}(x_j) \tag{2.36a}
$$

$$
p_i(x_i) \propto \phi_i(x_i) \prod_{k\in\mathcal{N}_i} m_{ki}(x_i) \tag{2.36b}
$$

ここで，$\phi_i(x_i), \psi_{ij}(x_i, x_j)$ はそれぞれ §2.1 で説明した (2.5a) 式, (2.5b) 式である．

最後に，(2.36a) 式, (2.36b) 式より次の関係を得る．

$$
m_{ij}(x_j) \propto \int \phi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{k\in\mathcal{N}_i\setminus j} m_{ki}(x_i) dx_i \tag{2.37}
$$

確率伝搬法では，はじめにすべてのメッセージ m_{ij} を適当な関数で初期化する．ほとんどの場合，初期値は 1(一様分布) である．次に，すべてのメッセージを (2.37) 式に従って

Algorithm 2 連続分布における確率伝搬法

for all m_{ij} **do**
  $m_{ij}^0(x_j) \leftarrow 1$
end for
for $t = 0$ to $T-1$ **do**
  for all m_{ij}^t **do**
    $m_{ij}^{t+1}(x_j) \leftarrow \int \phi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}^t(x_i) dx_i$
  end for
end for
for all i **do**
  $p_i(x_i) \leftarrow \left(\phi_i(x_i) \prod_{k \in \mathcal{N}_i} m_{ki}^T(x_i)\right) / Z_i$
end for

更新する．この手続きを十分な回数 (T 回) 繰り返す．最後に，周辺分布を (2.36b) 式によって推定する．具体的なアルゴリズムを Alg.2 に示す．ここで，Z_i を正規化定数とした．

確率伝搬法の重要な性質として， 対象のグラフィカルモデルの構造が木構造である (対象のグラフにループが存在しない) 場合，Alg.2 によって得られる周辺分布は厳密解に一致する性質が挙げられる．これは，反復計算によって得られる近似分布 P が，常に元の分布 Q に一致することを表している．

一方，対象のグラフィカルモデルの構造がループを含む場合，Alg.2 で求められる近似分布 P は一般に Q に等しくない．さらには，グラフィカルモデルの構造によっては (2.37) 式によるメッセージの更新が収束せずに，振動してしまう場合もある．ただし，たとえループを含んでいる場合でも，Alg.2 で得られる周辺分布の推定解は，多くの場合に Q の周辺分布と近いことが経験的に知られている．このようなループを含む確率伝搬法の手法は，一般に Loopy Belief Propagation(LBP)[40] と呼ばれている．

2.5.1 多値ラベル分類問題における確率伝搬法の導出

多値ラベル分類問題に関しての確率伝搬法も，連続分布と同様の手順で導出できる．なお，平均場近似とは異なり，二値ラベル分類問題では反復方程式を (2.24) 式のような簡略化された形で書くことができない．そのため，今回は多値ラベル分類問題における確率伝

搬法だけを導出する．二値ラベル分類問題は多値ラベル分類問題の特殊な場合と考えればよい．

$\psi_{ij}(x_i, x_j), \phi_i(x_i), m_{ij}(x_j), p_i(x_i), p_{ij}(x_i, x_j)$ を，それぞれ次の $\psi_{ij}^{st}, \phi_i^s, m_{ij}^t, p_i^s, p_{ij}^{st}$ を用いて表す．

$$\psi_{ij}^{st} \equiv \exp\left[-\beta f_{ij}(x_i = s, x_j = t)\right] \tag{2.38a}$$

$$\phi_i^s \equiv \exp\left[-\beta f_i(x_i = s)\right] \tag{2.38b}$$

$$m_{ij}^t \equiv m_{ij}(x_j = t) \tag{2.38c}$$

$$p_i^s \equiv p_i(x_i = s) \tag{2.38d}$$

$$p_{ij}^{st} \equiv p_{ij}(x_i = s, x_j = t) \tag{2.38e}$$

§2.5 と同様の手順で，$F[P]$ を $\sum_s p_i^s = 1, \sum_{s,t} p_{ij}^{st} = 1, \sum_t p_{ij}^{st} = p_i^s$ の制約条件の下で停留点を求めることで，次のメッセージ伝搬則と，周辺分布の計算式を得る．

$$m_{ij}^t \propto \sum_{s=1}^{L} \phi_i^s \psi_{ij}^{st} \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}^s \tag{2.39a}$$

$$p_{ij}^{st} \propto \psi_{ij}^{st} \phi_i^s \phi_j^t \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}^s \prod_{l \in \mathcal{N}_i \setminus i} m_{lj}^t \tag{2.39b}$$

$$p_i^s \propto \phi_i^s \prod_{k \in \mathcal{N}_i} m_{ki}^s \tag{2.39c}$$

実際のアルゴリズムは，Alg.2 の $m_{ij}(x_j), p_i(x_i)$ を上の m_{ij}^t, p_i^s に置き換えるだけでよい．

第3章　TAP方程式を用いた平均場近似の拡張とその応用

本章では，平均場近似及びそれを拡張したTAP(Thouless-Anderson-Palmer) 方程式について議論する．TAP 方程式は物性物理の分野にその起源を持ち，平均場近似と類似する反復計算を行うことで周辺分布を推定する手法である．本章では，これらの手法がコンピュータビジョンの問題– 特に MPM 推定問題–に対しても有用であるかを検証する．また，CV の問題に対する詳細な実装を示すとともに，多値ラベル分類問題 (物理の分野では対応する問題が存在しない) に対し，TAP 方程式およびそのアルゴリズムを新たに導出する．インタラクティブセグメンテーションとステレオマッチングを用いた実験では，高次の TAP 方程式は，精度と計算速度の両面で確率伝搬法を上回ることが分かった．

3.1　概要

本章では平均場近似と TAP 方程式の MPM 推定について焦点を当てる．TAP 方程式は元々物性物理学の分野にその起源を持ち，Sherrington-Kirkpatrick(SK) モデル [51] と呼ばれるモデルの厳密解を求める目的で，Thouless, Anderson, Palmer の 3 名が 1977 年に提案した方程式である [58] [1]．ここで，SK モデルとはスピングラスと呼ばれる，スピンがランダムな方向を向いているが時間的には変化しない，乱雑な秩序を持った物質を記述するための近似モデルである．元の TAP 方程式はエネルギー関数内の変数を平均値で置き換え，粒子数 N に対して $N \to \infty$ の極限をとることで解を導出する手法であったが，後に Plefka が自由エネルギーを用いた別の方法によっても，同様の方程式を導けることを示した [46]．この手法は一般に Plefka 展開と呼ばれる．

Plefka 展開を用いた TAP の手法は，平均場近似や確率伝搬法と同様に，自由エネルギー最小の変分原理に基づく．このとき，TAP の手法での自由エネルギー (TAP 自由エネル

[1] なお，TAP 方程式と同様の解法は既に森田らが 1976 年に提案していた [39] ものの，物性物理学分野では TAP 方程式がその初出であると一般的に理解されている．

ギー) は Taylor 展開の形をした式として導出される．そして，一次の TAP 自由エネルギーは平均場近似における自由エネルギーと一致することが知られている．ゆえに，TAP 方程式は平均場近似を理論的に拡張した手法とみなすこともできる．事実，複数の実験によって，二次の TAP 方程式が平均場近似を上回る精度の周辺分布を得られていることが報告されている [21, 20, 61, 34]．しかしながら，コンピュータビジョンの領域では，この TAP 方程式を用いた周辺分布の推定手法は全く知られていない．コンピュータビジョンでは，MPM 推定を行うための手法としては Loopy Belief Propagation(LBP)[40] に代表される確率伝搬法の手法 [63] が主に使用されていた．

前述のとおり，平均場近似と TAP 方程式は双方とも物性物理学にそのルーツを持つ．コンピュータビジョンの分野では，90 年代初期に平均場近似の応用がいくつか見られたが [35, 64]，その後の確率伝搬法の隆盛もあって主流にはならなかった．[60] で確率伝搬法と平均場近似の性能が比較されているが，これは木構造を持つ (ループのない)MRF に対してである．しかも，ここで用いている平均場近似は最も単純な 1 次の TAP 方程式であり，高次の TAP 方程式に関しては実験されていない．

このあいだ，ニューラルネットワークや情報理論などの各分野で，平均場近似について重要な発展がいくつかあった．一つは自由エネルギー最小化の観点から，平均場近似を再解釈する新たな枠組みが出来たことである [44, 22, 1]．この枠組みの詳細は既に §2.4 で述べた．自由エネルギー最小化の観点によると，確率伝搬法および平均場近似は，ともに自由エネルギーの局所解を探索する方法であると捉えられる (ただし前述のとおり，近似分布の確率モデルには若干の違いがある)．さらに，平均場近似の精度を改善する方法として従来から知られていた TAP 方程式の方法 [58] に，新しい理論的基礎 [44, 46, 1] が与えられた．これによると TAP 自由エネルギーは最小自由エネルギーの Taylor 展開に基づき，平均場近似のより高精度な解を得ることを目指す方法であるとみなせる．この新しい TAP 方程式の観点が前述した Plefka 展開である．以下，平均場近似と TAP 方程式の手法を総称して，MF-TAP と呼ぶ．

われわれは，この MF-TAP の方法を，コンピュータビジョンの諸問題に応用することを考えた．本章では，コンピュータビジョンの問題のためのアルゴリズムの導出を行い，その実装方法を示した上で，いくつかの実用的な問題を用いて MF-TAP の性能を検証する．

MF-TAP の確率伝搬法に対する (可能性を含めた) 長所として次の 3 つが挙げられる．

- MRF モデルの選択自由度の高さ．確率伝搬法と異なり，扱い得る確率密度関数のモ

デルはガウシアン分布や離散分布だけに制約されない.

- 計算がより高速に行えること. 少なくとも両者ともにナイーブな実装を行った場合, サイト 1 つあたりのエッジの数と同じ倍率で, 計算量が小さい. また, GPGPU などの並列計算にも適している.

- 推定精度の向上. 一般に平均場近似は確率伝搬法より精度が劣るが, MF-TAP の高次の項まで利用する場合では, 確率伝搬法より高い精度を達成しうる.

上の 3 項目の中では最初の項目が最も魅力的だが, その検証は別の研究に譲ることとし, 本章では離散ラベル分類問題に限定して, MF-TAP の有効性を調べる.

そのために本章では, 多値ラベル分類問題に適用可能な MF-TAP のアルゴリズムを導出する. TAP 方程式は方法論であり, 各問題 (MRF のモデル, エネルギー関数, 周辺分布の表現) ごとに, アルゴリズムを導出する必要がある. コンピュータビジョンでの二値ラベル分類問題については既に MF-TAP のアルゴリズムが物性物理学の方面から導出されている. しかしながら, 多値ラベル分類問題–これはコンピュータビジョンの領域で頻繁に用いられる, 二値ラベル分類問題の一般形である–に対する MF-TAP のアルゴリズムは, コンピュータビジョンだけでなく異分野でもこれまで導出されていない.

本研究の貢献は以下の 2 点に集約される:

- コンピュータビジョンの問題に対する MF-TAP の方法の実装を詳細に示す. 特に, 多値ラベル分類問題のための MF-TAP のアルゴリズムを新しく導出する.

- MF-TAP の方法が, 確率伝搬法よりも優れた性能を示す場合のあることを, 理論と実験の両面で示す.

本章の構成は下記の通りである. §3.2 では連続的なデータ項, 平滑化項から成る一般的な TAP 方程式を導出した後に, 具体的な二値ラベル, 多値ラベルでの TAP 方程式を導出する. §3.3 ではコンピュータビジョンの領域で主に用いられる確率伝搬法に対する, TAP 方程式の利点について説明する. §3.4 ではインタラクティブセグメンテーションとステレオマッチングの実験を通して, 提案手法の有効性を示す. §3.5 は本章の結論である.

3.2 TAP 方程式の導出

本節では TAP 方程式の理論の概要と，その具体的な導出手法について説明する．なお，本節の TAP 方程式の導出には，Geoirges ら [15] と Plefka[46] 双方の導出方法を参考にした．

3.2.1 TAP 方程式の基本概念

前述のとおり，平均場近似，確率伝搬法ではまず近似分布 P の確率モデルを適当に設定する．具体的には，平均場近似はすべてのサイトが独立であるという確率モデル ((2.17) 式) を近似分布として採用しており，確率伝搬法は平均場近似をより拡張した確率モデル ((2.28) 式) を近似分布として採用している．そして，両者ともに近似分布 P を真の分布 Q に近づけることで，周辺分布を近似的に求める戦略を用いていた．Plefka 展開による TAP 方程式も同様の枠組みで近似分布を求めるが，その求め方は前述の手順とは多少異なる．具体的には，Plefka 展開による TAP 方程式は P を動かすことで近似分布を求めるのではなく，後述する制約条件の $\hat{P}$ を動かすことで近似分布を求める．

以下に具体的な導出手順を示す．まず，Plefka 展開では P に関する制約条件として，P の周辺分布 $p_i(x_i)$ が，ある適当な分布 $\hat{p}_i(x_i)$ に等しいものとする．つまり，

$$p_i(x_i) = \hat{p}_i(x_i) \tag{3.1}$$

である．この制約条件の下で自由エネルギーを最適化するために，Lagrange 乗数 $\lambda_i(x_i)$ を付与した次の Lagrange 関数 $F[P, \hat{P}]$ を導入する．

$$\begin{aligned} F[P,\hat{P}] &= \beta \langle E \rangle_P - S[P] + \sum_i \int \lambda_i(x_i) \left\{ \hat{p}_i(x_i) - p_i(x_i) \right\} dx_i \\ &= \beta \langle E \rangle_P - \sum_i \int \lambda_i(x_i) p_i(x_i) dx_i - S[P] + \sum_i \int \lambda_i(x_i) \hat{p}_i(x_i) dx_i \end{aligned} \tag{3.2}$$

次に，$F[P, \hat{P}]$ を P に関して最小化する．ここで，(3.2) 式の第二項を P に関する λ_i の期待値と見ると，(3.2) 式の第二項は第一項の中に含められる．つまり，

$$F[P,\hat{P}] = \left\langle \beta E(\mathbf{x}) - \sum_i \lambda_i(x_i) \right\rangle_P - S[P] + \sum_i \int \lambda_i(x_i) \hat{p}_i(x_i) dx_i \tag{3.3}$$

と書ける．ここで (3.3) 式の第一項と第二項に注目する．この二項を新しい自由エネルギー $F'[P] \equiv \langle \beta E(\mathbf{x}) - \sum_i \lambda_i(x_i) \rangle_P - S[P]$ とみなすと，(2.15) 式より $F'[P]$ を最小にする P は

次の分布 $Q_\beta(\mathbf{x})$ と等価である．

$$Q_\beta(\mathbf{x}) = \frac{1}{Z}\exp\left(-\beta E(\mathbf{x}) + \sum_i \lambda_i(x_i)\right) \tag{3.4}$$

ゆえに，(2.15) 式の関係から P を最小化する $F[P,\hat{P}]$ は次の $F[\hat{P}]$ となる．

$$F[\hat{P}] = \sum_i \int \lambda_i(x_i)\hat{p}_i(x_i)dx_i - \ln\left[\int \exp\left(\beta E(\mathbf{x}) + \sum_i \lambda_i(x_i)\right)d\mathbf{x}\right] \tag{3.5}$$

この $F[\hat{P}]$ は制約条件である $\hat{P}$ を引数にとる汎関数である．以上から，TAP 方程式では制約条件である $\hat{P}$ を動かすことで，最適となる $\hat{P}$ を求める．

しかしながら，(3.5) 式は第二項の多重積分によって，その停留点を解析的に求めることがほとんどの場合不可能である．しかしながら，この式は $\beta = 0$ という場合のみ解析的に計算できる．ゆえに，Plefka 展開は (3.5) 式を $\beta = 0$ まわりでの Taylor 展開によって，$F[\hat{P}]$ の解析解を近似的に求める．

$$F[\hat{P}] = F^0[\hat{P}] + \beta F^1[\hat{P}] + \frac{\beta^2}{2}F^2[\hat{P}]\cdots \tag{3.6}$$

ここで，$F^n[\hat{P}]$ を次式のように定義した．

$$F^n[\hat{P}] = \left.\frac{\partial^n F[\hat{P}]}{\partial \beta^n}\right|_{\beta=0} \tag{3.7}$$

あとは $F^n[\hat{P}]$ の式を計算し，Euler-Lagrange 方程式を用いて $F[\hat{P}]$ の停留点を求めればよい．以下は $F^n[\hat{P}]$ に関する具体的な計算である．

3.2.2　導出のための基本定理

$F^n[\hat{P}]$ を求めるためにははじめにいくつかの基本定理を導出する必要がある．また，計算を簡略化する目的でいくつかの新しい表記を導入する．

添字を用いた積分の簡略化

$F^n[\hat{P}]$ の計算には多くの積分計算を必要とするため，以降の計算では積分計算を簡略化した以下の表記を採用する．

まず，任意の関数 $\mathcal{O}(\mathbf{x})$ に対する $\hat{p}_i(x_i)$ の積分を次のように表す．

$$\langle \mathcal{O} \rangle_i \equiv \int \hat{p}_i(x_i)\mathcal{O}(\mathbf{x})dx_i \tag{3.8}$$

同様に，$\mathcal{O}(\mathbf{x})$ に対する $\hat{p}_i(x_i), \hat{p}_j(x_j)$ の積分を次のように表す．

$$\langle \mathcal{O} \rangle_{ij} \equiv \iint \hat{p}_i(x_i)\hat{p}_j(x_j)\mathcal{O}(\mathbf{x})dx_i dx_j \tag{3.9}$$

次に，右下に何も表記されていないブラケット記法 $\langle \cdot \rangle$ を，確率分布 $\hat{p}(\mathbf{x}) = \prod_i \hat{p}(x_i)$ に対する期待値とする．

$$\langle \mathcal{O} \rangle \equiv \int \left(\prod_i \hat{p}_i(x_i) \right) \mathcal{O}(\mathbf{x})d\mathbf{x} \tag{3.10}$$

例えば，$\langle f_{ij} \rangle$ は分布 $\hat{p}(\mathbf{x})$ に対する関数 f_{ij} の積分を表し，$\langle f_{ij} \rangle_{ij}$ に等しい．

$$\langle f_{ij} \rangle = \langle f_{ij} \rangle_{ij} = \iint \hat{p}_i(x_i)\hat{p}_j(x_j)f_{ij}(x_i, x_j)dx_i dx_j \tag{3.11}$$

$\beta \to 0$ の極限における $F[\hat{P}]$(0 次の TAP 自由エネルギー)

最も簡単な例として，$\beta \to 0$ の極限における $F[\hat{P}]$ の式を導出する．これは (3.6) 式における 0 次の TAP 自由エネルギーに等しい．$\beta \to 0$ の極値では $E(\mathbf{x})$ は消失するため，多重積分は次の簡単な形へ変形できる．

$$\begin{aligned}
\lim_{\beta \to 0} F[\hat{P}] &= \sum_i \int \lambda_i(x_i)\hat{p}_i(x_i)dx_i - \ln \left[\int \exp\left(\sum_i \lambda_i(x_i) \right) dx_i \right] \\
&= \sum_i \int \lambda_i(x_i)\hat{p}_i(x_i)dx_i - \sum_i \ln \left[\int \exp\left(\lambda_i(x_i) \right) dx_i \right]
\end{aligned} \tag{3.12}$$

Euler-Lagrange 方程式より，$\lambda_i(x_i)$ の停留点は次の関係を持つ．

$$\hat{p}_i(x_i) = \frac{\exp\left(\lambda_i(x_i) \right)}{\int \exp\left(\lambda_i(x_i) \right) dx_i} \tag{3.13}$$

(3.13) 式より，$\lambda_i(x_i)$ の停留点は $\lambda_i(x_i) = \ln \hat{p}_i + C$ (C は不定値) の関係を持つため一意でないが，この不定値は $F[\hat{P}]$ の形に影響を受けない．ゆえに $\lambda_i(x_i) = \ln \hat{p}_i(x_i)$ とすると，(3.12) 式は次式で表せられる．

$$\lim_{\beta \to 0} F[\hat{P}] = \sum_i \int \hat{p}_i(x_i) \ln \hat{p}_i(x_i)dx_i = -S[\hat{P}] \tag{3.14}$$

これは，0 次の TAP 自由エネルギーが分布 $\hat{P}$ のエントロピーに等しいことを示している．

$\langle \cdot \rangle_\beta$, $U_\beta(\mathbf{x})$ の導入

はじめに，任意の関数 $\mathcal{O}(\mathbf{x})$ における，(3.4) 式で導入した分布 $Q_\beta(\mathbf{x})$ の期待値を $\langle \mathcal{O} \rangle_\beta$ と表現する．

$$\langle \mathcal{O} \rangle_\beta \equiv \int Q_\beta(\mathbf{x})\mathcal{O}(\mathbf{x})d\mathbf{x} \tag{3.15}$$

ここで，$\beta \to 0$ の極限では λ_i の停留点は $\lambda_i(x_i) = \ln \hat{p}_i(x_i)$ となる．ゆえに，$\beta \to 0$ の極限において $\langle \mathcal{O} \rangle_\beta$ は分布 $\hat{P}$ の期待値に一致する．すなわち

$$\lim_{\beta \to 0} \langle \mathcal{O} \rangle_\beta = \langle \mathcal{O} \rangle_0 = \langle \mathcal{O} \rangle = \int \left(\prod_i \hat{p}_i(x_i) \right) \mathcal{O}(\mathbf{x})d\mathbf{x} \tag{3.16}$$

の関係を持つ．

次に，$\langle \mathcal{O} \rangle_\beta$ を β に関して微分したときの関係について導出する．

$$\begin{aligned}\frac{\partial \langle \mathcal{O} \rangle_\beta}{\partial \beta} = \left\langle \frac{\partial \mathcal{O}}{\partial \beta} \right\rangle_\beta + \left\langle \mathcal{O}(\mathbf{x}) \left(-E(\mathbf{x}) + \sum_i \frac{\partial \lambda_i}{\partial \beta} \right) \right\rangle_\beta \\ - \langle \mathcal{O} \rangle_\beta \left\langle -E(\mathbf{x}) + \sum_i \frac{\partial \lambda_i}{\partial \beta} \right\rangle_\beta\end{aligned} \tag{3.17}$$

ここで，新しい関数 $U_\beta(\mathbf{x})$ を次式で定義する．

$$U_\beta(\mathbf{x}) \equiv E - \langle E \rangle_\beta - \sum_i \left(\frac{\partial \lambda_i}{\partial \beta} - \left\langle \frac{\partial \lambda_i}{\partial \beta} \right\rangle_\beta \right) \tag{3.18}$$

(3.18) 式によって，(3.17) 式は次式で書き直せる．

$$\frac{\partial \langle \mathcal{O} \rangle_\beta}{\partial \beta} = \left\langle \frac{\partial \mathcal{O}}{\partial \beta} \right\rangle_\beta - \langle \mathcal{O} U_\beta \rangle_\beta + \langle \mathcal{O} \rangle_\beta \langle U_\beta \rangle_\beta \tag{3.19}$$

ここで，明らかに $\langle U_\beta \rangle_\beta = 0$ であるため，(3.19) 式は次式で表せられる．

$$\frac{\partial \langle \mathcal{O} \rangle_\beta}{\partial \beta} = \left\langle \frac{\partial \mathcal{O}}{\partial \beta} \right\rangle_\beta - \langle \mathcal{O} U_\beta \rangle_\beta \tag{3.20}$$

$\beta \to 0$ の極限下における $U_\beta(\mathbf{x})$

次に，$\beta \to 0$ の極限下における (3.18) 式の $U_\beta(\mathbf{x})$ の性質について調べる．はじめに，$\partial \lambda_i / \partial \beta$ は (3.5) 式の関係より次式で表せられる．

$$\frac{\partial \lambda_i(x_i)}{\partial \beta} = \frac{\partial^2 F[\hat{P}]}{\partial \hat{p}_i(x_i) \partial \beta} \tag{3.21}$$

ここで，$F[\hat{P}]$ が $\beta, \hat{p}_i(x_i)$ に対して十分滑らかであると仮定すると，$\partial\lambda_i/\partial\beta$ の $\beta \to 0$ に関する極限は次式で表現できる．

$$\left.\frac{\partial\lambda_i(x_i)}{\partial\beta}\right|_{\beta=0} = \frac{\partial\langle E\rangle}{\partial\hat{p}_i(x_i)} = f_i(x_i) + \sum_{j\in\mathcal{N}_i}\int\hat{p}_j(x_j)f_{ij}(x_i,x_j)dx_j$$
$$= f_i + \sum_{j\in\mathcal{N}_i}\langle f_{ij}\rangle_j \tag{3.22}$$

(3.22) 式を用いることで，U_β は次の関係を持つ．

$$\begin{aligned}\lim_{\beta\to0}U_\beta &= \sum_i f_i + \sum_{i,j} f_{ij} - \sum_i\langle f_i\rangle - \sum_{i,j}\langle f_{ij}\rangle \\ &\quad - \sum_i\left(f_i + \sum_{j\in\mathcal{N}_i}\langle f_{ij}\rangle_j - \langle f_i\rangle - \sum_{j\in\mathcal{N}_i}\langle\langle f_{ij}\rangle_j\rangle_i\right) \\ &= \sum_{i,j} f_{ij} - \sum_{i,j}\langle f_{ij}\rangle - \sum_{i,j}\langle f_{ij}\rangle_j - \sum_{i,j}\langle f_{ij}\rangle_i + 2\sum_{i,j}\langle\langle f_{ij}\rangle_j\rangle_i \\ &= \sum_{i,j} f_{ij} - \sum_{i,j}\langle f_{ij}\rangle_j - \sum_{i,j}\langle f_{ij}\rangle_i + \sum_{i,j}\langle f_{ij}\rangle \end{aligned} \tag{3.23}$$

ここで，平滑化項の対称性 $f_{ij}(x_i,x_j) = f_{ji}(x_j,x_i)$ より，

$$\sum_i\sum_{j\in\mathcal{N}_i} f_{ij}(x_i,x_j) = 2\sum_{i,j} f_{ij}(x_i,x_j)$$

の関係を用いた．以上より，U_0 は関数 $U_{ij}(x_i,x_j)$ を用いた

$$U_0 = \sum_{i,j} U_{ij}(x_i,x_j) = \sum_{i,j}\Big(f_{ij} - \langle f_{ij}\rangle_j - \langle f_{ij}\rangle_i + \langle f_{ij}\rangle\Big) \tag{3.24}$$

の形で表現できる．

3.2.3　1 次の TAP 自由エネルギーの導出

以上の関係を用いて，1 次の TAP 自由エネルギーを導出する．まず，(3.5) 式を β で微分することで，次式を得る．

$$\frac{\partial F[\hat{P}]}{\partial\beta} = \sum_i\left\langle\frac{\partial\lambda_i}{\partial\beta}\right\rangle_0 + \langle E\rangle_\beta - \sum_i\left\langle\frac{\partial\lambda_i}{\partial\beta}\right\rangle_\beta \tag{3.25}$$

$\beta \to 0$ の極限では，(3.25) 式の第一項と第三項は打ち消し合う．また，(3.14) 式の関係から 0 次の TAP 方程式は $-S[\hat{P}]$ である．ゆえに，1 次の TAP 自由エネルギーは次式で表せられる．

$$F[\hat{P}] \approx -S[\hat{P}] + \langle E \rangle_{\hat{P}} \tag{3.26}$$

(2.18) 式から，これは平均場近似での自由エネルギーの形 ((2.18) 式) に等しい．ゆえに，TAP 自由エネルギーは平均場近似での自由エネルギーを理論的に拡張した式とみなすこともできる．

3.2.4　2 次の TAP 自由エネルギーの導出

2 次の TAP 方程式に関しても 1 次の TAP 方程式と同様に導出する．(3.5) 式を β に関して 2 回微分することで，次式を得る．

$$\frac{\partial^2 F[\hat{P}]}{\partial \beta^2} = \sum_i \left\{ \left\langle \frac{\partial^2 \lambda_i}{\partial \beta^2} \right\rangle_0 - \left\langle \frac{\partial^2 \lambda_i}{\partial \beta^2} \right\rangle_\beta + \left\langle \frac{\partial \lambda_i}{\partial \beta} U_\beta \right\rangle_\beta \right\} - \langle E U_\beta \rangle_\beta \tag{3.27}$$

ここで，(3.27) 式の第三項について $\beta \to 0$ の極限をとると，次式を得る．

$$\left\langle \frac{\partial \lambda_i}{\partial \beta} U_\beta \right\rangle_\beta \to \left\langle (f_i + \sum_{j \in \mathcal{N}_i} \langle f_{ij} \rangle_j)(f_{ij} - \langle f_{ij} \rangle_i - \langle f_{ij} \rangle_j + \langle f_{ij} \rangle_{ij}) \right\rangle$$

上の式を具体的に計算する．まず f_i の項について，

$$\begin{aligned} &\langle f_i (f_{ij} - \langle f_{ij} \rangle_i - \langle f_{ij} \rangle_j + \langle f_{ij} \rangle_{ij}) \rangle \\ &\qquad = \langle f_i \langle f_{ij} \rangle_j \rangle_i - \langle f_i \rangle_i \langle f_{ij} \rangle_{ij} - \langle f_i \langle f_{ij} \rangle_j \rangle_i + \langle f_i \rangle_i \langle f_{ij} \rangle_{ij} = 0 \end{aligned}$$

次に $\langle f_{ij} \rangle_j$ の項について，

$$\begin{aligned} &\langle \langle f_{ij} \rangle_j f_{ij} \rangle_{ij} - \langle \langle f_{ij} \rangle_j \langle f_{ij} \rangle_i \rangle_{ij} - \langle \langle f_{ij} \rangle_j^2 \rangle_{ij} + \langle \langle f_{ij} \rangle_j \langle f_{ij} \rangle_{ij} \rangle_{ij} \\ &\qquad = \langle \langle f_{ij} \rangle_j^2 \rangle_i - \langle f_{ij} \rangle_{ij}^2 - \langle \langle f_{ij} \rangle_j^2 \rangle_i + \langle f_{ij} \rangle_{ij}^2 = 0 \end{aligned}$$

以上から，(3.27) 式の第三項は $\beta \to 0$ の極限において

$$\left\langle \frac{\partial \lambda_i}{\partial \beta} U_\beta \right\rangle_\beta \to 0$$

の関係を得る．以上から，$\beta \to 0$ の極限において (3.27) 式の第一項と第二項は打ち消し合い，第三項は 0 となることが分かる．ゆえに，$\partial^2 F[\hat{P}]/\partial\beta^2$ は $\beta \to 0$ の極限下で次の関係を持つ．

$$\left.\frac{\partial^2 F}{\partial \beta^2}\right|_{\beta=0} = -\langle EU_0\rangle = -\langle U_0^2\rangle \tag{3.28}$$

上式の $\langle EU_0\rangle = \langle U_0^2\rangle$ の関係については，(3.27) 式の第三項の計算と同様の手順で導出できる．以上から，$\langle U_0^2\rangle$ は (3.24) 式の関係より

$$\langle U^2\rangle = \sum_{i,j}\sum_{k,l}\langle U_{ij}U_{kl}\rangle \tag{3.29}$$

の形で書ける．ここで，$(i,j),(k,l)$ の総和を次の 3 つの場合に分ける．

i, j, k, l がすべて異なる場合

定義より U_{ij}, U_{kl} はそれぞれ独立となる．また，$\langle U_{ij}\rangle = 0$ の関係が得られるため，

$$\langle U_{ij}U_{kl}\rangle = \langle U_{ij}\rangle\langle U_{kl}\rangle = 0 \tag{3.30}$$

$(i,j) = (k,l)$ の場合

計算によって，$\left\langle U_{ij}^2\right\rangle$ は最終的に次の関係を満たす．

$$\begin{aligned}
\langle U_{ij}U_{ij}\rangle &= \left\langle\left(f_{ij} - \langle f_{ij}\rangle_j - \langle f_{ij}\rangle_i + \langle f_{ij}\rangle\right)^2\right\rangle \\
&= \langle f_{ij}^2\rangle + \langle\langle f_{ij}\rangle_i^2\rangle_j + \langle\langle f_{ij}\rangle_j^2\rangle_i + \langle f_{ij}\rangle^2 \\
&\qquad - 2\left\langle f_{ij}\langle f_{ij}\rangle_j\right\rangle_{ij} - 2\left\langle f_{ij}\langle f_{ij}\rangle_i\right\rangle_{ij} \\
&= \langle f_{ij}^2\rangle - \langle\langle f_{ij}\rangle_i^2\rangle_j - \langle\langle f_{ij}\rangle_j^2\rangle_i + \langle f_{ij}\rangle^2
\end{aligned} \tag{3.31}$$

$i = k, j \neq l$ の場合

計算によって，$\langle U_{ij}U_{il}\rangle$ の各項はすべて打ち消し合い，0 となる．

$$\begin{aligned}
\langle U_{ij}U_{il}\rangle &= \left\langle\left(f_{ij} - \langle f_{ij}\rangle_j - \langle f_{ij}\rangle_i + \langle f_{ij}\rangle\right)\left(f_{il} - \langle f_{il}\rangle_l - \langle f_{il}\rangle_i + \langle f_{il}\rangle\right)\right\rangle \\
&= \langle f_{ij}f_{il}\rangle - \left\langle f_{ij}\langle f_{il}\rangle_l\right\rangle - \left\langle\langle f_{ij}\rangle_j f_{il}\right\rangle + \left\langle\langle f_{ij}\rangle_j\langle f_{il}\rangle_l\right\rangle \\
&= \langle f_{ij}f_{il}\rangle - \langle f_{ij}f_{il}\rangle - \langle f_{ij}f_{il}\rangle + \langle f_{ij}f_{il}\rangle = 0
\end{aligned} \tag{3.32}$$

(3.30) 式, (3.31) 式, (3.32) 式の結果より，$\langle U_0^2 \rangle$ は最終的に次式の形を取る．

$$\begin{aligned}\left\langle U_0^2 \right\rangle &= \sum_{i,j} \left\langle U_{ij}^2 \right\rangle \\ &= \sum_{i,j} \Big(\langle f_{ij}^2 \rangle - \langle \langle f_{ij} \rangle_i^2 \rangle_j - \langle \langle f_{ij} \rangle_j^2 \rangle_i + \langle f_{ij} \rangle^2 \Big) \end{aligned} \tag{3.33}$$

ゆえに，(3.28) 式, (3.33) 式の関係から，2 次の TAP 方程式は次式の形で求められる．

$$\begin{aligned} F[\hat{P}] &\approx -S[\hat{P}] + \langle E \rangle_{\hat{P}} - \frac{\beta^2}{2} \langle U_0^2 \rangle_{\hat{P}} \\ &= -S[\hat{P}] + \langle E \rangle_{\hat{P}} - \frac{\beta^2}{2} \sum_{i,j} \langle U_{ij}^2 \rangle_{\hat{P}} \end{aligned} \tag{3.34}$$

3.2.5　3 次の TAP 自由エネルギーの導出

3 次の TAP 方程式に関しても 2 次と同様に導出する．まず，(3.5) 式を β に関して 3 回微分することで，次式を得る．

$$\begin{aligned} \frac{\partial^3 F[\hat{P}]}{\partial \beta^3} = \sum_i \Bigg\{ & \left\langle \frac{\partial^3 \lambda_i}{\partial \beta^3} \right\rangle_0 - \left\langle \frac{\partial^3 \lambda_i}{\partial \beta^3} \right\rangle_\beta + 2 \left\langle \frac{\partial^2 \lambda_i}{\partial \beta^2} U_\beta \right\rangle_\beta \\ & - \left\langle \frac{\partial \lambda_i}{\partial \beta} U_\beta^2 \right\rangle_\beta + \left\langle \frac{\partial \lambda_i}{\partial \beta} \frac{\partial U_\beta}{\partial \beta} \right\rangle_\beta \Bigg\} - \left\langle E \frac{\partial U_\beta}{\partial \beta} \right\rangle_\beta + \left\langle E U_\beta^2 \right\rangle_\beta \end{aligned} \tag{3.35}$$

§3.2.4 と同様の計算を行うことで，$\beta \to 0$ の極限において，(3.35) 式の第一項，第二項は打ち消し合って 0 となり，第三項から六項は 0 へ収束する．ゆえに，(3.35) 式は $\beta \to 0$ の極限で次の関係を得る．

$$\left. \frac{\partial^2 F}{\partial \beta^2} \right|_{\beta=0} = \left\langle E U_0^2 \right\rangle_0 = \left\langle U_0^3 \right\rangle = \sum_{i,j} \left\langle U_{ij}^3 \right\rangle + 6 \sum_{i,j,k} \langle U_{ij} U_{jk} U_{ki} \rangle \tag{3.36}$$

ここで，総和の添字 (i, j, k) は，グラフ内におけるあらゆる隣接サイト i, j, k の三つ組を表す (ただし $i \neq j \neq k$ で，重複は含まない)．以上から，3 次の TAP 自由エネルギーは次式となる．

$$\begin{aligned} F[\hat{P}] &\approx -S[\hat{P}] + \langle E \rangle_{\hat{P}} - \frac{\beta^2}{2} \langle U_0^2 \rangle_{\hat{P}} + \frac{\beta^3}{6} \langle U_0^3 \rangle_{\hat{P}} \\ &= -S[\hat{P}] + \langle E \rangle_{\hat{P}} - \frac{\beta^2}{2} \sum_{i,j} \langle U_{ij}^2 \rangle_{\hat{P}} \\ &\qquad + \frac{\beta^3}{6} \sum_{i,j} \langle U_{ij}^3 \rangle_{\hat{P}} + \beta^3 \sum_{i,j,k} \langle U_{ij} U_{jk} U_{ki} \rangle_{\hat{P}} \end{aligned} \tag{3.37}$$

3.2.6　二値ラベル分類問題における TAP 方程式

以上の関係を用いて，二値ラベル分類における 3 次の TAP 方程式を導出する．まず，§2.4.2 で定義された関係を (3.37) 式の $F[\hat{P}]$ に代入し，整理することで次の関係を得る．

$$\begin{aligned} F[\mathbf{m}] = \sum_i \left[\frac{1+m_i}{2} \ln\left(\frac{1+m_i}{2}\right) + \frac{1-m_i}{2} \ln\left(\frac{1-m_i}{2}\right) \right] \\ - \beta \sum_i h_i m_i - \beta \sum_{i,j} J_{ij} m_i m_j - \frac{\beta^2}{2} \sum_{i,j} J_{ij}^2 (1-m_i^2)(1-m_j^2) \\ - \frac{2\beta^3}{3} \sum_{i,j} J_{ij}^3 m_i m_j (1-m_i^2)(1-m_j^2) \\ - \beta^3 \sum_{i,j,k} J_{ij} J_{jk} J_{ki} (1-m_i^2)(1-m_j^2)(1-m_k^2) \end{aligned} \tag{3.38}$$

この TAP 自由エネルギーは平均場近似の (2.23) 式を拡張した形の式となる．ゆえに，平均場近似と同様に，$F[\hat{P}]$ を m_i について微分することで，以下の固定点方程式を得る．

$$\begin{aligned} m_i = \tanh\Bigg[\beta h_i + \beta \sum_{j \in \mathcal{N}_i} J_{ij} m_j - \beta^2 \sum_{j \in \mathcal{N}_i} J_{ij}^2 m_i (1-m_j^2) \\ + \frac{2\beta^3}{3} \sum_{j \in \mathcal{N}_i} J_{ij}^3 (1-3m_i^2) m_j (1-m_j^2) \\ - 2\beta^3 \sum_{(j,k) \in \mathcal{N}_i^2} J_{ij} J_{jk} J_{ki} m_i (1-m_j^2)(1-m_k^2) \Bigg] \end{aligned} \tag{3.39}$$

ここで，$(j,k) \in \mathcal{N}_i^2$ はサイト i に隣接する，あらゆるサイトのペアを表す (ただし $j \neq k$ で，重複は含まない)．この式を二値ラベル分類問題における TAP 方程式 (TAP equation) と呼ぶ [15]．この式は (2.24) 式と類似しているが，β に関しての高次の項が含まれている点が異なる．この項の影響によって，TAP 方程式で推定された周辺分布の精度は平均場近似のそれを上回るものと期待される．ただし，この固定点方程式は $\beta \approx 0$ という条件下では正しい周辺分布の値へ収束するものの，一般のコンピュータビジョンの領域で用いられる条件下で収束するかは定かでなく，実験的に確かめるほかない．

3.2.7　多値ラベル分類問題における TAP 方程式

同様に，多値ラベル分類問題における TAP 方程式を導出する．ただし，3 次の TAP 方程式に関しては計算が非常に複雑になるため，今回は 2 次の TAP 方程式に関してのみ導

出する．まず，§2.4.3 で定義された関係を (3.34) 式の $F[\hat{P}]$ に代入し，整理することで次の関係を得る．

$$F[\mathbf{p}] = \sum_i \sum_s p_i^s \ln p_i^s + \beta \sum_i \sum_s p_i^s f_i^s + \beta \sum_{i,j} \sum_{s,t} p_i^s p_j^t f_{ij}^{st} - \frac{\beta^2}{2} \sum_{i,j} \Bigg(\sum_{s,t} p_i^s p_j^t (f_{ij}^{st})^2 - \sum_t p_j^t \big(\sum_s p_i^s f_{ij}^{st} \big)^2 - \sum_s p_i^s \big(\sum_t p_j^t f_{ij}^{st} \big)^2 + \big(\sum_{s,t} p_i^s p_j^t f_{ij}^{st} \big)^2 \Bigg) \quad (3.40)$$

ここで，簡単のため $\hat{\mathbf{p}}$ を $\mathbf{p}$ の形に置換した．この TAP 自由エネルギーは平均場近似の (2.26) 式を拡張した形の式となる．ゆえに，(3.40) 式の $F[\mathbf{p}]$ を $\sum_s p_i^s = 1$ の制約条件の下で最小化するために，Lagrange 乗数を付与した Lagrange 関数を導入し，これを p_i^s に関して微分し停留点を求めることで次の固定点方程式を得る．

$$p_i^s \propto \exp\Bigg[-\beta f_i^s - \beta \sum_{j \in \mathcal{N}_i} \sum_t p_j^t f_{ij}^{st} + \frac{\beta^2}{2} \sum_{j \in \mathcal{N}_i} \Bigg(\sum_t p_j^t (f_{ij}^{st})^2 - \big(\sum_t p_j^t f_{ij}^{st} \big)^2 + 2 \sum_{u,v} p_i^u p_j^v f_{ij}^{uv} (g_{ij}^s - f_{ij}^{sv}) \Bigg) \Bigg] \quad (3.41)$$

ここで，g_{ij}^s を次のようにおいた．

$$g_{ij}^s = \sum_t p_j^t f_{ij}^{st} \quad (3.42)$$

この式を多値ラベル分類問題における TAP 方程式と呼ぶ．二値と同様にこの式も (2.27) 式と類似しているが，β に関しての高次の項が含まれている点が異なる．この項の影響によって，TAP 方程式で推定された周辺分布の精度は二値と同様に，平均場近似のそれを上回るものと期待される．

Potts モデルに対する TAP 方程式

(3.41) 式は一般的な多値ラベル分類問題に対しての TAP 方程式を表しているが，コンピュータビジョンで多く用いられる Potts モデルを採用すると，計算はずっと単純になる．ここで Potts モデルとは，平滑化項 f_{ij}^{st} を次式で単純化したモデルである．

$$f_{ij}^{st} \equiv -J\delta^{st} \quad (3.43)$$

ここで δ^{st} は Kronecker のデルタである．平滑化項を行列で表すと，Potts モデルでは非対角成分が零で対角成分のみ $-J$ の成分を取る正方行列で表現できる．この場合，(3.43) 式を (3.41) 式に代入し，計算することで次の固定点方程式を得る．

$$p_i^s \propto \exp\left[\beta f_i^s + \beta J \sum_{j\in\mathcal{N}_i} p_j^s + \frac{\beta^2}{2} J^2 \sum_{j\in\mathcal{N}_i} p_j^s \left((1-p_j^s) + 2\left(\sum_u p_i^u p_j^u - p_i^s\right)\right)\right] \tag{3.44}$$

このモデルを採用することの利点は，ラベル数を L とすると 1 回毎の反復計算のコストが L^2 ではなく L に比例することにある．すなわち，例え対象の問題が大きなラベル数であったとしても，Potts モデルの下では比較的高速に解の推定値を求められる．

3.2.8　Steffensen の反復法による更新

平均場近似，TAP を問わず上述の導出では，自由エネルギーの停留点は

$$\hat{P} = G[\hat{P}] \tag{3.45}$$

の固定点方程式によって与えられる．平均場近似の場合，反復法による更新は必ず自由エネルギーの局所解に収束することが知られているが，TAP 方程式に関してはその限りでない．事実，われわれの実験では，TAP 方程式の次数を高次に設定した場合，周辺分布の推定解が収束せずに発散してしまう現象がしばしば観測された．これは，TAP 方程式が高次になるに従って，G がより複雑な形をとるためと考えられる．反復法による更新は，$G[\hat{P}]$ の各成分 $\hat{p}_i^{t+1}$ が $|\partial G/\partial \hat{p}_i| < 1$ の条件を満たさなければ発散してしまう．この性質により，TAP 方程式の次数の増加に従って，収束条件はより満たされにくくなる．また，反復法による更新は収束が遅いという欠点もある．この反復は 1 次収束しかないので，収束に至るまでに多くの反復を要する．

そこで以下の実験では，Steffensen の反復法 [19, 7] を用いて周辺分布の更新を行った．Steffensen の反復法は Aitken の Δ^2 法 [7] を改良したアルゴリズムであり，$|\partial G/\partial \hat{p}_i| < 1$ の条件を必要とせず，また 2 次収束する．具体的なアルゴリズムは次の通りである．$t+1$ が 3 の倍数でない場合は通常の反復法と同じく，$\hat{P}^{t+1} = G[\hat{P}^t]$ を更新式とすることで更新を行う．一方，$t+1$ が 3 の倍数である場合は，$\hat{P}$ の各成分における過去の値 $\hat{p}_i^t, \hat{p}_i^{t-1}, \hat{p}_i^{t-2}$ を用いて

$$\hat{p}_i^{t+1} = \hat{p}_i^{t-2} - \frac{(\hat{p}_i^{t-1} - \hat{p}_i^{t-2})^2}{\hat{p}_i^t - 2\hat{p}_i^{t-1} + \hat{p}_i^{t-2}} \tag{3.46}$$

と更新する．

3.3 平均場近似，TAP 方程式と確率伝搬法の比較

本節では，平均場近似ならびに TAP 方程式の確率伝搬法に対する長所を議論する．

3.3.1 MRF モデル選択のより高い自由度

MF-TAP では，確率伝搬法と比べて，選択できる MRF モデル，中でも周辺分布の表現の自由度が高い．変数が連続値をとる場合，確率伝搬法で扱える分布は事実上ガウス分布に固定される一方で，MF-TAP ではそのような制限がない．確率伝搬法に課されるその制約は，(2.37) 式によるメッセージの更新則に由来する．連続的なメッセージを少数のパラメトリックな分布で更新するには，(2.37) 式によって更新されたメッセージが，また同一のパラメータによって表現される必要がある．混合ガウス分布 [52] やパーティクルフィルタ [18] を用いてこの制約を解消する方法もあるが，計算時間や精度維持の難しさなど他の課題が生じる．MF-TAP にはその制約はなく，分布の離散表現はもちろん，様々なパラメトリックな確率密度関数を扱いうる．ただし，扱う確率密度関数ごとに別のアルゴリズムを導出する必要があり，その導出されたアルゴリズムがきちんと収束するかどうかはまた別の問題である．

3.3.2 より高速な計算

離散変数の集合であっても，1 回の反復に要する計算コストに関して言えば，MF-TAP の手法は少なくとも確率伝搬法のナイーブな実装より高速である．MF-TAP では 1 つのサイトを更新する際に，平均場近似と同様，隣接サイトの状態を参照し，そのサイトの状態を更新するだけでよい．その計算コストは，確率伝搬法のステップ (b) におけるメッセージ 1 つの計算にほぼ相当する．従って，確率伝搬法では，サイト 1 つに隣接するサイトの数の倍数だけ，MF-TAP より遅くなる．さらに確率伝搬法ではメッセージ 1 つを計算するのにすべての隣接サイトにアクセスする必要があるので，MF-TAP に比べて必要なメモリアクセス総数が同じ倍数だけ多くなる．GPU 等の並列システム上に実装する場合，これらのシステムでは全体の計算速度がメモリアクセス回数に支配される傾向があるため，両者の計算速度の差はさらに大きくなり得る．

当然ながら，全体の計算コストは反復 1 回の計算速度だけに依存するものではない．同様に重要なファクターとして，収束に至るまでに要する反復回数が挙げられる．これは基本的に各々の問題及びデータに依存するため，実験的にしか調べられない．われわれの実験では，必要な反復回数は平均場近似，TAP 方程式，確率伝搬法とでほぼ同じであるという結果を得ている．これは，確率伝搬法と比較して MF-TAP が隣接サイト数の倍数だけ早いことを表している．

3.3.3　より高い精度

上述の通り，平均場近似，TAP 方程式，確率伝搬法のいづれも，各サイトの周辺分布を特定の反復方程式に従って近似的に計算する．従って各手法の問題は，その近似の精度ということになる．上述したように，平均場近似では各サイトが独立であるという確率モデル ((2.17) 式) に基づく．一方，確率伝搬法ではより自由度の高い確率モデル ((2.28) 式) に基づいているため，周辺分布の精度は多くの場合に平均場近似を上回ることが知られている．

一方，TAP 方程式も確率伝搬法と同様に平均場近似を理論的に拡張した手法であるが，その近似手法は確率伝搬法とは異なる方法で導出されているため，確率モデルの自由度だけで性能を比較することは困難である．ゆえに，実際の問題を用いて実験的に各手法の精度を比較することが重要となる．

ニューラルネットワークの分野では，2 次の TAP 方程式の精度は平均場近似と確率伝搬法の中間にあることが報告されている [61]．その一方で，画像復元の分野では興味深いことに，2 次の TAP 方程式が確率伝搬法を僅かながら上回ることが報告されている [55]．この結果は，TAP 方程式は問題によって確率伝搬法と同等以上の精度が得られることを表している．さらに，双方の実験とも 2 次の TAP 方程式だけの比較であり，より高次の TAP 方程式に関しての実験については，われわれの知る限り存在していない．われわれの実験結果によれば，3 次の TAP 方程式は，一般に確率伝搬法より高精度であった．

3.4　実験

本節では MF-TAP の手法を CV の問題に対する性能を評価するため，2 種類の実験を行った．具体的には，適用対象として二値ラベル分類問題であるインタラクティブセグメ

Table 3.1: 各手法での，100 回の反復に要する計算時間

	CPU[ms]	GPU[ms]	CPU/GPU
平均場近似	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

ンテーションと，多値ラベル分類問題であるステレオマッチングを選んだ．比較のため，LBP も実行したが，元の sum-product ルール [63] をそのまま実装したものを用いた．

実験では，CPU および GPU 上での実行時間を計測し，各手法間で比較した．CPU では Intel Core i7 2.67GHz を，GPU には nVidia GeForce GTX480 を使用した．CPU 上でのコンパイルには MSVC2010 を使用し，最大限の最適化を行う設定に加えて，浮動小数点演算を高速化する `/fp:fast` オプションを指定した上で，OpenMP を用いて 4 コア 8 スレッドの並列計算を行うようにした．

3.4.1 二値ラベル分類問題 (インタラクティブセグメンテーション)

各画素につき 8 つの隣接画素間にエッジがある MRF モデルを対象に，次の手順でインタラクティブセグメンテーションの実験を行った．まず，GrabCut[49] 同様，対象となる画像に対し，前景と背景を人手でマウスを使っておおまかに指定する (指定の一例を Fig. 3.1 中に示す)．その後，この指定を元に前景及び背景の色モデルが作られ，エネルギー関数が生成される．これを用いて，画像の各画素 i が前景であることを示す周辺分布 $p_i(x_i)$ を各手法で計算する．エネルギー関数の生成には OpenCV 2.3 の `calcNWeights()` と `constructGCGraph()` を利用し，そこで指定するパラメータはデフォルトの $\gamma = 50$, $\lambda = 450$ を選んだ．また，パラメータ T は経験的に $T = 80$ とした．画像は，[43] および [37] のデータセットの中から幾つか選んだ．

オリジナルの GrabCut では，最適化と前景及び背景の色モデルの更新を何度か反復することで，より精度の高い結果を得ている．これと同じように各手法を実行したところ，どの方法も (周辺分布をしきい値 $p = 0.5$ で二値化した後の結果が) グラフカットを用いた場合とほぼ同一となり，差異を評価しにくかったので，今回の実験では最適化を一回だけ行うこととした．なお，この場合の色モデルは指定した前景・背景から作ったもので，平均場近似，TAP 方程式，LBP ともに共通である．また，推定された周辺分布の精度を評価

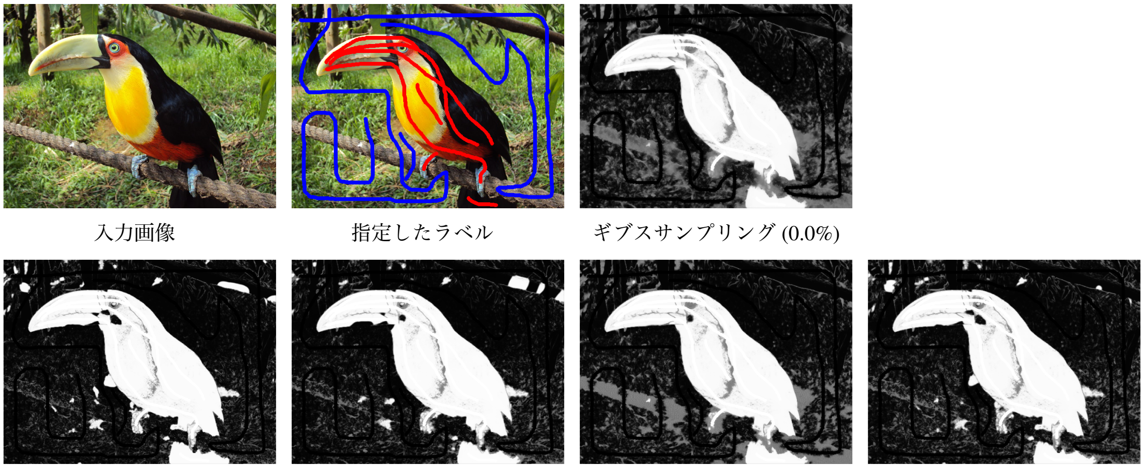


Fig. 3.1: データセット *Bird* を用いた，インタラクティブセグメンテーションの結果．括弧内の数字は収束後の，真の周辺分布からの残差を表す．精度は平均場近似，2 次 TAP，LBP，3 次 TAP の順で改善されている．

するため，ギブスサンプリング [28] を用いて周辺分布を別途計算し，これを真の周辺分布の解とみなした．そこでは 1 画素につき 2 万サンプルを生成することで，十分な時間をかけて高精度な解を求めた．

Fig. 3.1 に一例として，画像 *Bird* ([43]) に対する結果を示す．画像の大きさは 640×480 ピクセルである．計算結果を示す濃淡画像は，濃淡値がその画素が前景である確率を表す．すなわち，白の画像は前景である周辺分布の確率が 1.0 であることを表し，黒は 0.0 であることを表す．また，図の真下にある手法の名称の隣にある括弧内の数字は，ギブスサンプリングによる解に対する誤差を表す．平均場近似 (13.0%)，2 次 TAP(12.0%)，LBP(10.6%)，3 次 TAP(6.1%) の順に精度が良くなっていることが分かる．Fig. 3.2 は，同じ画像に対する各手法の 1 画素あたりの誤差と反復回数の関係を表す．TAP 方程式の次数を 1(平均場近似)，2，3 と上げるに従って精度が向上していることは，TAP 方程式の狙い通りに自由エネルギーの精度が向上していることによると考えられる．また LBP は，2 次の TAP 方程式よりは良いが，3 次の TAP 方程式には劣ることが分かる．

Table 3.1 に，同じ画像に対して各手法 100 回の反復を行ったときの計算時間を示す．まず CPU および，GPU を問わず，平均場近似及び TAP の手法は，LBP に比べて 6-10 倍程度高速であることが見て取れる．これは，LBP が 1 回の反復につき各画素で 8 方向のメッセージを計算する必要がある一方，平均場近似及び TAP の手法は代替その 1 方向分の計算

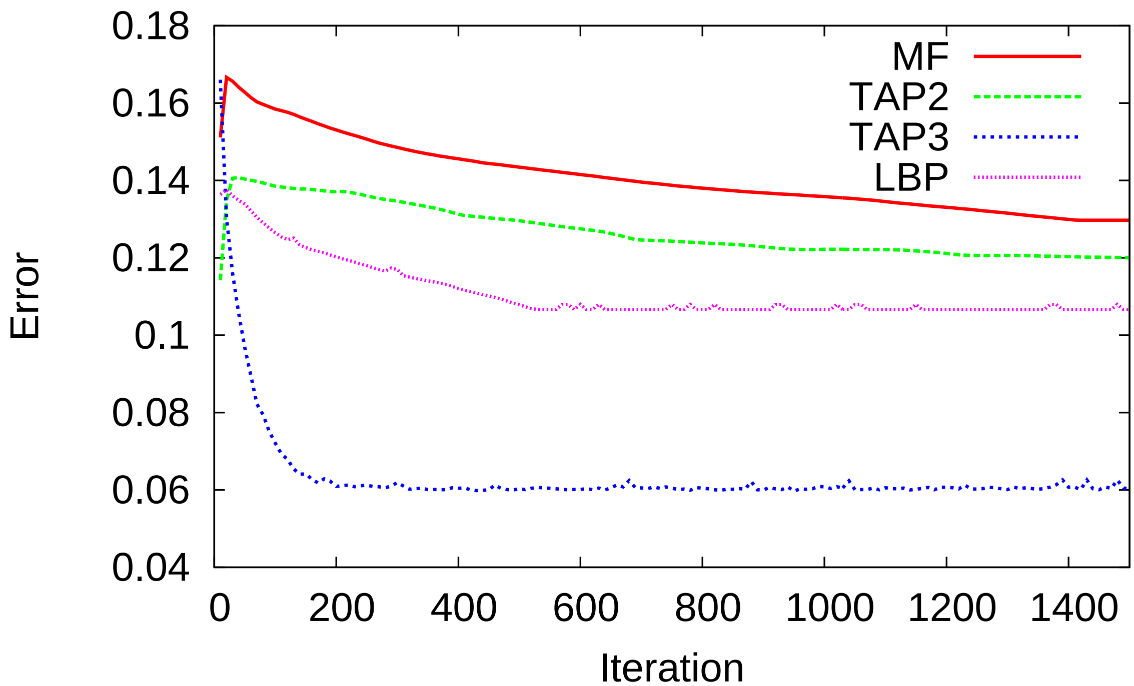


Fig. 3.2: 反復回数毎における，画像 *Bird* を用いたインタラクティブセグメンテーションでの 1 画素辺りの平均誤差.

で済むことによる．また，平均場近似及び TAP の手法は，GPU 上で実装した場合，CPU 上に実装した場合と比較して 40-50 倍の高速化を達成できている．これは，平均場近似及及び TAP の手法が LBP 以上に並列計算 (特に現代の GPU アーキテクチャ) に適していることを示している．

同様の実験を色々な画像を用いて行ったところ，ほぼ同様の結果を得た，Fig. 3.3 に，[37] 中のデータセットから適当に選んだ 4 枚の画像 *Flower*, *Horse*, *Starfish*, *Tiger* に対する，各手法の結果を示す．また，Fig. 3.5 に収束後の 1 画素あたりの平均誤差を示す．これらの図より，画像ごとに若干の違いはあるものの，平均場近似，2 次 TAP，LBP，3 次 TAP の順に精度が向上していることが分かる．

3.4.2　多値ラベル分類問題 (ステレオマッチング)

次に，多値ラベル分類問題の例としてステレオマッチングを行った．各画素につき 8 隣接画素間にエッジのある MRF を対象に，Middlebury の MRF エネルギー最小化ライブラリ [54] を用いてエネルギー関数の生成を行い，平滑化項のパラメータを $|L| = 16, \lambda = 15,$

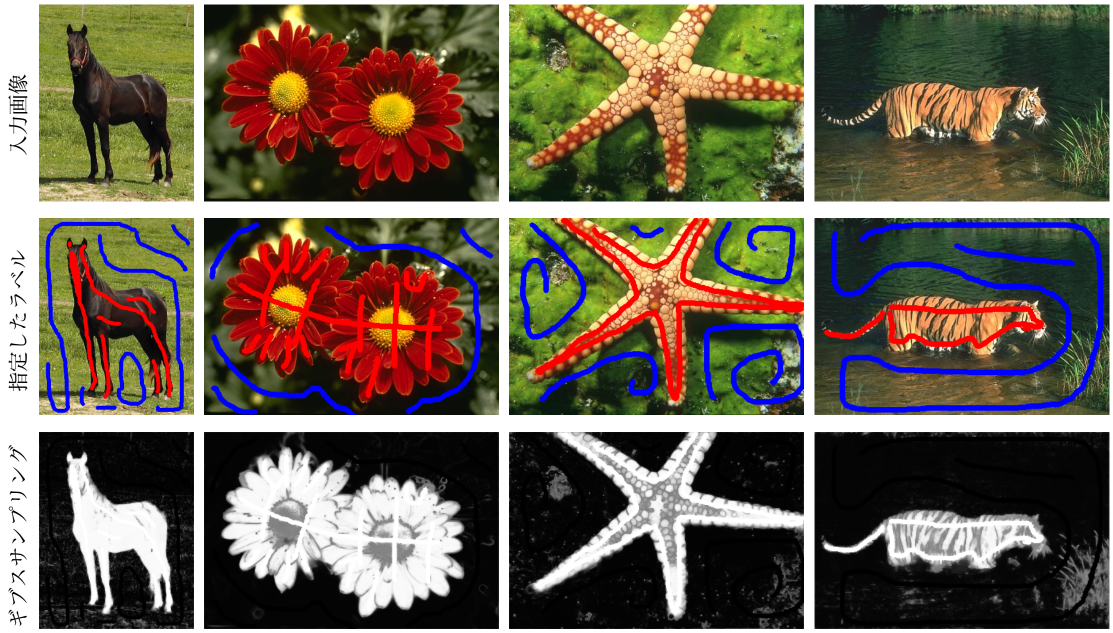


Fig. 3.3: インタラクティブセグメンテーションの実験に用いた，その他の 4 枚の画像: (*Horse, Flower, Starfish, Tiger*)．上段: 入力画像．中段: ラベル画像．下段: ギブスサンプリングで生成した周辺分布の画像．

`truncated` $= 1$ とした．すなわち，このエネルギー分布は Potts モデルを選択したものと等価である．また，温度は経験的に $T = 20$ とした．前節と同様，ギブスサンプリング法を用いて周辺分布を高精度に推定し (1 画素あたり 2 万サンプルを生成)，これを真の解とみなして各手法の誤差を評価した．

Tsukuba データに対してステレオマッチングを行い，十分収束した後の推定結果を Fig. 3.6 に示す．ここでは，$p_i(x_i)$ の最大値を与える x_i(ラベル値) を表示している．各図の下の括弧内の数値は各手法の精度を示すが，前節の結果同様に，平均場近似，2 次の TAP 方程式，LBP の順に誤差が小さく，誤差の大小も同様の傾向がある．また，Fig. 3.7 は同じ画像に対する各手法の 1 画素あたりの誤差と反復回数の関係を表す．まだ試してはいないものの，多値ラベル分類問題についても 3 次の TAP 方程式を導出して計算に用いれば，前節同様に LBP の精度を上回るものと期待される．各手法について，100 回の反復に要する計算時間を測定したところ，平均場近似，2 次の TAP 方程式，LBP の順にそれぞれ 2.81, 3.94, 14.74 秒であった．今考えているマルコフ確率場のモデルでは，各画素あたり 8 つのエッジが存在するので，理論上 MF-TAP は LBP よりも 8 倍高速なはずであり，実測値としてはそれに近くなっている．

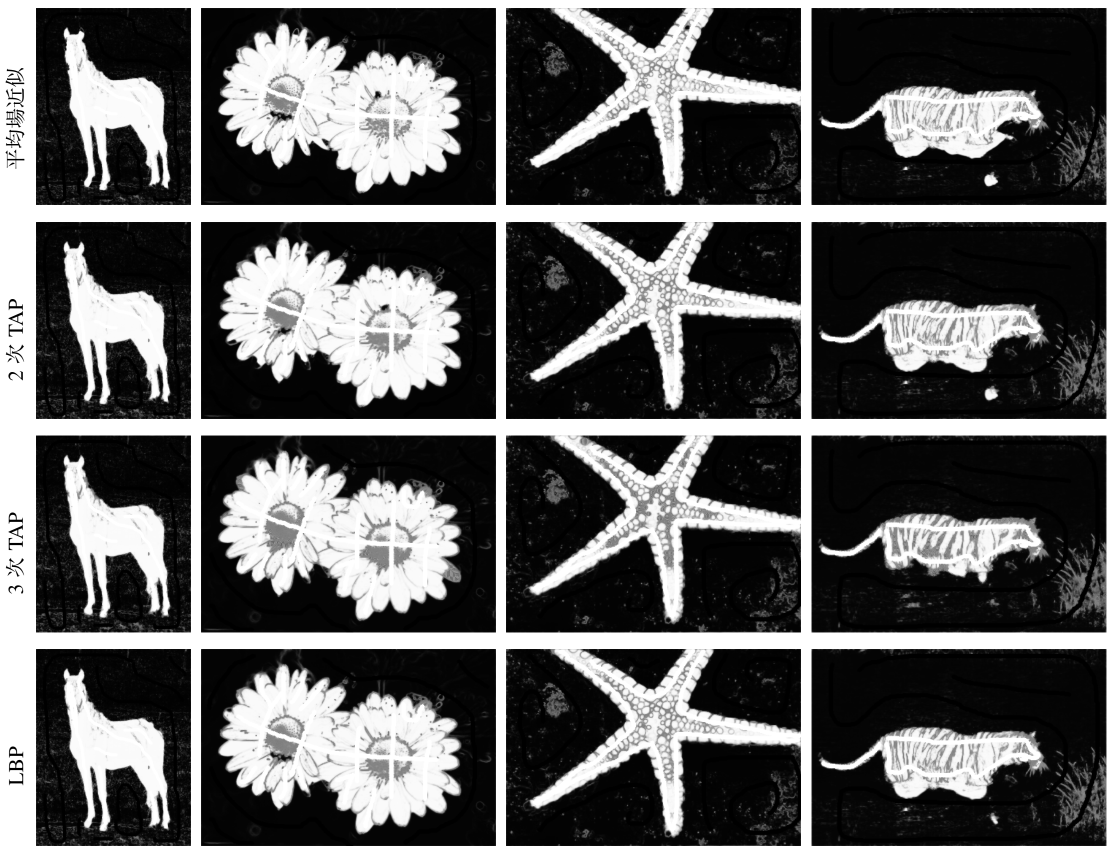


Fig. 3.4: 4 枚の画像に対する，インタラクティブセグメンテーションの結果 (*Horse, Flower, Starfish, Tiger*).

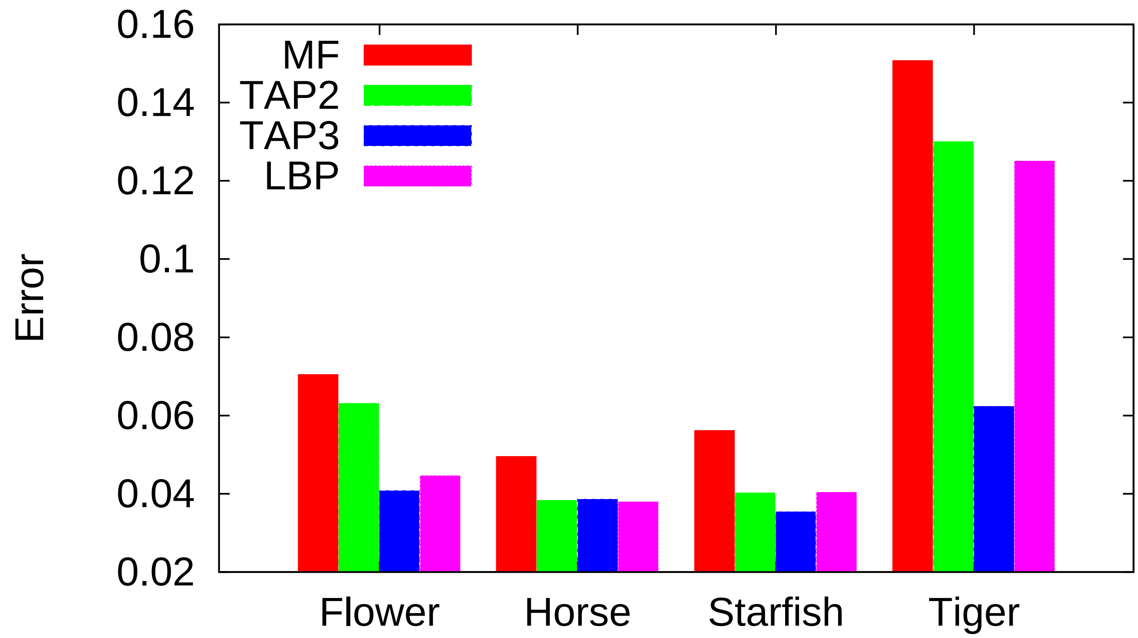


Fig. 3.5: 4 枚の画像を用いた場合の，1 画素辺りの誤差

3.5　本章の結論

本章では，MF-TAP の方法をコンピュータビジョンの問題に適用することを考えた．主に変数が離散値を取る場合の最適化問題を対象に，アルゴリズムおよび詳細な実装方法を示した．そこでは，新しく多値ラベル分類問題のための TAP 方程式および具体的な数値アルゴリズムを導出した．また，MF-TAP と LBP の理論的な比較を行い，MF-TAP には，MRF モデルの選択自由度の高さ，精度の高さおよび計算速度の速さという，3 つの長所があることを述べた．

さらに，導出した MF-TAP のアルゴリズムを，インタラクティブセグメンテーションとステレオマッチングに適用し，その性能を評価する実験を行った．われわれの実験によれば，MF-TAP は精度の面で LBP に少なくとも匹敵し，3 次の TAP 方程式は LBP を上回る．また計算速度の面では，MF-TAP は LBP より理論上，サイトに接続するエッジの数と同じ倍数だけ高速であり，実験でもそれは確かめられた．GPU 上に実装した場合，その差はさらに広がった．

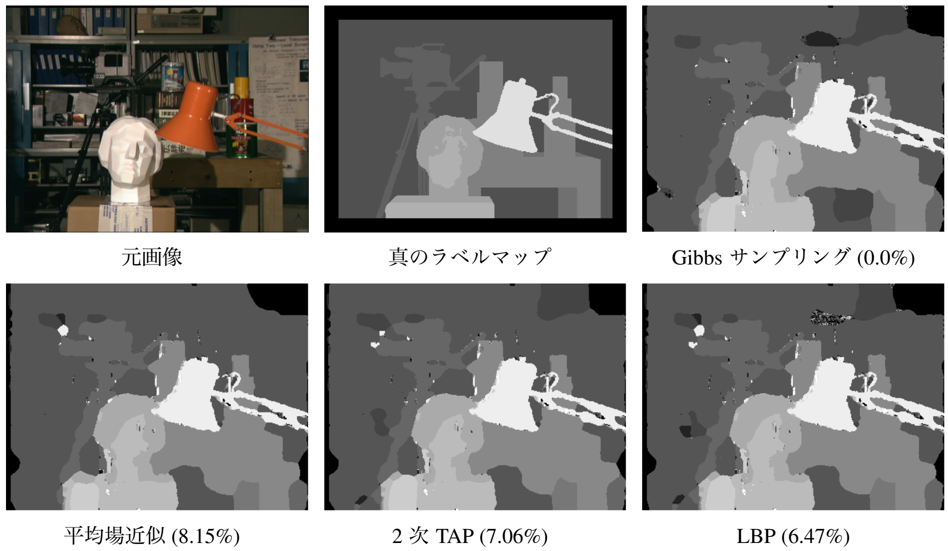


Fig. 3.6: ステレオマッチングの結果．括弧内の数値は，収束後における各手法の誤差を表す．

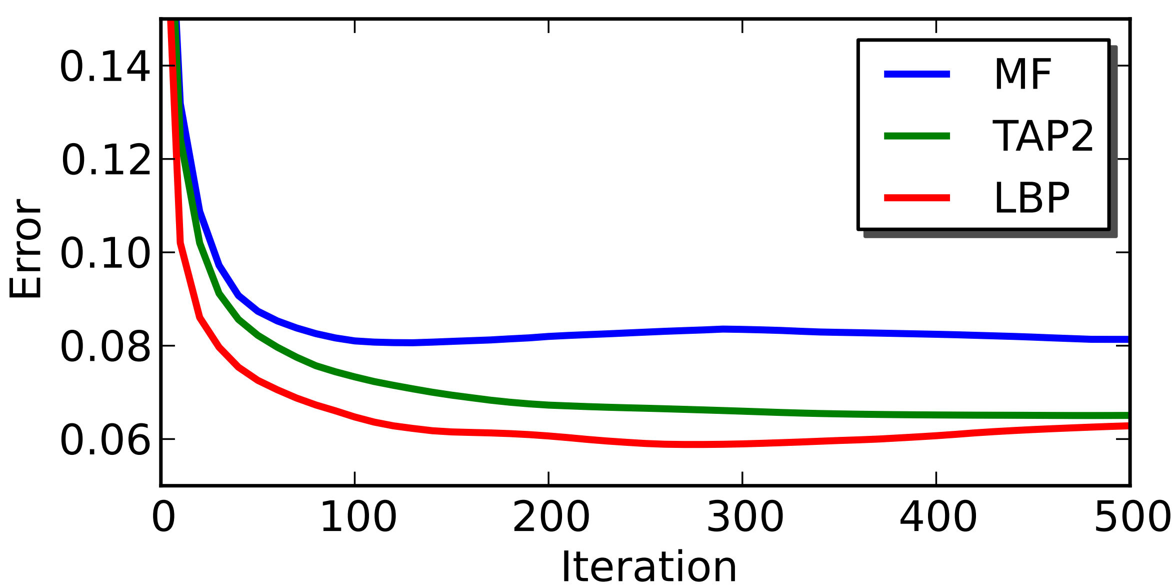


Fig. 3.7: 反復回数毎における，画像 *Tsukuba* を用いたステレオマッチングでの 1 画素辺りの平均誤差．

第4章　非一様に離散化された変数空間上でのマルコフ確率場の推定手法

本章で，われわれは連続的なマルコフ確率場の問題を離散的なマルコフ確率場の問題に離散化するための方法を新しく提案する．連続的なマルコフ確率場がもつ制約により，その問題を解くための最適化アルゴリズムの種類は限られる一方，離散的なマルコフ確率場ではそのような制約がない．そのため，コンピュータビジョンの領域では例え連続的なマルコフ確率場として解くべき問題 (例: ステレオマッチング) に対しても，連続的なエネルギー関数を離散的なエネルギー関数に変換することでマルコフ確率場を離散化し，その上で離散的なマルコフ確率場の最適化問題を解くことでサイト毎の周辺分布を推定していた．一方，提案手法ではエネルギー関数を離散的な関数に変換するのではなく，連続的な関数として取り扱う．その代わり，提案手法では周辺分布それ自体を離散化することで最適化を行う．このような定式化の下で，我々は新しい平均場近似と確率伝搬法のアルゴリズムを新しく導出する．これらのアルゴリズムは変数空間が非一様に離散化した場合でも正しく周辺分布を推定できる．このような非一様な離散化によって，計算速度を保ちつつ精度の高い周辺分布を求めることが可能となる．さらに非一様な離散化を効率的に行うため，われわれは動的離散化という，疎から密に変数空間を離散化するための手法を新しく提案する．最後に複数の実験を通して，われわれの提案手法が従来手法と比較して効果的であることを示す．

4.1　概要

本章では，連続的なマルコフ確率場に対する MPM 推定について焦点を当てる．§2.4, §2.5 で示した通り，周辺分布を推定するためのアルゴリズムは，変数 x_i が連続的であるか離散的であるかによってそれぞれ異なる．変数が連続的である場合，周辺分布の密度関数は少数のパラメータで構成されるパラメトリックな連続分布として表現される．平均場

近似では，周辺分布の更新は (2.20) 式に従って対象のパラメータを更新する．確率伝搬法では，メッセージも同様にパラメトリックな連続分布として扱い，同様に (2.37) 式に従って対象のパラメータを更新する．しかしながら，このような更新は反復方程式によって計算された分布がまた同じ分布族に属している必要があるため，扱える問題の範囲は非常に限られる．確率伝搬法の場合は，ガウシアン分布のような単純な分布しか事実上扱うことができない[1]．平均場近似の場合，この制約はいくらか緩和されるのもの，良い収束性能を持つ反復アルゴリズムを導出することは一般に難しい．

一方，変数が離散的である場合には，そのような制約は存在しない．これは，(2.27) 式もしくは (2.39c) 式に従って計算された分布が，必ず同じ離散的な分布族となることに起因する．つまり，離散的なマルコフ確率場は連続的なマルコフ確率場では解くことができないような，非常に幅広い範囲の問題を扱える．そのため，コンピュータビジョンの領域では離散的な領域で定義された問題 (例: 多値ラベルの画像セグメンテーション) だけでなく，本来連続的な領域で定義された問題に対しても，離散的なマルコフ確率場の問題に変換し，推定値を求めることが一般的であった．具体的には，ステレオマッチングやオプティカルフローの問題ではサイトの変数は視差やフローベクトルといった連続的な値となるが，このような問題に対しても一般には離散化によってエネルギー関数を離散的な関数に変換することで，離散的なマルコフ確率場の問題を推定していた [54, 53, 56, 62]．

本研究では，連続的な MRF の問題を離散的な枠組みで解くための新しい手法を提案する．従来手法では，連続的な (2.7) 式のエネルギー分布を離散化によって離散的なエネルギー分布へ変換し，その上で離散的な平均場近似あるいは確率伝搬法の推定アルゴリズムを用いて，サイト毎の周辺分布を推定していた．一方提案手法では，連続的なエネルギー分布はそのまま連続的なものとして扱い，離散化は行わない．その代わり，提案手法ではサイト毎の周辺分布それ自体を『離散化』することで最適化を行う．厳密には，近似分布 P を離散化し，その下で元の分布 Q へと近づける．この操作によって，連続的な平均場近似や確率伝搬法のアルゴリズムは最終的に，離散的な平均場近似や確率伝搬法に似たアルゴリズムへ変換される．

提案手法では近似分布 P を離散化するために，P を混合矩形分布の積によって表現する．この表現の下では，各々の矩形分布の中心は従来手法におけるエネルギー分布の離散値に相当し，矩形分布の高さは離散的な周辺分布での確率値に相当する．このような定式

[1] このような，マルコフ確率場の結合分布がガウシアン分布に従う確率分布を，一般にガウシアンマルコフ確率場 (Gaussian MRF) と呼ぶ [50, 55].

化の元で，新しい平均場近似と確率伝搬法のアルゴリズムを導出する．この定式化による利点は，(矩形分布同士が変数空間上で重ならない限り) 各サイト毎に任意の位置や大きさの矩形分布関数を配置できることが挙げられる．従来手法においても非等間隔なエネルギー関数の離散化によって周辺分布を推定することもできるが，その手法で得られた周辺分布の形は明らかに本来の連続的な周辺分布の形とは異なる．提案手法では，そのような非等間隔な離散化であっても，本来の連続的な周辺分布に近い形の周辺分布を求められる．さらに，同じ大きさの混合矩形分布を等間隔のグリッド上に配置した場合，提案手法で導出された平均場近似と確率伝搬法のアルゴリズムは従来手法のそれらと一致する．これは後述する補正項の存在によって説明できる．提案手法で求められる離散的な平均場近似と確率伝搬法の反復方程式は，従来手法には無い補正項が一部に含まれている．この補正項は矩形分布関数の非一様性を補正するための役割を果たす．矩形分布関数が等間隔に配置された場合，補正項は互いに打ち消し合うため影響を与えないが，非一様な離散化を行った場合にはこの項による補正が行われる．従来手法での平均場近似と確率伝搬法はこの補正項が含まれていないために，非一様な離散化の際に連続分布とは本質的に異なる形の周辺分布を求めてしまう．

提案手法の利点は以下の 2 点に集約される:

- 提案手法では変数空間を非一様に離散化できる．これによって，例えば一部の重要な区間は密に矩形分布を配置し，そうでない区間は疎に配置することで，計算コストをかけずに推定精度を向上できる．
- 提案手法では変数空間が非ユークリッド空間でグリッドを一様に区切ることが難しい場合でも，正しい周辺分布の推定解を求められる (例えば，双方向反射率分布関数 (BRDF)[42] の推定問題では，変数空間は球面の形状を取る)．

前者の利点は，変数空間の次元が 2 次元以上の高次元である場合に対して特に有効である．可能であれば適当な事前知識を利用して矩形分布を非一様に取ることで，周辺分布の計算を効率的に行うこともできる．

本研究では，これらの提案手法をさらにもう一歩進めた動的な離散化手法についても新しく提案する．この手法では疎から密 (coarse-to-fine) の戦略を採用している．すなわち，初めは変数空間を粗く分割することで高速に計算を行い，大まかな周辺分布の形状を得る．そして混合矩形分布のウェイトが最も高い箇所を分割し，再度計算を行う．この手順

を一定数繰り返すことで，周辺分布が低い箇所は疎に，高い箇所は密にとった混合矩形分布を構成できる．

本章の構成は下記の通りである．§4.2 では変数空間を非一様に離散化することのできる，新しい平均場近似と確率伝搬法のアルゴリズムを導出する．§4.3 では疎から密の方法による，変数空間の動的離散化の手法について説明する．§4.4 では複数の実験を通して，提案手法の有効性を示す．§4.5 は本章の結論である．

4.2 非一様な平均場近似，確率伝搬法の導出

本節では非一様に変数空間を離散化することのできる，新しい平均場近似と確率伝搬法のアルゴリズムを導出する．

4.2.1 新しい平均場近似アルゴリズムの導出

平均場近似は各々のサイトがそれぞれ独立であるという近似分布 P の確率モデルの下で KL 距離を本来の分布 Q に対して最小化させる手法であることを，われわれは既に §2.4 を通して述べた．そして，KL 距離の最小化は自由エネルギー ((2.14) 式) の最小化と等価であり，P の停留点は (2.20) 式の固定点方程式を反復して解くことで求められることも同様に述べた．なお，本章では温度 T ならびに逆温度 $\beta = 1/T$ は議論に入らないため，$T = 1$ と設定する．

ここで，従来手法での離散化の過程について説明する．従来手法では，各々の変数 x_i がとる値を連続的なものではなく，S 個の固定された値をとる，離散的な変数と定義する．すなわち，x_i は $[x^1, \ldots, x^S]$ の離散的な値をとるものとする．そのような表式を用いた場合，周辺分布も同様に S 個の値をとる離散変数となる．ここで，$p_i^s \equiv p_i(x_i = x^s)$ と定義すると，(2.14) 式で定義された自由エネルギーは次のように表現できる．

$$F[\mathbf{p}] = \sum_i \sum_s p_i^s f_i(x^s) + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} p_i^s p_j^t f_{ij}(x^s, x^t) + \sum_i \sum_s p_i^s \ln p_i^s \qquad (4.1)$$

これは多値ラベル分類問題における，平均場近似の自由エネルギーの表現に等しい．ゆえに，$\sum_i p_i^s = 1$ の制約条件の下で (4.1) 式の停留点を求めると，(2.27) 式より固定点方程式

は次式の形をとる．

$$p_i^s \propto \exp\left[-\left(f_i(x^s) + \sum_{j\in\mathcal{N}_i}\sum_{t=1}^{S} f_{ij}(x^s, x^t)p_j^t\right)\right] \tag{4.2}$$

これは周辺分布 $[p_i^1, \ldots, p_i^S](i = 1, \ldots, N)$ の更新則に対応する．

前述のとおり，x_i が連続変数の場合，$p_i(x_i)$ はガウシアン分布のようなパラメトリック分布として表現される．この分布の元では，(2.20) 式はパラメータに関する反復方程式となる．しかしながら，パラメータに関する反復方程式を導くためには，(2.20) 式で導出される分布が同じ分布族に属していなければならない．

ここで，われわれは連続的なマルコフ確率場を離散化するための，新しい定式化について説明する．その定式化の元では，近似分布 $p_i(x_i)$ は次の，S_i 個の混合矩形分布として表現される．

$$p_i(x_i) = \sum_{s=1}^{S_i} \alpha_i^s h_i^s(x_i) \tag{4.3}$$

ここで，h_i^s は位置と大きさが固定された矩形分布であり，α_i^s は混合矩形分布のウェイトを表す．$h_i^s(x_i)$ の具体的な形を以下に示す．まず，$\mathcal{X}$ を変数空間，d を変数空間の次元と定義する．同様に，d 次元の超直方体を $\mathcal{X}_i^s \subset \mathcal{X}$ と定義する．ただし，$\mathcal{X}_i^s$ は任意の $s \neq t$ について次の関係を満たすものとする．

$$\mathcal{X}_i^s \cap \mathcal{X}_i^t = \emptyset \tag{4.4}$$

これは，$\mathcal{X}_i^s$ が変数空間上で互いに交わらない部分集合であることを表している．このような場合，$h_i(x_i)$ は次式の形をとる．

$$h_i^s(x_i) = \begin{cases} 1/\mathcal{V}_i^s & \text{if } x \in \mathcal{X}_i^s \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

ここで，$\mathcal{V}_i^s$ は $\mathcal{X}_i^s$ の超体積を表す．つまり，(4.5) 式は $\int h_i^s(x_i)dx_i = 1$ の関係を満たす．これは $\int p_i(x_i) = 1$ の関係から，α_i^s が次の関係を満たすことも表している．

$$\sum_{s=1}^{S_i} \alpha_i^s = 1 \tag{4.6}$$

(4.3) 式の定義では，$h_i^s(x_i)$ は ($\mathcal{X}_i^s \cap \mathcal{X}_i^t = \emptyset$ の関係を満たす限り) 任意の位置や大きさの矩形分布をとって良い．例えば変数空間 $\mathcal{X}$ の重要性に従って，矩形分布 $[h_i^1(x_i), \ldots, h_i^{S_i}]$ を

非一様に配置することも可能である．また，各サイト毎の矩形分布の形は共通していなくても良い (すなわち，$h_i^s(x) \neq h_j^s(x)$ でも良い)．さらには，各サイト毎に混合矩形分布の数が異なっても良い．これによって，例えば重要なサイトは混合矩形分布の数を増やし，そうでないサイトは数を減らして計算量を削減するといった処理も可能となる．

次に，提案手法ではウェイト α_i^s だけを動かすことで，近似分布 P を Q に近づける．ただし，これは (4.3) 式を (2.20) 式にそのまま代入するだけでは導出できない．なぜなら，(2.20) 式は $p_i(x_i)$ が混合矩形分布でない，制約のない確率分布と仮定した上で導出された，ある点 $p_i(x_i)$ についての自由エネルギーの固定点方程式だからである．ゆえに，α_i^s についての固定点方程式を求めるためには議論を自由エネルギーの時点まで遡る必要がある．まず，(4.3) 式を (2.14) 式に代入することで，自由エネルギーの第一項 ($P(\mathbf{x})$ に関する $E(\mathbf{x})$ の期待値) は次のように計算される．

$$
\begin{aligned}
\langle E\rangle_P &= \sum_i \int p_i(x_i) f_i(x_i) dx_i + \sum_{(i,j)\in\mathcal{E}} \iint p_i(x_i) p_j(x_j) f_{ij}(x_i, x_j) dx_i dx_j \\
&= \sum_i \sum_s \alpha_i^s \int f_i(x_i) h_i^s(x_i) dx_i \\
&\qquad + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} \alpha_i^s \alpha_j^t \iint f_{ij}(x_i, x_j) h_i^s(x_i) h_j^t(x_j) dx_i dx_j \\
&= \sum_i \sum_s \alpha_i^s f_i^s + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} \alpha_i^s \alpha_j^t f_{ij}^{st}
\end{aligned} \tag{4.7}
$$

ここで，f_i^s, f_{ij}^{st} を次のようにおいた．

$$
f_i^s = \int f_i(x_i) h_i^s(x_i) dx_i \tag{4.8a}
$$

$$
f_{ij}^{st} = \iint f_{ij}(x_i, x_j) h_i^s(x_i) h_j^t(x_j) dx_i dx_j \tag{4.8b}
$$

(4.8) 式より，f_i^s, f_{ij}^{st} はそれぞれ矩形分布 $h_i^s(x_i)$, $h_i^s(x_i)h_j^t(x_j)$ における，f_i, f_{ij} についての期待値を表す．同様に，(4.3) 式を用いて，自由エネルギーの第二項 ($P(\mathbf{x})$ のエントロ

ピー) は次のように計算される．

$$
\begin{aligned}
S[P] &= -\sum_i \sum_s \alpha_i^s \int h_i^s(x_i) \ln \left(\sum_{s'} \alpha_i^{s'} h_i^{s'}(x_i) \right) dx_i \\
&= -\sum_i \sum_s \alpha_i^s \int h_i^s(x_i) \ln \left(\alpha_i^s h_i^s(x_i) \right) dx_i \\
&= -\sum_i \sum_s \alpha_i^s \ln \alpha_i^s - \sum_i \sum_s \alpha_i^s \int h_i^s(x_i) \ln h_i^s(x_i) dx_i \\
&= -\sum_i \sum_s \alpha_i^s \ln \alpha_i^s + \sum_i \sum_s \alpha_i^s B_i^s
\end{aligned} \tag{4.9}
$$

ここで，B_i^s は矩形分布 h_i^s のエントロピーであり，次式で定義される．

$$
B_i^s \equiv -\int h_i^s(x_i) \ln h_i^s(x_i) dx_i = \ln \mathcal{V}_i^s \tag{4.10}
$$

本来，混合分布の Shannon エントロピーを解析的に求めることは困難であるが，今回の場合は (4.4) 式による制約から $\mathcal{X}_i^s$ は他の $\mathcal{X}_i^t$ に干渉しないため，混合分布のエントロピーは単純なエントロピーの和として分解できる．ゆえに，(4.7) 式, (4.9) 式から，最終的な自由エネルギー ((2.14) 式) は次のように表現できる．

$$
F_{\mathrm{MF}}[\alpha] = \sum_i \sum_s \alpha_i^s (f_i^s - B_i^s) + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} \alpha_i^s \alpha_j^t f_{ij}^{st} + \sum_i \sum_s \alpha_i^s \ln \alpha_i^s \tag{4.11}
$$

提案手法では (4.6) 式の制約条件の下で (4.11) 式の α_i^s に関する固定点方程式を導出する．これは従来手法における，$\sum_s p_i^s = 1$ の制約条件の下での (4.1) 式の p_i^s に関する固定点方程式と類似している．すなわち，提案手法と従来手法では $\alpha_i^s \leftrightarrow p_i^s$, $(f_i^s - B_i^s) \leftrightarrow f_i(x^s)$ ($f_i^s \leftrightarrow f_i(x^s)$ でない), $f_{ij}^{st} \leftrightarrow f_{ij}(x^s, x^t)$ という類似性があり，従来手法の周辺分布，データ項，平滑化項を上のルールに従って置換することで，提案手法と全く同じ式と制約条件を得られる．ゆえに，(4.2) 式より α_i^s についての固定点方程式は次式となる．

$$
\alpha_i^s \propto \exp\left[-\left((f_i^s - B_i^s) + \sum_{j\in\mathcal{N}_i} \sum_t f_{ij}^{st} \alpha_j^t \right) \right] \tag{4.12}
$$

なお §2.4.3 では，サイト i はすべて同一のラベル数を取るものと仮定していたが ($S_i = S$), S_i がそれぞれ異なる場合においても導出の流れは変わらないため，ここでは省略する．

提案手法における (4.12) 式の反復方程式は，従来手法における (4.2) 式の反復方程式と類似している．基本的には $\alpha_i^s \leftrightarrow p_i^s$, $f_i^s \leftrightarrow f_i(x^s)$, $f_{ij}^{st} \leftrightarrow f_{ij}(x^s, x^t)$ のような類似性が存在するが，唯一の違いは B_i^s の有無である．もしサイト i の混合矩形分布 $[h_i^1(x_i), \ldots, h_i^{S_i}(x_i)]$

の大きさが同一であれば，B_i^s はすべての s について同一の値をとるため，(4.12) 式より提案手法の反復式は B_i^s による影響を受けない．すなわち，混合矩形分布の大きさがすべて同じであれば，従来手法と提案手法の反復式は一致する[2]．以上から，B_i^s は変数空間 $\mathcal{X}$ の離散化の『非一様性』を補正するための，補正項としての役割をもつことが分かる．

4.2.2　新しい確率伝搬法アルゴリズムの導出

平均場近似と同様に，確率伝搬法に関しても新しい非一様に離散化された場合の反復方程式を導出する．確率伝搬法は近似分布 P が (2.28) 式という確率モデルの形を取るとした上で，元の分布 Q に近づけるアルゴリズムであることを，われわれは §2.5 を通して既に述べた．

ここで，従来手法での離散化の過程について述べる．平均場近似と同様に，従来手法では各々の変数 x_i がとる値を連続的なものではなく，S 個の固定された値をとる，離散的な変数と定義する．この定義の下では，確率伝搬法のベーテ自由エネルギー ((2.30) 式) は次の離散的な形を取る．

$$F[\mathbf{p}] = \sum_i \sum_s p_i^s f_i(x^s) + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} p_{ij}^{st} f_{ij}(x^s, x^t) \\ - \sum_i (z_i - 1) \sum_s p_i^s \ln p_i^s + \sum_{(i,j)\in\mathcal{E}} p_{ij}^{st} \ln p_{ij}^{st} \quad (4.13)$$

ここで，$\mathbf{p}$ はすべての p_i^s と p_{ij}^{st} を含むベクトルである．また，p_i はサイト i の周辺分布，p_{ij} はサイト i と j の結合確率分布を表すため，p_i および p_{ij} の制約条件は次の 3 つの関係を満たす．

$$\sum_s p_i^s = 1 \quad (4.14a)$$

$$\sum_{s,t} p_{ij}^{st} = 1 \quad (4.14b)$$

$$\sum_s p_{ij}^{st} = p_j^t \quad (4.14c)$$

[2]厳密には，$f_i(x^s)$ は x^s における f_i の値を表す一方で，f_i^s は部分空間 $\mathcal{X}_i^s$ における f_i の期待値を表すため，両者の値は一致しない．しかし，離散的な周辺分布としてはほぼ同じ形の分布を取る．

これらの制約条件の下で $F[\mathbf{p}]$ の停留点を求めることによって，次の離散的なメッセージの反復方程式を得る．

$$m_{ij}^t \leftarrow \sum_s \phi_i^s \psi_{ij}^{st} \prod_{k \in \mathcal{N}_i \setminus j} m_{ki}^s \tag{4.15}$$

ここで，ϕ_i^s, ψ_{ij}^{st} は次式で定義される．

$$\phi_i^s = \exp[-f_i(x^s)] \tag{4.16a}$$

$$\psi_{ij}^{st} = \exp[-f_{ij}(x^s, x^t)] \tag{4.16b}$$

次に，提案手法では $p_i(x_i)$ と $p_{ij}(x_i, x_j)$ を，次の混合矩形分布を用いて定義する．

$$p_i(x_i) = \sum_{s=1}^{S_i} \alpha_i^s h_i^s(x_i) \tag{4.17a}$$

$$p_{ij}(x_i, x_j) = \sum_{s=1}^{S_i} \sum_{t=1}^{S_j} \alpha_{ij}^{st} h_i^s(x_i) h_j^t(x_j) \tag{4.17b}$$

(4.17) 式を $\int p_i(x_i)dx_i = 1$, $\iint p_{ij}(x_i, x_j)dx_i dx_j = 1$, $\int p_{ij}(x_i, x_j)dx_i = p_j(x_j)$ の関係式にそれぞれ代入することで，次の 3 つの関係式を得る．

$$\sum_s \alpha_i^s = 1 \tag{4.18a}$$

$$\sum_{s,t} \alpha_{ij}^{st} = 1 \tag{4.18b}$$

$$\sum_s \alpha_{ij}^{st} = \alpha_j^t \tag{4.18c}$$

これらの関係式は従来手法における，(4.14) 式 の関係と一致する．

次に，平均場近似と同様に，(4.17) 式を用いてベーテ自由エネルギーを書き直す．(4.17) 式を (2.30) 式に代入することで，ベーテ自由エネルギーの第一項 (分布 P における $E(\mathbf{x})$ の期待値) は次のように計算できる．

$$\begin{aligned}\langle E \rangle_P &= \sum_i \sum_s \alpha_i^s \int f_i(x_i) h_i^s(x_i) dx_i \\ &\qquad + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} \alpha_{ij}^{st} \iint f_{ij}(x_i, x_j) h_i^s(x_i) h_j^t(x_j) dx_i dx_j \\ &= \sum_i \sum_s \alpha_i^s f_i^s + \sum_{(i,j)\in\mathcal{E}} \sum_{s,t} \alpha_{ij}^{st} f_{ij}^{st}\end{aligned} \tag{4.19}$$

同様の手順で，ベーテ自由エネルギーの第二項 (分布 P のエントロピー) を計算する．まず，(4.9) 式の議論から，p_i の負のエントロピーは次のように計算できる．

$$\int p_i(x_i)\ln p_i(x_i)dx_i = \sum_s \alpha_i^s \ln \alpha_i^s - \sum_s \alpha_i^s B_i^s \tag{4.20}$$

p_{ij} についても同様の計算を行う．本来，混合分布の Shannon エントロピーを解析的に求めることは困難であるが，$\mathcal{X}_i^t \cap \mathcal{X}_i^t = \emptyset$ の制約から，p_{ij} のエントロピーは次の簡単な形に書き直せる．

$$\begin{aligned}
&p_{ij}(x_i,x_j)\ln p_{ij}(x_i,x_j)dx_i dx_j \\
&\quad= \sum_{s,t}\alpha_{ij}^{st}\iint h_i^s(x_i)h_j^t(x_j)\ln\left(\sum_{s',t'}\alpha_{ij}^{s't'}h_i^{s'}(x_i)h_j^{t'}(x_j)\right)dx_i dx_j \\
&\quad= \sum_{s,t}\alpha_{ij}^{st}\iint h_i^s(x_i)h_j^t(x_j)\ln\left(\alpha_{ij}^{st}h_i^s(x_i)h_j^t(x_j)\right)dx_i dx_j \\
&\quad= \sum_{s,t}\alpha_{ij}^{st}\ln\alpha_{ij}^{st} - \sum_{s,t}\alpha_{ij}^{st}(B_i^s + B_j^t)
\end{aligned} \tag{4.21}$$

(4.20) 式, (4.21) 式から，ベーテ自由エネルギーの第二項は次式の形をとる．

$$\begin{aligned}
S[\alpha] = \sum_i (z_i - 1)\sum_s \alpha_i^s \ln\alpha_i^s - \sum_i (z_i-1)\sum_s \alpha_i^s B_i^s \\
- \sum_{(i,j)\in\mathcal{E}}\sum_{s,t}\alpha_{st}^{ij}\ln\alpha_{st}^{ij} + \sum_{(i,j)\in\mathcal{E}}\sum_{s,t}\alpha_{ij}^{st}(B_i^s + B_j^t)
\end{aligned} \tag{4.22}$$

(4.19) 式, (4.22) 式から，ベーテ自由エネルギーは最終的に次式となる．

$$\begin{aligned}
F_{\mathrm{BP}}[\alpha] = \sum_i\sum_s \alpha_i^s(f_i^s + (z_i-1)B_i^s) + \sum_{(i,j)\in\mathcal{E}}\sum_{s,t}\alpha_{ij}^{st}(f_{ij}^{st} - B_i^s - B_j^t) \\
- \sum_i (z_i-1)\sum_s \alpha_i^s\ln\alpha_i^s + \sum_{(i,j)\in\mathcal{E}}\sum_{s,t}\alpha_{ij}^{st}\ln\alpha_{ij}^{st}
\end{aligned} \tag{4.23}$$

提案手法では (4.18) 式の制約条件の下で (4.23) 式の停留点を求めるための反復方程式を導出する．これは従来手法における，(4.14) 式の制約条件の下での (4.13) 式の自由エネルギーの固定点方程式と類似する．すなわち，提案手法と従来手法では $\alpha_i^s \leftrightarrow p_i^s, \alpha_{ij}^{st} \leftrightarrow p_{ij}^{st}$, $f_i^s + (z_i-1)B_i^s \leftrightarrow f_i(x^s), f_{ij}^{st} - B_i^s - B_j^t \leftrightarrow f_{ij}(x^s, x^t)$ という類似性があり，従来手法の分布，データ項，平滑化項を上のルールに従って置換することで，提案手法と全く同じ式

と制約条件を求められる．ゆえに，提案手法における離散的なメッセージの反復方程式は (4.15) 式より次式となる．

$$m_{ij}^t \propto \sum_s \phi_i^s \psi_{ij}^{st} \prod_{k\in\mathcal{N}_i \setminus j} m_{ki}^s$$

ここで，ϕ_i^s と ψ_{ij}^{st} はそれぞれ次式で定義される．

$$\phi_i^s = \exp\left[-(f_i^s + (z_i - 1)B_i^s)\right] \tag{4.24a}$$

$$\psi_{ij}^{st} = \exp\left[-(f_{ij}^{st} - B_i^s - B_j^t)\right] \tag{4.24b}$$

最後に，サイト i の周辺分布ならびにサイト i, j の周辺結合確率はそれぞれ次の形で表現される．

$$\alpha_{ij}^{st} \propto \psi_{ij}^{st} \phi_i^s \phi_j^t \prod_{k\in\mathcal{N}_i \setminus j} m_{ki}^s \prod_{l\in\mathcal{N}_j \setminus i} m_{lj}^t \tag{4.25a}$$

$$\alpha_i^s \propto \phi_i^s \prod_{k\in\mathcal{N}_i} m_{ki}^s \tag{4.25b}$$

もしサイト i の混合矩形分布 $[h_i^1(x_i), \ldots, h_i^{S_i}(x_i)]$ の大きさが同一であれば，B_i^s はすべての s について同一の値をとるため，(4.25) 式ならびに (4.24) 式より提案手法の反復式は平均場近似と同様に B_i^s による影響を受けない．ゆえに，B_i^s は離散化の非一様性を補正するための，補正項の役割をはたす．

4.3　変数空間上での動的離散化の提案

4.3.1　非一様な離散化の有用性

新しい平均場近似と確率伝搬法の手法は，非一様に変数空間を離散化した場合でも，元の連続的なエネルギー分布に近い周辺分布を正しく求められる．これによって，変数空間の重要箇所は密に混合矩形分布を配置し，そうでない箇所は疎に配置することで，周辺分布の推定精度を大きく低下させることなく計算速度を向上できる．変数空間が 2 次元以上の高次元空間である場合，この効果はより劇的なものとなる．なぜなら，平均場近似と確率伝搬法の双方において，反復計算のコストは主にラベル数によって決定するからである (変数空間が D 次元で，1 次元あたりのラベル数が L である場合，計算コストは双方とも L^{2D} に比例する)．

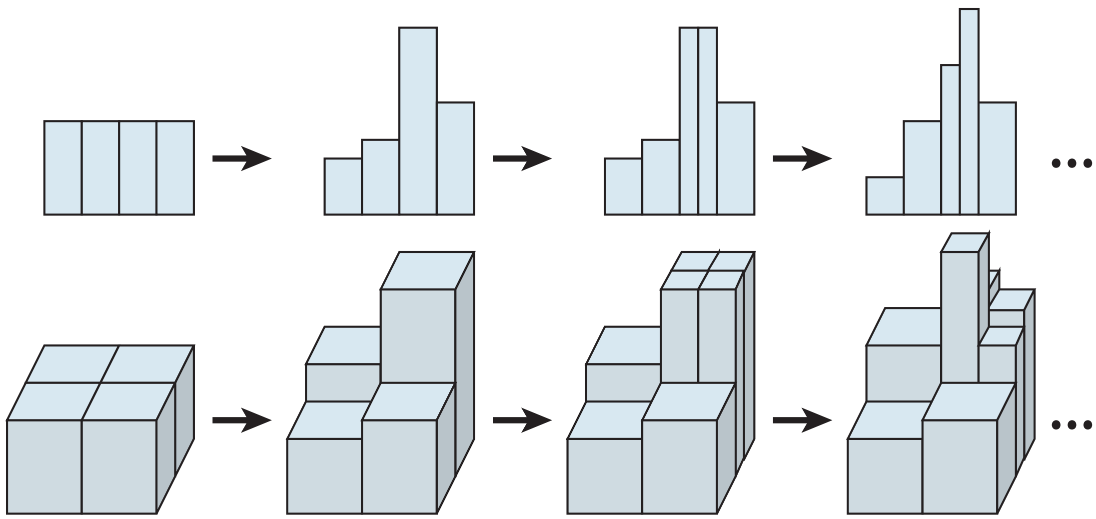


Fig. 4.1: 変数間の動的離散化．矩形の集合は真の周辺分布を近似する混合周辺分布を表す．動的離散化の手法では，最も大きいウェイトを持つ矩形分布は複数の混合矩形分布によって分割される．

しかしながら，この議論だけではどのように変数空間を離散化すれば効率的な離散化が行えるだろうという疑問に対しては，未だに明確な回答が得られていない．もし対象のマルコフ確率場の問題に対して適当な事前知識 (例: 全体の変数空間の中で，この領域であれば恐らく周辺分布のピークが表れるだろうという事前知識) を持っているのであれば，その区間に対し集中的に矩形分布を配置することで良好な離散化を行える．一方，ほとんどの問題ではそのような事前知識が得られない．本節ではそのような場合でも，後述する変数空間の動的離散化によって効果的な離散化を行う手法を新しく提案する．

4.3.2　疎から密へのブロック分割

MPM 推定の精度をできる限り高めるのであれば，周辺分布が高くない箇所はさほど重要でない一方，周辺分布のピーク位置付近は精度の良い分布を求める必要がある．これは，周辺分布の最大値付近を細かく離散化することで達成できる．

前もって周辺分布の密度が最大となる位置を知ることは一般的に不可能であるため，新しい提案手法では動的に変数空間を離散化することで，ピーク位置を推定する．Fig. 4.1 に動的離散化の概要を示す．動的離散化でははじめに，少数の混合矩形分布を用いて変数空間を疎に離散化する．この混合矩形分布は真の周辺分布を近似するための近似分布であ

るため，非一様な平均場近似もしくは確率伝搬法を複数回反復させることで，(この混合矩形分布の表現の下での) 真の周辺分布に近い分布を求められる．次に，各々のサイト i に対し，混合矩形分布のウェイト $[\alpha_i^1, \ldots, \alpha_i^{S_i}]$ の中から最も高い値の α_i^s を選ぶ．そして，この α_i^s を構成している矩形分布を複数の混合矩形分布に分割する．最後に，再度非一様な平均場近似もしくは確率伝搬法を適用することで，分割後の混合矩形分布をさらに真の分布へと近づける．これら 3 つの手順を一定数繰り返すことで，ピーク位置を細かく離散化した混合矩形分布を構成できる．

4.3.3　矩形分布の分割

Fig. 4.1 のように，適当な混合矩形分布を複数の混合矩形分布に分割することで混合矩形分布の構成は変化する．すなわち，分割時には平均場近似における α_i^s，確率伝搬法における m_{ij}^t をそれぞれ更新する必要がある．これらのパラメータの更新には，分割する前と後で混合矩形分布の形を変化させないように更新することが望ましい．この原理を基にすると，パラメータの分割手法は平均場近似と確率伝搬法でそれぞれ異なる．

平均場近似における具体的なパラメータの更新手順は下記の通りである．最もウェイトが大きい矩形分布を，K 個の矩形分布で分割する場合を考える．上の原理を基にすると，新しい混合矩形分布のウェイトは元のウェイトを K 分割することで与えられる．これによって，分割の前と後で対象の分布の形が変化しないことが保証される．例として，サイト i における s 番目の矩形分布を，2 つの矩形分布に分割する場合を考える．この分割手法によれば，分割前に $[\alpha_i^1, \ldots, \alpha_i^{s-1}, \alpha_i^s, \alpha_i^{s+1}, \ldots, \alpha_i^{S_i}]$ というウェイトを表していた混合矩形分布は，分割によって $[\alpha_i^1, \ldots, \alpha_i^{s-1}, \alpha_i^2/2, \alpha_i^s/2, \alpha_i^{s+1}, \ldots, \alpha_i^{S_i}]$ という，$S_i + 1$ 個の混合矩形分布で表現される．

確率伝搬法における具体的なパラメータの更新手順は下記の通りである．サイト i に対しての混合矩形分布の分割を考える．周辺分布のウェイトを求めるため，まず i 近傍のメッセージ $\{m_{ki}^s | k \in \mathcal{N}_i\}$ を利用して，(4.25) 式に従って周辺分布のウェイト α_i^s を求める．次に，これらのウェイトの中で最も大きいウェイト α_i^s を選び，上の手順に従って変数空間を分割する．平均場近似の場合には，混合矩形分布のウェイト $[\alpha_i^1, \ldots, \alpha_i^{S_i}]$ を分割することで変数空間の分割を行うことができた．一方，確率伝搬法の場合には，k 近傍のメッセージ $[m_{ki}^1, \ldots, m_{ki}^{S_i}]$ に対し，平均場近似と同様の方法で分割することで変数空間の分割を行う．

4.4 実験

本節では 2 種類の実験を通して，提案手法の有効性を確認した．

4.4.1 非一様な離散化に対する影響の測定

変数空間を非一様に離散化した場合での従来手法と提案手法のふるまいを調べるため，われわれは周辺分布の厳密解を求められる単純なガウシアンマルコフ確率場の問題を用いて実験を行った [50]．具体的には，われわれは次のエネルギー分布に従う，5×5 の 4 近傍 MRF モデルを用いて実験を行った．

$$E(\mathbf{x}) = \sum_i x_i^2 + \sum_{(i,j)\in\mathcal{E}} (x_i - x_j)^2 \tag{4.26}$$

明らかに，(4.26) 式におけるすべてのサイトの周辺分布はガウシアン分布となり，その平均は 0 である．

この MRF モデルを用いて，われわれは変数空間を $x = 0$ に関して非対称的に離散化した．以下に具体的な離散化の手順を示す．まず，離散化の際にわれわれは $x \in [-2, 2]$ の変数空間だけを考慮した．そして，負の領域 ($[-2, 0]$) は 64 個の区間に分けて離散化する一方，正の領域 ($[0, 2]$) は 16 個の区間に分け，離散化を行った．すなわち，本来連続的な MRF の問題を，合計 80 個の離散ラベルに分割した．

さらに，われわれは従来手法での平均場近似，確率伝搬法のアルゴリズムと提案手法での平均場近似，確率伝搬法のアルゴリズムを上の離散的な MRF に対して適用した．Fig. 4.2 と Fig. 4.3 にこれらの結果を示す．これらの図はすべて，5×5 の MRF モデルにおける左上のサイトの周辺分布を表している．また，従来手法での推定値は赤い点で表現する一方，提案手法での推定値は青いヒストグラムで表現する．そして，周辺分布の厳密解は連続的な赤のカーブで表現する．なお，従来手法における周辺分布は離散分布 ($[p_i^1, \ldots, p_i^S]$) として表現される．ゆえに，Fig. 4.2，Fig. 4.3 では連続的な領域での密度と比較するために，従来手法の周辺分布を適当にスケール倍することで周辺分布の高さを調節している．また，提案手法では周辺分布は混合矩形分布として表現されるため，Fig. 4.2，Fig. 4.3 では同様の表現を用いて表現している．

Fig. 4.2，Fig. 4.3 を見ても分かる通り，従来手法における平均場近似と確率伝搬法は両者ともバイアスが含まれている．これは，従来手法での MPM 推定解が真の推定解 ($x = 0$)

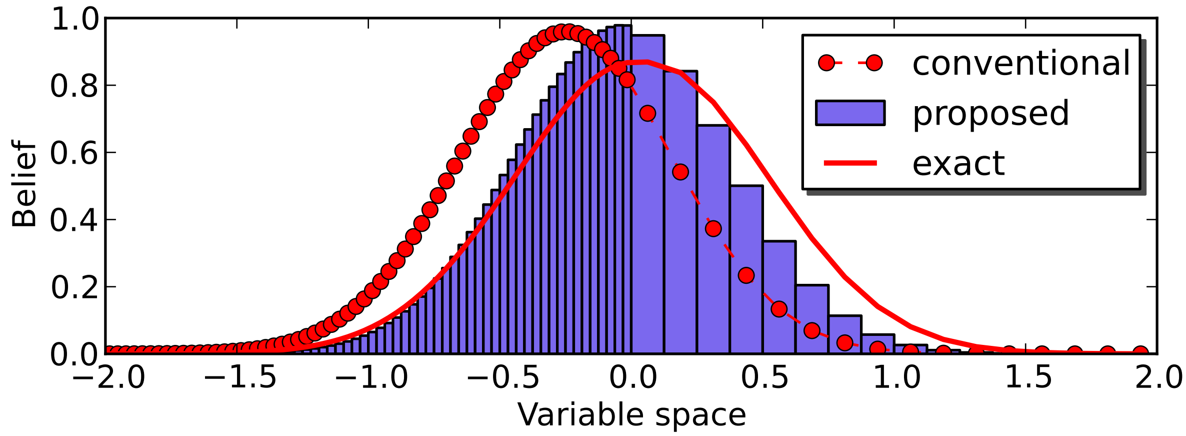


Fig. 4.2: 非一様な離散化を行った場合の，従来手法と提案手法の平均場近似アルゴリズムの結果．赤い点は従来手法での平均場近似で推定した周辺分布を表す．青いヒストグラムは提案手法での平均場近似で推定した周辺分布の密度関数を表す．連続的な赤の曲線は真の周辺分布を表す．

から負の方向へそれていることを示している．これは非一様な離散化の影響による．MRFにおいて，対象の状態のエネルギーが低ければ低いほど，その結合確率密度の値は指数的に上昇する．すなわち，密に離散化した負の領域は多くの低エネルギー状態を含むことになるため，結果として周辺分布はエネルギーが低い負の領域へそれてしまう．一方，提案手法における平均場近似と確率伝搬法は両者とも，従来手法より正確な推定がなされている．これは補正項である B_i^s の影響による．平均場近似の結果は少々バイアスが含まれているものの，その精度は従来手法よりもはるかに良い．確率伝搬法では，その平均場近似よりもさらに良好な推定結果が得られている．なお，Fig. 4.2，Fig. 4.3 で推定された分散の推定値は双方とも厳密解とは異なる値を示しているが，これはガウシアンマルコフ確率場における平均場近似，確率伝搬法の理論的な限界によるものであり，非一様な離散化によるものではない．たとえ変数空間の離散化が対称的であったとしても，平均場近似，確率伝搬法は両者とも周辺分布の分散の厳密な値を推定できない．

4.4.2　ステレオマッチング

われわれはステレオマッチングの問題を提案手法である動的離散化のアルゴリズムに適用することで，その有効性を確認した．エネルギー関数の生成に，今回は Middlebury MRF

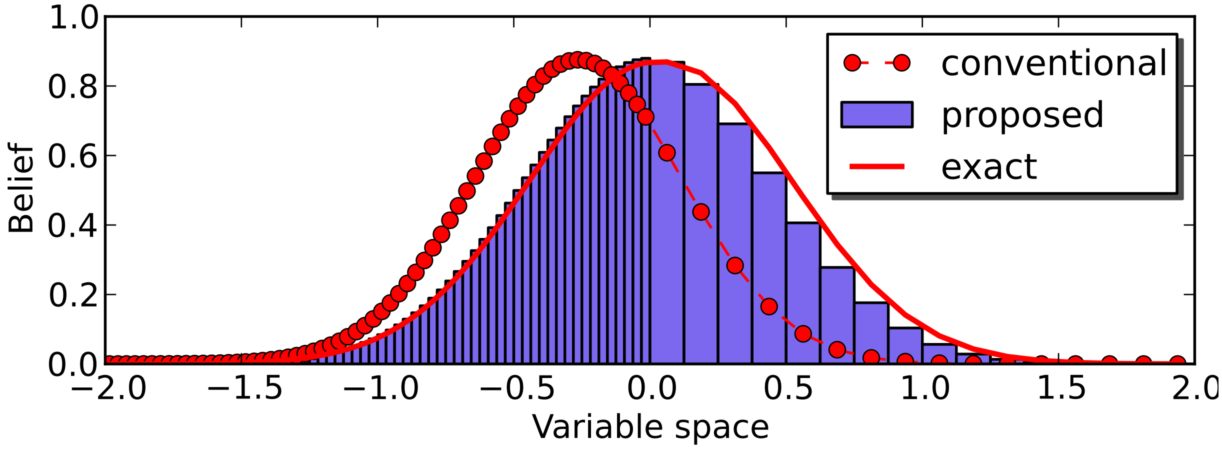


Fig. 4.3: 非一様な離散化を行った場合の，従来手法と提案手法の確率伝搬法アルゴリズムの結果．凡例はすべて Fig. 4.2 と同じである．

のライブラリ [54] を利用した．エネルギー関数を生成するためのパラメータは $|L| = 128$, $\lambda = 2$, $\texttt{smoothmax} = 20$, $\texttt{truncated} = 2$ とした．また，このようにして生成したエネルギー関数をそのまま利用した場合，確率分布は非常に鋭いピーク関数の形を取るため，今回の実験の目的には不適であった．以上から，われわれはエネルギー分布関数に $1/10$ 倍した関数を使用することで，周辺分布の形がよりゆるやかになるよう調節した．

動的離散化の実験は以下の手順で行った．はじめに，8 個の混合矩形分布を用いて，変数空間を疎に離散化した．その際の矩形分布の大きさは，すべて同じとした．次に，非一様な変数空間での平均場近似，確率伝搬法のアルゴリズムを適用した．その際の反復回数はそれぞれ 100 回とした．次に，最も大きいウェイトの矩形分布を抽出し，その矩形分布を 2 つの矩形分布に分割した．最後に，これらの手順を 8 回繰り返した．なお，実装上の理由から，最小の矩形分布の幅は 1 と設定した．つまり，もし最も大きなウェイトである矩形分布の幅が 1 であれば，矩形分布の分割は二番目に大きい矩形分布に対して行う．もしこの矩形分布の幅も 1 であれば，矩形分布の分割は三番目に大きい矩形分布に対して行う．

上の再帰的な矩形分布の分割によって，初めに各サイト毎に 8 個であった矩形分布の数は，最終的に 16 個となる．Fig. 4.4 と Fig. 4.5 にデータセット *Aloe* (641×555 px) を用いた場合の，周辺分布の遷移を示す．なお，比較のために，従来手法での平均場近似，確率伝搬法の結果についても同様に Fig. 4.4，Fig. 4.5 に示す．この場合の離散化は等間隔に 16 分割するものとし，反復回数は両者とも 1000 回とした．

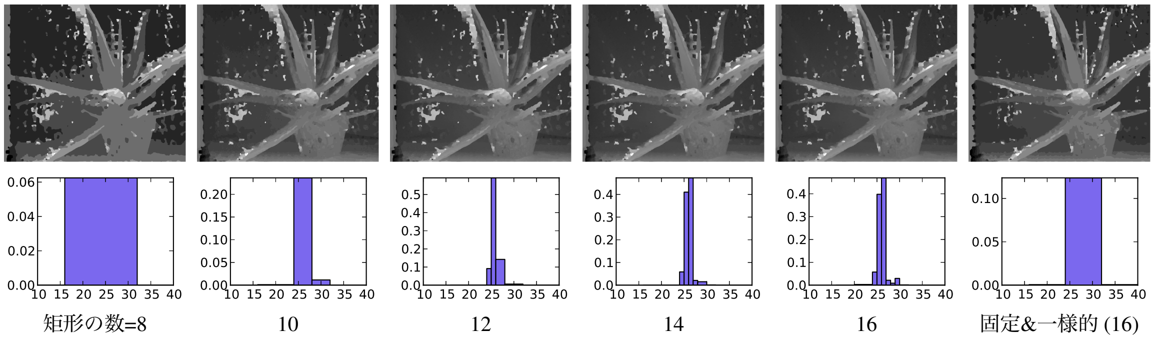


Fig. 4.4: 動的離散化を用いた平均場近似アルゴリズムの *Aloe* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

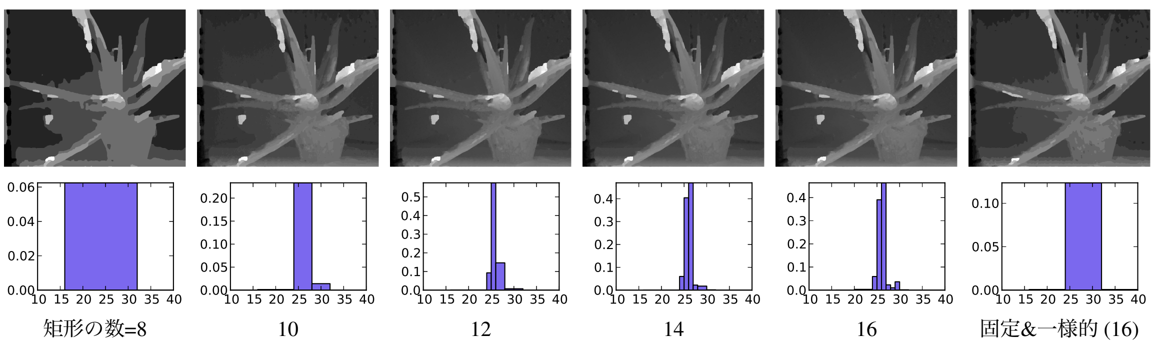


Fig. 4.5: 動的離散化を用いた確率伝搬法アルゴリズムの *Aloe* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

両者の結果とも，矩形分布の増加に従って，周辺分布の精度が次第に改善されていくことが分かる．なお，下段の水平軸は変数区間の範囲 ($[0, 128]$) を表す．すなわち，変数空間を 8 分割，16 分割した場合の矩形分布の幅は，それぞれ $128/8 = 16, 128/16 = 8$ となる．従来手法の，矩形の幅が固定された状態の周辺分布の結果と比較すると，動的離散化による周辺分布の推定精度は非常に良好であることが分かる．これは，動的離散化の手法で推定した結果の視差が，ラベル数の増加に従ってより滑らかに繋がっていることからも明らかである．さらに，Fig. 4.4，Fig. 4.5 は提案手法の矩形分布の数が同じ場合 ($S = 16$) だけでなく，より少ない場合においても同様の結果を示している．これは，従来手法の離散化ではラベル数の不足によって，不十分な視差の結果が得られていることに起因する．

その他のデータセットに対しても同様の実験を行った．選択した 3 枚の画像 (*Cloth1*,

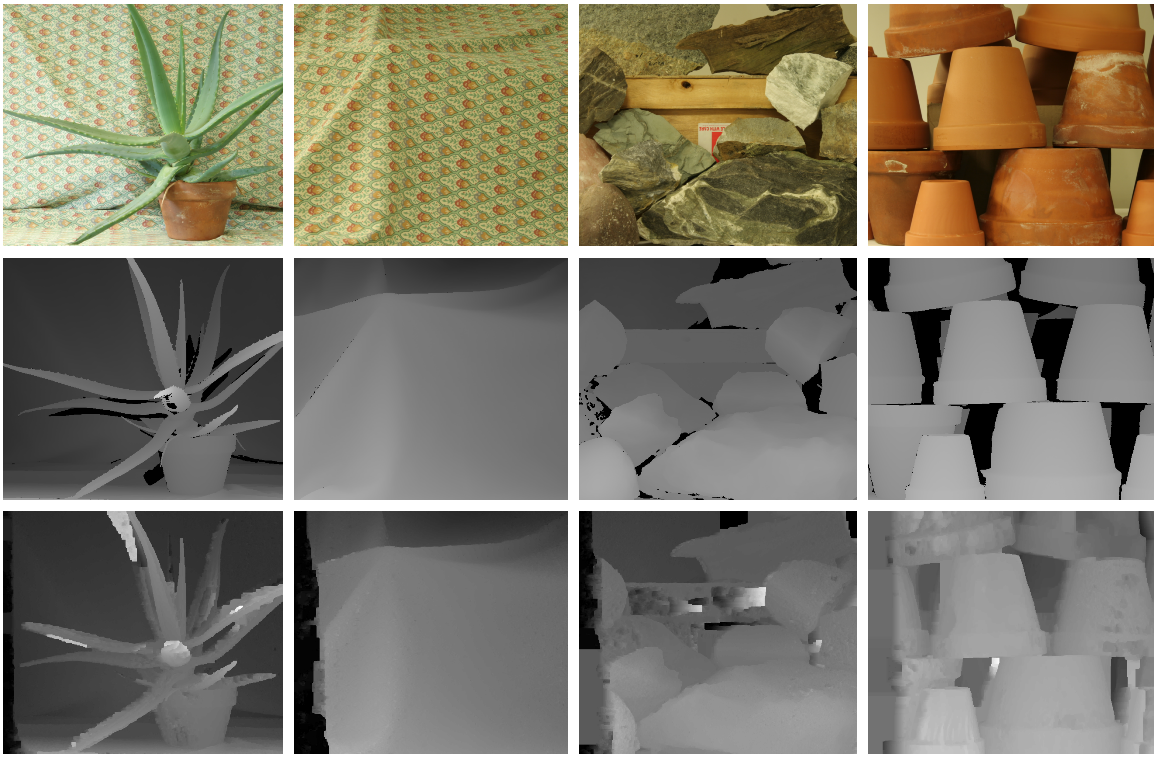


Fig. 4.6: 実験で用いた 4 つのデータセット: *Aloe*, *Cloth1*, *Rocks1*, *Flowerpots*．上段: 入力画像の左側の画像．中段: 真の画像．下段: α-拡張アルゴリズムによって推定された視差画像．

Rocks1, *Fowerpots*) は *Aloe* と同様に，すべて Middlebury MRF のライブラリ [54] から取得した．なお，パラメータはすべて *Aloe* と同様の値を使用した．Fig. 4.6 に (*Aloe* を含む) 入力画像と真の視差画像の値を示す．また，MAP 推定で得られた結果と平均場近似，確率伝搬法の結果を比較する目的で，α-拡張アルゴリズム [26] で得られた視差画像についても Fig. 4.6 に示す．

すべての結果を図 4.7-4.12 に示す．これらの図はすべて *Aloe* と同様の傾向を示した．具体的には，すべてのデータセットとも矩形分布の増加に従って，周辺分布の精度が次第に改善されている．また，提案手法の結果は従来手法と比較して，少ないラベル数でもより良好な視差画像が得られている．これは同時に，周辺分布の推定が従来手法よりもよりよい精度で求められていることを意味する．

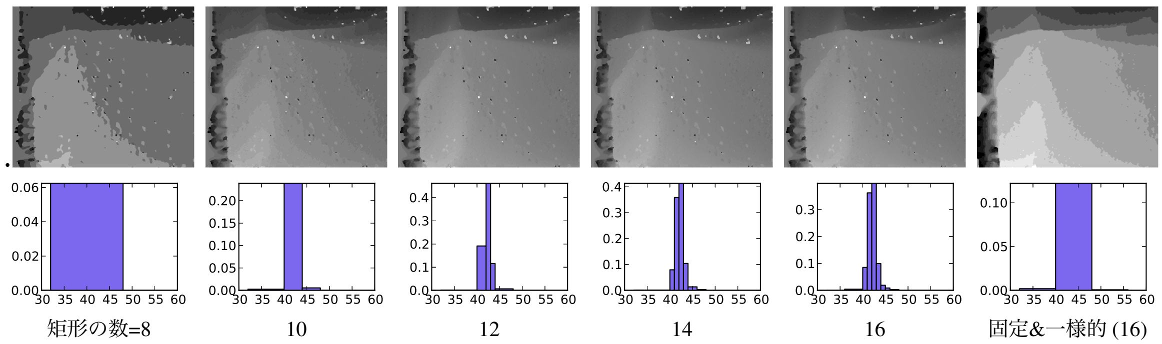


Fig. 4.7: 動的離散化を用いた平均場近似アルゴリズムの *Cloth1* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

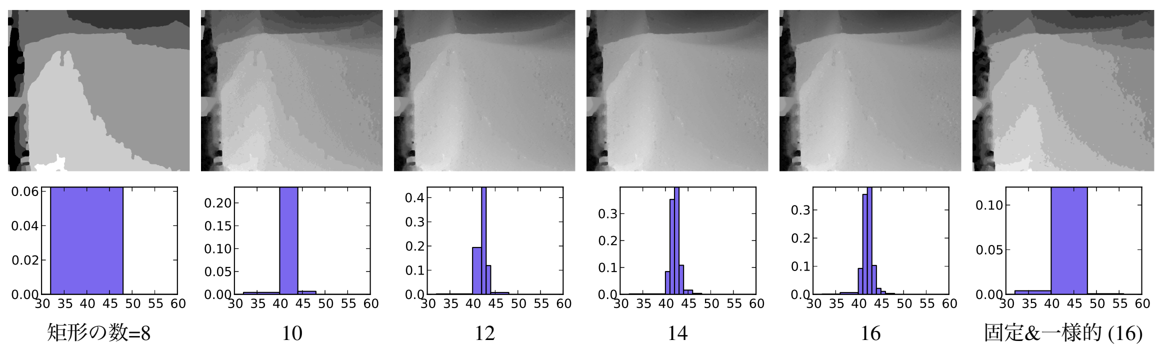


Fig. 4.8: 動的離散化を用いた確率伝搬法アルゴリズムの *Cloth1* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

4.5　本章の結論

本章で，われわれは連続的な MRF の問題を離散的な MRF の問題に離散化するための方法を新しく提案した．この定式化では，周辺分布は混合矩形分布によって表現される連続的な分布となる．これによって，変数空間を非一様に離散化した場合での平均場近似と確率伝搬法のアルゴリズムを新しく導出した．さらに，われわれは変数空間を疎から密に離散化することで，変数空間を動的に離散化するための方法についても新しく提案した．これによって，周辺分布のピーク位置を効率的に推定することが可能となった．複数の実験を通して，これらの手法が実際の問題に対して効果的であることを確認した．

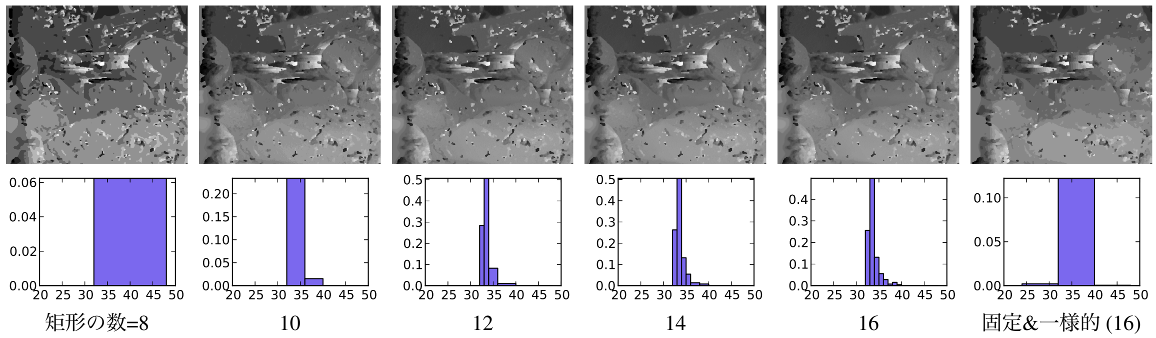


Fig. 4.9: 動的離散化を用いた平均場近似アルゴリズムの *Rocks1* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

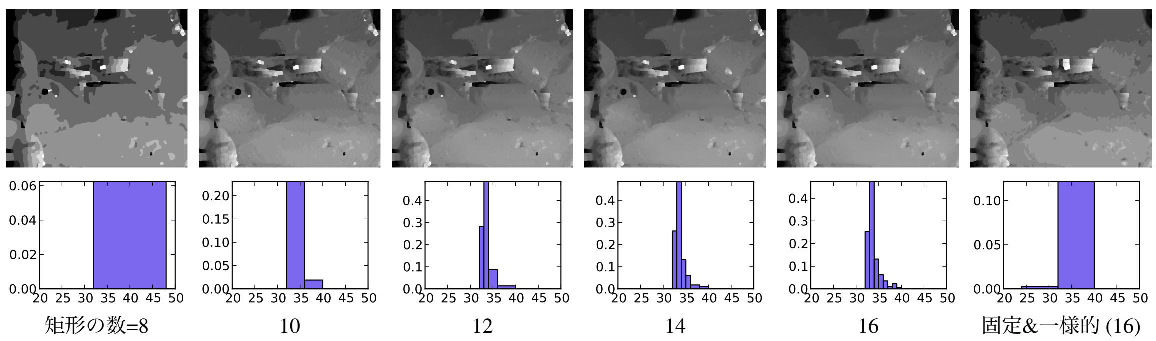


Fig. 4.10: 動的離散化を用いた確率伝搬法アルゴリズムの *Rocks1* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

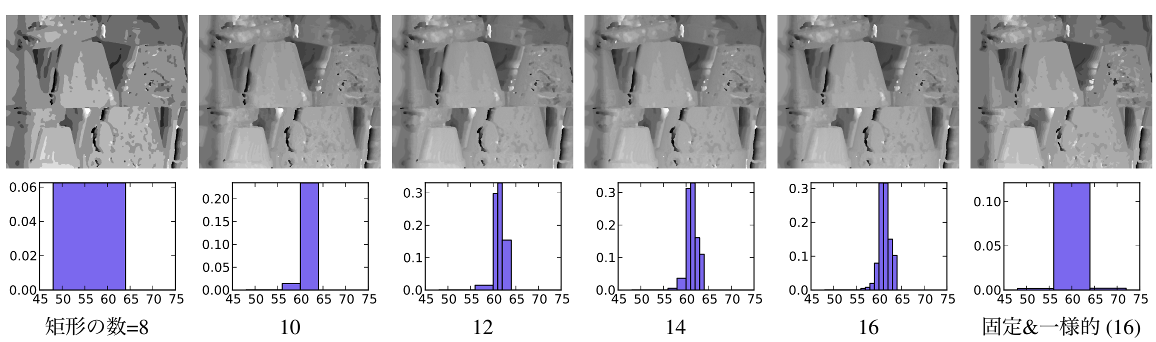
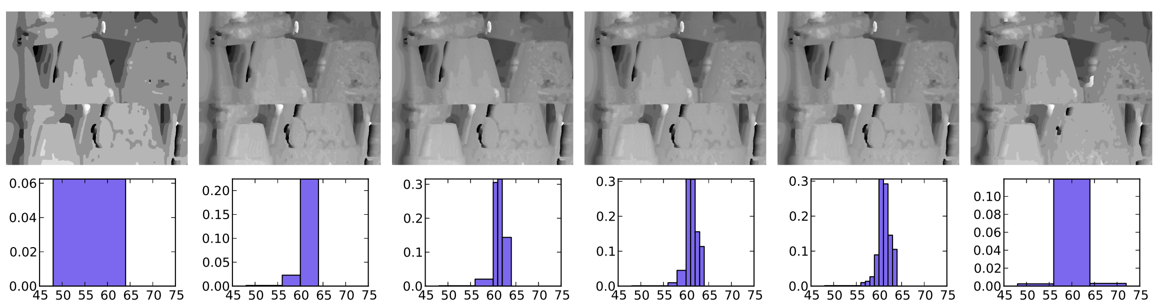


Fig. 4.11: 動的離散化を用いた平均場近似アルゴリズムの *Flowerpots* に対する結果．上段: 視差画像．下段: $(100, 100)$ の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．



矩形の数=8　10　12　14　16　固定&一様的 (16)

Fig. 4.12: 動的離散化を用いた確率伝搬法アルゴリズムの *Flowerpots* に対する結果．上段: 視差画像．下段: (100, 100) の位置でのサイトにおける，混合矩形分布で近似された周辺分布のグラフ．

第5章　結論

本研究では，コンピュータビジョンの最適化手法として特に重要な位置を占めるマルコフ確率場と呼ばれる確率モデルに焦点を当て，その最適化手法について主に2つの方向から拡張を行った．

第1章では，マルコフ確率場が多くのコンピュータビジョンの問題に対して用いられる理由について3つの点を指摘し，最適化手法の拡張がコンピュータビジョンの分野において重要であることを説明した．

第2章では，2つの研究に共通しているマルコフ確率場のモデルについてその詳細を説明した．次に，主要なMPM推定の手法である平均場近似と確率伝搬法について触れ，両者とも自由エネルギー最小の変分原理を基としていることについても説明し，実際に導出を行った．

第3章では，TAP方程式を用いた平均場近似の拡張手法について議論した．具体的には，平均場近似を拡張したTAP方程式が物性物理の分野で古くから用いられてるにもかかわらず，コンピュータビジョンの分野では殆ど用いられていない点について指摘し，その応用の有効性について説明した．次に，Plefka展開を用いたTAP方程式の導出方法について述べ，3次の次数までのTAP自由エネルギーを導出した．さらに，二値ならびに多値のTAP方程式を導出した．多値のTAP方程式に関しては物性物理の分野で対応する問題が存在しないため，新しく導出した．次に，平均場近似とTAP方程式をコンピュータビジョンの分野で主に用いられる確率伝搬法と比較した場合，主に3つの長所があることを指摘した．最後に，インタラクティブセグメンテーションとステレオマッチングの2つの実験を通して，TAP方程式のコンピュータビジョンに対する有効性を示した．

第4章では，非一様に離散化された変数空間上でのマルコフ確率場の推定手法について議論した．具体的には，従来は連続的なマルコフ確率場の問題を離散的なフレームワークで求める際に，エネルギー関数の離散化によって連続的なマルコフ確率場を離散的なものに変換した上で，平均場近似や確率伝搬法などの各種離散的な最適化を行なっていた．本研究ではこの離散化の手法を拡張し，変数空間が非一様な場合においても正しく周辺分布

を推定するための手法を新しく提案した．次に，この提案手法によって計算速度を保ちつつ精度の高い周辺分布を求めることが可能となる点について具体的に説明し，さらに非一様な離散化を効率的に行うための動的離散化の手法についても新しく提案した．最後に複数の実験を通して，本研究の提案手法が従来手法と比較して効果的であることを示した．

現在，多値ラベル分類問題に関しての TAP 方程式は 2 次に関してのみ導出されており，3 次以降に関しては導出されていない．ゆえに，3 次以降の多値ラベル分類問題に関する TAP 方程式を導出し，それを実際のコンピュータビジョンの問題に対して応用することで，実際の精度がどのような程度であるか確認する必要がある．

また，非一様な変数空間における現状の動的離散化については，最も大きなウェイトを取る矩形分布を分割していく戦略を取っているが，これが最良の周辺分布の推定に繋がるかどうかは理論的に明らかでないため，厳密な考察を行う必要がある．加えて §2.3.3 で述べたように，MPM 推定は MAP 推定を含む．これは，MAP 推定に関しても動的離散化の手法を用いて効率的な推定を行える可能性があることに繋がるため，より多くのコンピュータビジョンの問題について動的離散化の手法を適用できる可能性がある．これについても今後の研究で明らかにしていく所存である．

参考文献

[1] 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.

[2] A. Barbu. Learning Real-Time MRF Inference for Image Denoising. In *Computer Vision and Pattern Recognition*, 2009.

[3] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B(Methodological)*, 36:192–236, 1974.

[4] J. Besag. On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.

[5] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[6] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[7] R. L. Burden and J. Douglas Faires. *Numerical Analysis*. Brooks/Cole Pub Co, 2010.

[8] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, 2000.

[9] P. Clifford. Markov Random Fields in Statistics. In G. Grimmett and D. Welsh, editors, *Disorder in Physical Systems*, pages 19–32. Oxford University Press, 1990.

[10] D. Crandall, A. Owens, N. Snavely, and D. P. Huttenlocher. Discrete-Continuous Optimization for Large-Scale Structure from Motion. In *Computer Vision and Pattern Recognition*, 2011.

[11] R. C. Dubes, A. K. Jain, S. G. Nadabar, and C. C. Chen. MRF model-based algorithms for image segmentation. In *International Conference on Pattern Recognition*, volume 1, pages 808–814, 1990.

[12] A. E. Gelfand and A. F. M. Smith. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[13] D. Geman. Random Fields and Inverse Problems in Imaging. *In Saint-Flour Lectures 1988, Lecture Notes in Mathematics*, pages 113–193, 1990.

[14] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[15] A. Georges and J. S. Yedidia. How to expand around mean-field theory using high-temperature expansions. *Journal of Physics A: Mathematical and General*, 24(9):2173, 1991.

[16] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279, 1989.

[17] P. L. Hammer. Some network flow problems solved with pseudo-Boolean programming. *Operations Research*, 13:388–399, 1965.

[18] A. Ihler and D. McAllester. Particle Belief Propagation. In D. van Dyk and M. Welling, editors, *Artificial Intelligence and Statistics*, pages 256–263, Clearwater Beach, Florida, 2009.

[19] L. W. Johnson and D. R. Scholz. On Steffensen's Method. *SIAM Journal on Numerical Analysis*, 5:296–302, 1968.

[20] Y. Kabashima and D. Saad. The TAP approach to intensive and extensive connectivity systems. In *Advanced Mean Field Methods: Theory and Practice*, chapter 5. The MIT Press, 2001.

[21] H. J. Kappen and F. B. Rodriguez. Boltzmann Machine learning using mean field theory and linear response correction. In *Neural Information Processing Systems*, pages 280–286. The MIT Press, 1998.

[22] 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.

[23] D. Koller and N. Friedman. Exact Inference: Variable Elimination. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 9, pages 287–343. The MIT Press, 2009.

[24] D. Koller and N. Friedman. Inference as Optimization. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 11, pages 381–485. The MIT Press, 2009.

[25] D. Koller and N. Friedman. Learning Undirected Models. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 20, pages 943–1006. The MIT Press, 2009.

[26] D. Koller and N. Friedman. MAP Inference. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 13, pages 551–604. The MIT Press, 2009.

[27] D. Koller and N. Friedman. Partially Observed Data. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 19, pages 849–942. The MIT Press, 2009.

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

[29] D. Koller and N. Friedman. The Bayesian Network Representation. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 3, pages 45–102. The MIT Press, 2009.

[30] D. Koller and N. Friedman. Undirected Graphical Models. In *Probabilistic Graphical Models: Principles and Techniques*, chapter 4, pages 103–156. The MIT Press, 2009.

[31] N. Komodakis, N. Paragios, and Georgios Tziritas. MRF Optimization via Dual Decomposition: Message-Passing Revisited. In *International Conference on Computer Vision*, 2007.

[32] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *International Conference on Computer Vision*, pages 1018–1025, 2005.

[33] P. Krähenbühl and V. Koltun. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In *Neural Information Processing Systems*, pages 109–117. 2011.

[34] M. A. R. Leisink and H. J. Kappen. Learning in Higher Order Boltzmann Machines using Linear Response. *Neural Networks*, 13:2000, 1999.

[35] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition, 2009.

[36] M. Malfait and D. Roose. Wavelet-based image denoising using a Markov random field a priori model. *IEEE Transactions on Image Processing*, 6(4):549–565, 1997.

[37] 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 *International Conference on Computer Vision*, pages 416–423, 2001.

[38] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[39] T. Morita and T. Horiguchi. Exactly solvable model of a spin glass. *Solid State Communications*, 19(9):833–835, 1976.

[40] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy Belief Propagation for Approximate Inference: An Empirical Study. In *Uncertainty in Artificial Intelligence*, pages 467–475, 1999.

[41] R. M. Neal. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical report, University of Toronto, 1993.

[42] F. E. Nicodemus. Directional Reflectance and Emissivity of an Opaque Surface. *Applied Optics*, 4(7):767, 1965.

[43] 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*, 45(3):1159–1179, 2012.

[44] 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.

[45] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *American Association of Artificial Intelligence National Conference on AI*, pages 133–136, 1982.

[46] T. Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *Journal of Physics A: Mathematical and General*, 15(6):1971, 1982.

[47] D. Rajan and S. Chaudhuri. An MRF-Based Approach to Generation of Super-Resolution Images from Blurred Observations. *Journal of Mathematical Imaging and Vision*, 16(1):5–15, 2002.

[48] S. Roth and M. J. Black. Fields of Experts. *International Journal of Computer Vision*, 82(2):205–229, 2009.

[49] C. Rother, V. Kolmogorov, and A. Blake. ”GrabCut”: interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, 2004.

[50] H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 2005.

[51] D. Sherrington and S. Kirkpatrick. Solvable Model of a Spin-Glass. *Physical Review Letters*, 35:1792, 1975.

[52] E. Sudderth, A. Ihler, W. Freeman, and A. Willsky. Nonparametric Belief Propagation. In *Computer Vision and Pattern Recognition*, pages 605–612, 2003.

[53] J. Sun, N.-N. Zheng, and H.-Y. Shum. Stereo Matching Using Belief Propagation. *Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.

[54] 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 *European Conference on Computer Vision*, pages 16–29, 2006.

[55] K. Tanaka. Generalized Belief Propagation Formula in Probabilistic Information Processing Based on Gaussian Graphical Model. *The Institute of Electronics, Information and Communication Engineers*, 88:2368–2379, 2005.

[56] M. F. Tappen and W. T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In *International Conference on Computer Vision*, pages 900–907, 2003.

[57] 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 International Conference on Computer Vision*, pages 900–907, 2003.

[58] D. J. Thouless, P. W. Anderson, and R. G. Palmer. Solution of solvable model of a spin glass. *Philosophical Magazine*, 35:593–601, 1977.

[59] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *International Conference on Computer Vision*, pages 839–846, 1998.

[60] 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.

[61] M. Welling and Y. W. Teh. Approximate inference in Boltzmann machines. *Artificial Intelligence*, 143(1):19–50, 2003.

[62] L. Xu, J. Jia, and Y. Matsushita. Motion Detail Preserving Optical Flow Estimation. In *Computer Vision and Pattern Recognition*, 2010.

[63] 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 239–236. Morgan Kaufmann, 2003.

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

[65] C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. In *German Association for Pattern Recognition*, number 1, pages 214–223, 2007.

謝辞

修士課程 2 年間という短い期間にも関わらず，様々な方々にお世話になりました．簡潔ながらこの場をお借りしてお礼申し上げます．

出口光一郎教授には，研究に対する心構えやプレゼンテーションの仕方など，ゼミを通して幅広くご教授頂きました．常に物事の在り方について疑問を持ち本質を捉えようと探求する心構えは，研究だけでなく他の分野でも重要な事柄であるように思います．ありがとうございました．

岡谷貴之准教授には，研究の進展について様々なアドバイスを頂き，まだ研究に不慣れであった僕にあらゆる面で多大なバックアップをして頂きました．恐らくこの研究室でなければ，博士課程へ進むことはなかったのではないかと思います．ありがとうございました．

嵯峨智助教授には，発表内容のご指摘やコンピュータの設定，車の運転など，研究室生活の様々な面でお世話になりました．ありがとうございました．

田中・和泉研究室の田中和之教授には，専門的な立場からマルコフ確率場についての重要な示唆を複数頂きました．厚く御礼申し上げます．また，同研究室の安田宗樹助教授にはお忙しい中お会いして頂き，専門で研究されている TAP 方程式についていくつかの貴重な助言を頂きました．ありがとうございました．

秘書の坂根明美さんには事務手続きだけでなく，様々な内容の話し相手にもなって頂きました．ありがとうございました．

同輩の阿部君，柳澤君，馬君には，研究を共にする学友として様々な面でお世話になりました．お陰様で楽しい研究室生活を送ることができました．ありがとうございました．

最後に，共に大学生活を進めるにあたってお世話になったすべての方々と，生活を支えて頂いた家族へ，心から感謝申し上げます．