

MAP-MRFに基づく最適化(II)

BP法

元々のアルゴリズムは，ループのないネットワーク
に対して考案されたもの。

大局的最適解 への収束が保証。

Loopy Belief Propagation :

大局的最適解への収束が保証されない。

画像の持つ情報の局所性
画像のMRFモデル表現 $\left\{ \begin{array}{l} \text{画像の持つ情報の局所性} \\ \text{画像のMRFモデル表現} \end{array} \right\} \Rightarrow \text{CVへのLoopy BPの適用}$

CVにおける有効性の確認

MRF画像モデル

サイト i における隠れ状態変数 : $x_i \Rightarrow \mathbf{x} = \{x_i\}_{i=1,N}$

サイト i における観測変数 : $y_i \Rightarrow \mathbf{y} = \{y_i\}_{i=1,N}$

事後確率 :

$$P(\mathbf{x} | \mathbf{y}) \propto P(\mathbf{y} | \mathbf{x})P(\mathbf{x})$$

$$= P(y_1, y_2, \dots, y_N | \mathbf{x})P(x_1, x_2, \dots, x_N)$$

$$= P(y_1 | \mathbf{x}) \cdots P(y_N | \mathbf{x})P(x_1, x_2, \dots, x_N)$$

(\because 観測変数は互いに独立)

$$= P(y_1 | x_1) \cdots P(y_N | x_N)P(x_1 | n(1)) \cdots P(x_N | n(N))$$

($n(i)$: サイト i の近傍)

各サイト毎の条件付確率 :

→ クリークポテンシャルによる表現

マルコフネットワークについて

最大クリークサイズが2のMRF(Pairwise MRF)の
グラフィカル・モデルの場合

$\mathbf{x} = \{\mathbf{x}_s\}$: 隠れ状態変数, $\mathbf{y} = \{\mathbf{y}_s\}$: 観測変数

$s = 1, \dots, N$ (N : サイト数), \mathbf{x}_i の次元 = M (M : ラベル数)

事後確率 $P(\mathbf{x} | \mathbf{y})$

$$P(\mathbf{x} | \mathbf{y}) \propto \prod_s \varphi_s(\mathbf{x}_s, \mathbf{y}_s) \prod_r \prod_{t \in N(r)} \varphi_{rt}(\mathbf{x}_r, \mathbf{x}_t)$$

$\varphi_s(\mathbf{x}_s, \mathbf{y}_s)$: ノード \mathbf{x}_s における local evidence function

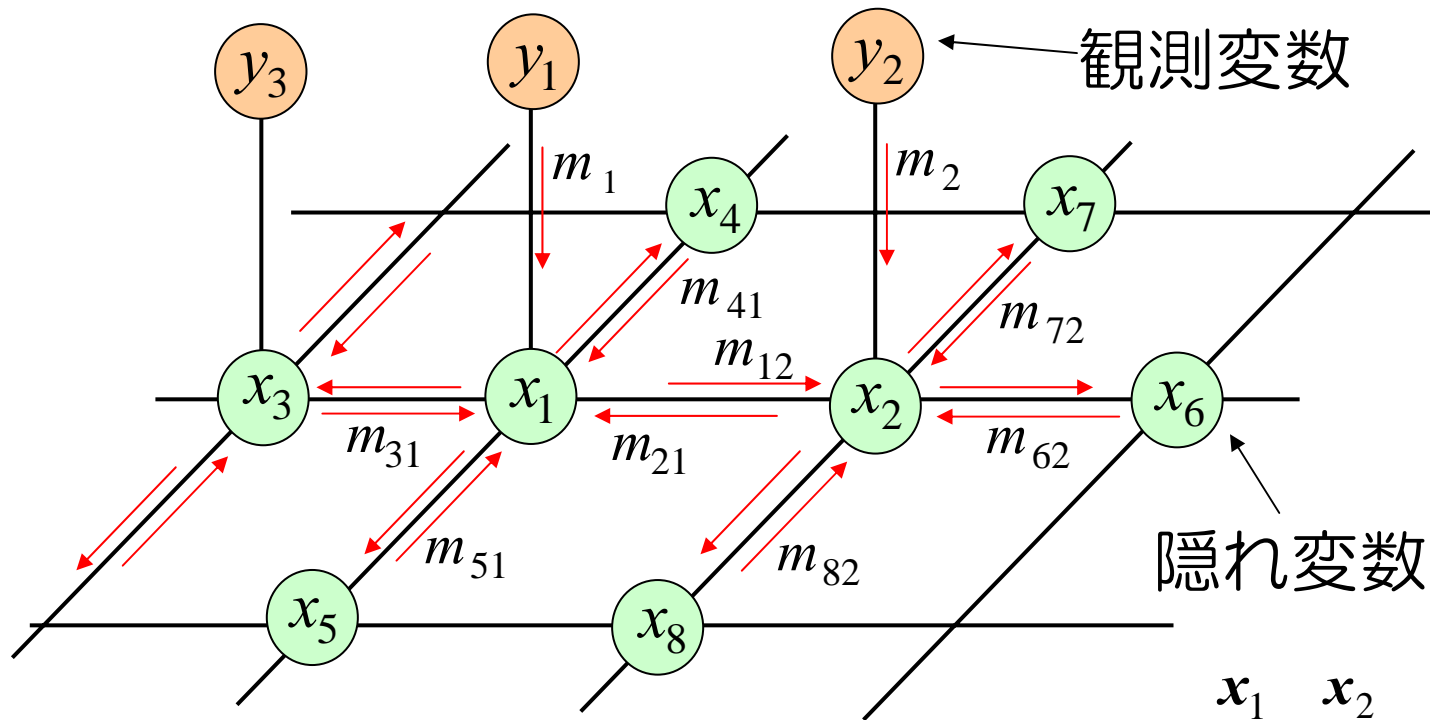
\Rightarrow 観測確率 $P(\mathbf{y}_s | \mathbf{x}_s)$ に等しい.

$\varphi_{rt}(\mathbf{x}_r, \mathbf{x}_t)$: ノード \mathbf{x}_r と \mathbf{x}_t の間の compatibility function

ネット表現された事後 確率の最大化

\Rightarrow BP(Belief Propagation)法

マルコフネットワークとBP

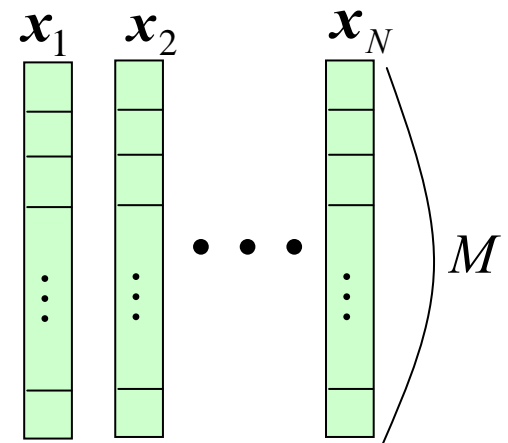


x_s から x_j へのメッセージ

$$\mathbf{x}_s : x_s(i)_{i=0, \dots, M-1} \rightarrow \mathbf{x}_t : x_t(j)_{j=0, \dots, M-1}$$

(各変数は M 個の要素からなる)

M : 例えば, 画像の階調数など



BP法による事後確率最大化

BP法：ネットワーク上でメッセージを繰返し伝播させながら推論を行うアルゴリズム

$m_{rt}(\mathbf{x}_r, \mathbf{x}_t)$: ノード \mathbf{x}_r からノード \mathbf{x}_t へのメッセージ.

$m_s(\mathbf{x}_s, \mathbf{y}_s)$: ノード \mathbf{x}_s で観測ノード \mathbf{y}_s から \mathbf{x}_s に送られるメッセージ.

$b_s(\mathbf{x}_s)$: ノード \mathbf{x}_s における確信度.

$\mathbf{x}_s, \mathbf{x}_r, \mathbf{x}_t, \mathbf{y}_s$: 全て M 次元ベクトル

BPアルゴリズムの更新規則

$\left\{ \begin{array}{l} \text{max - product アルゴリズム} \\ \text{sum - product アルゴリズム} \end{array} \right.$

標準的max-productアルゴリズム

1) 全てのメッセージ $m_{rt}(\mathbf{x}_r, \mathbf{x}_t)$ を均一分布に初期化

$$m_s(\mathbf{x}_s, \mathbf{y}_s) = \varphi_s(\mathbf{x}_s, \mathbf{y}_s) = P(\mathbf{y}_s | \mathbf{x}_s) \text{ とする.}$$

2) $m_{st}(\mathbf{x}_s, \mathbf{x}_t)$ を $i = 1, \dots, T$ で更新.

$$m_{st}^{(i+1)}(\mathbf{x}_s, \mathbf{x}_t) \leftarrow \kappa \cdot \max_{\mathbf{x}_s} \varphi_{st}(\mathbf{x}_s, \mathbf{x}_t) m_s^{(i)}(\mathbf{x}_s, \mathbf{y}_s) \prod_{\mathbf{x}_k \in N(\mathbf{x}_s)} m_{ks}^{(i)}(\mathbf{x}_k, \mathbf{x}_s)$$

3) 確信度の計算.

$$b_s(\mathbf{x}_s) \leftarrow \kappa \cdot m_s(\mathbf{x}_s, \mathbf{y}_s) \prod_{\mathbf{x}_k \in N(\mathbf{x}_s)} m_{ks}(\mathbf{x}_k, \mathbf{x}_s)$$

$$x_s^{MAP} = \arg \max_{x_s(i)} b_s(x_s(i))$$

得られた値 x_s^{MAP} は、ノード \mathbf{x}_s が取る最適ラベル

BP法によるDisparity検出

MRFにおけるモデル化：

サイト集合 $S = \{1, \dots, N\}$ (格子上の画素集合)

ラベル集合 $L = \{1, \dots, M\}$ (画素の視差値集合)

事後確率  事後エネルギー

max - product BPアルゴリズム

 min - sum BPアルゴリズム ($E(f)$ の最小化)

$$E(f) = \sum_{i \in S} \sum_{j \in N_i} V_2(f_i, f_j) + \sum_{i \in S} V_1(f_i), \quad f_i, f_j \in L$$

$V_2(f_i, f_j)$: サイト i と j のラベル f_i と f_j の間のコスト

$V_1(f_i)$: サイト i のラベル f_i のコスト

事後エネルギーの最小化 = 事後確率の最大化

以上のことを考慮すると，BPアルゴリズムは次のように表される．

$$m_{pq}^t(f_p, f_q) \leftarrow \min_{f_p} \left(V_2(f_p, f_q) + V_1(f_p) + \sum_{s \in N_p \setminus q} m_{sp}^{t-1}(f_s, f_p) \right)$$

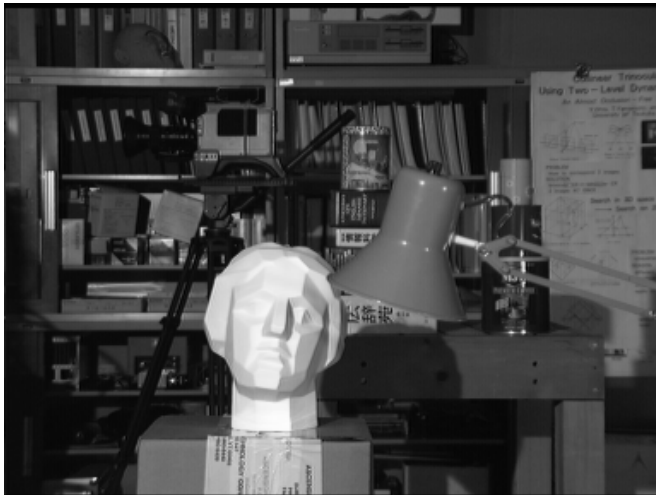
$N_p \setminus q$ は， q を除く p の近傍サイト集合

T 回の繰り返し処理の後，確信度は

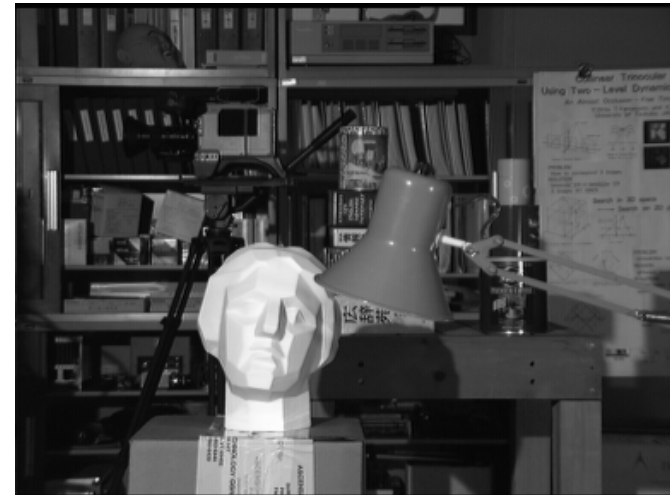
$$b_q(f_q) = V_1(f_q) + \sum_{p \in N_q} m_{pq}^T(f_p, f_q)$$

と表される． [Pedro F. Felzenszwalb et.al.]

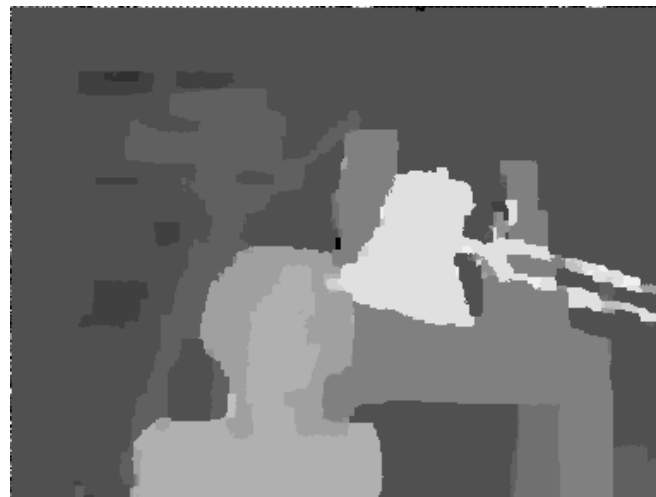
BP法によるDisparity計算の例



左画像



右画像



視差画像

マルコフチェーン（連鎖）とは

マルコフ性をもつランダム変数の系列： $\{X_0, X_1, X_2, \dots\}$

- 状態空間 $S = \{1, 2, \dots, N\}$
- 初期分布 π_0 : 時刻 t_0 でのマルコフ連鎖の分布確率

各状態 i に対して, $\pi_0(i) = P(X_0 = i) \geq 0$, $\sum_i \pi_0(i) = 1$

- 確率遷移行列： $\mathbf{Q} = (Q_{ij})$ 状態数 N が有限なら, $N \times N$ 行列

$$Q_{ij} = Q(i, j) = P\{X_{n+1} = j \mid X_n = i\}$$

一般に, Q_{ij} は時刻 n に依存.

マルコフ連鎖の例：「1匹のカエルが3つの池の間を飛び回る」

状態 $S = \{1, 2, 3\}$ は3つの池を表し，時間と共に変化．

カエルは，ある確率分布 π_0 に従って初期状態を選択．

$$\pi_0 = (1/2, 1/4, 1/4)$$

初期状態の選択： $U_0 \sim U(0,1)$ を生成

$$\mathbf{Q} = \begin{pmatrix} 1/8 & 3/4 & 1/8 \\ 1/3 & 1/9 & 5/9 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

$$X_0 = \begin{cases} 1 & \text{if } 0 \leq U_0 \leq 1/2 \\ 2 & \text{if } 1/2 < U_0 \leq 3/4 \\ 3 & \text{if } 3/4 < U_0 \leq 1 \end{cases}$$

とする．

乱数生成結果が $U_0 = 0.520$ であると， $X_0 = 2$ ．従って，カエルは行列 \mathbf{Q} の2行目の確率に従って，初期状態の選択と同様，次の状態 X_1 に移動． $U_1 \sim U(0,1)$ を生成した結果， $U_1 = 0.235$ であったとすると， $X_1 = 1$ となる． X_2 の状態は \mathbf{Q} の1行目に従って決まる．

$U_2 = 0.35$ が得られたとすると $X_2 = 2$ となり，次の状態 X_3 への遷移確率は $P(X_3 = j \mid X_2 = 2)$ と表され，状態 X_2 のみに依存．
一般に，

$$P_{X_{t(n)} \mid X_{t(n-1)}, \dots, X_{t(0)}}(x_n \mid x_{n-1}, \dots, x_1) = P_{X_{t(n)} \mid X_{t(n-1)}}(x_n \mid x_{n-1})$$

であるとき，この確率過程はマルコフ過程と呼ばれる．

マルコフ過程を遷移行列で表現すると，

$$\begin{aligned} & P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) \\ &= P(X_0 = i_0)P(X_1 = i_1 \mid X_0 = i_0)P(X_2 = i_2 \mid X_1 = i_1) \cdots \\ & \quad \cdots P(X_n = i_n \mid X_{n-1} = i_{n-1}) \\ &= \pi_0(i_0)Q(i_0, i_1)Q(i_1, i_2)Q(i_2, i_3) \cdots Q(i_{n-1}, i_n) \end{aligned}$$

結合密度関数が1ステップの条件付密度関数の積で表現．

π_0 : 初期分布 $\Rightarrow \pi_n$: 時刻 n での分布を表すとする.

即ち, $\pi_n(i) = P(X_n = i)$

状態空間が N 状態と有限であるとき,

$$\pi_{n+1}(i) = \sum_{j=1}^N \pi_n(j)Q(j,i), \quad \text{行列表現では} \quad \pi_{n+1} = \pi_n Q$$

ただし, $\pi_n = (\pi_n(1), \pi_n(2), \dots, \pi_n(N))$

これより, 帰納的に $\pi_n = \pi_0 Q^n$ が導かれる.

$c \geq 0$ に対して次式が成り立つとき, 移動不変な結合密度関数と呼ばれる.

$$\begin{aligned} P_{X_{t(1)}X_{t(2)}\cdots X_{t(n)}}(x_1, x_2, \dots, x_n) \\ = P_{X_{t(1)+c}X_{t(2)+c}\cdots X_{t(n)+c}}(x_1, x_2, \dots, x_n) \end{aligned}$$

基本極限定理：

X_0, X_1, \dots が既約で非周期的，かつ定常分布 $\pi(\cdot)$ を持つマルコフ連鎖であるとき， X_0 の任意の初期分布 π_0 に対して

$$\lim_{n \rightarrow \infty} \pi_n(i) = \pi(i), \text{ for } \forall i$$

が成り立つ．

マルコフ連鎖の性質

規約性： 全ての状態が遷移行列 Q で互いに繋がる

非周期性： 有限期間中，状態間で回りあうことがない

定常分布： ある分布 π を初期分布 π_0 として選んだとき，

$\pi_1 = \pi$ となるとき， π を定常分布と呼ぶ．

Markov Chain Monte Carlo Method

Markov Chain Monte Carlo (MCMC)法は、目的とする分布（例えば、ベイズ推定で重要な役割を果たす事後分布など）を表すような定常分布 π を持つマルコフ連鎖を設計し、そのサンプルを生成する方法である。

ある確率変数 X の関数 $g(X)$ の期待値 $E\{g(X)\}$ を求めることを考える。但し、 $X \sim \pi(x)$ とする。もし、 X_1, X_2, \dots, X_n が同一の分布 $\pi(x)$ に従う互いに独立な n 個の確率変数とすれば、

$$E\{g(X)\} = \frac{1}{n} \sum_{i=1}^n g(X_i) = \hat{g}(X)$$

で推定することができる。

MCMC法の代表的手法 $\left\{ \begin{array}{l} \text{Metropolis - Hastings Algorithm} \\ \text{Gibbs Sampler} \end{array} \right.$

多変量密度分布からのサンプル抽出が必要な場合：

ギブスサンブラが有効.

(条件付単一変数分布のみを扱うため)

$\mathbf{X} = (X_1, X_2, \dots, X_p) \in R^p$ を, 結合密度関数 $f(x_1, x_2, \dots, x_p)$ を持つ確率変数ベクトルとする.

このとき, f からの標本列がマルコフ連鎖を構成 (仮定) .

ギブスサンプリングを用いるための前提：

$$f_i(x_i | x_j, j \neq i), i = 1, \dots, p$$

が既知.

ギブスサンプルの概要

$\mathbf{X}^{(t)} = (X_1^{(t)}, X_2^{(t)}, \dots, X_p^{(t)}) \in R^p$ を, 時刻 t でのマルコフ連鎖の値とする.

このとき, 以下のように $\mathbf{X}^{(t)}$ から $\mathbf{X}^{(t+1)}$ に更新する.

1. $X_1^{(t+1)} \sim f_1(x_1 | X_2^{(t)}, X_3^{(t)}, \dots, X_p^{(t)})$ を生成.
2. $X_2^{(t+1)} \sim f_2(x_2 | X_1^{(t+1)}, X_3^{(t)}, \dots, X_p^{(t)})$ を生成.
- \vdots
3. $X_p^{(t+1)} \sim f_p(x_p | X_1^{(t+1)}, X_2^{(t+1)}, \dots, X_{p-1}^{(t+1)})$ を生成.

p 次元ランダムベクトル $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots$ は, 分布が $f(x_1, x_2, \dots, x_p)$ のマルコフ連鎖を構成する.

抽出した標本列： $X^{(0)}, X^{(1)}, X^{(2)}, \dots$

初期標本から M 個までの標本：

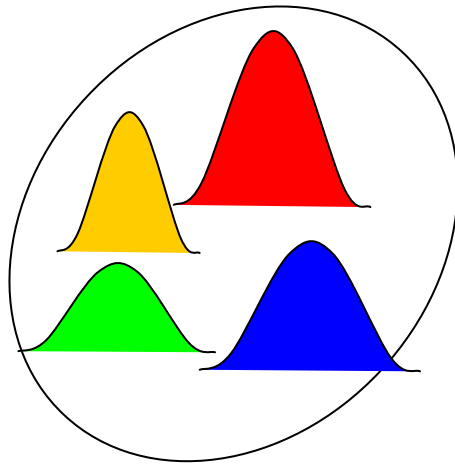
目的の分布から抽出されたものとは見なせない。

\therefore $\left(\begin{array}{l} \text{初期状態の記憶を失うまでに, ある時間を} \\ \text{必要とする.} \Rightarrow \text{緩和時間 (バーンイン時間)} \end{array} \right)$

緩和時間以降の標本は定常分布から抽出したものと見することが可能。

$f(x_1, x_2, \dots, x_p)$ の定常分布化 \Rightarrow 目的とする分布

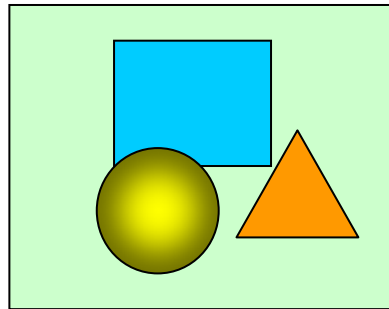
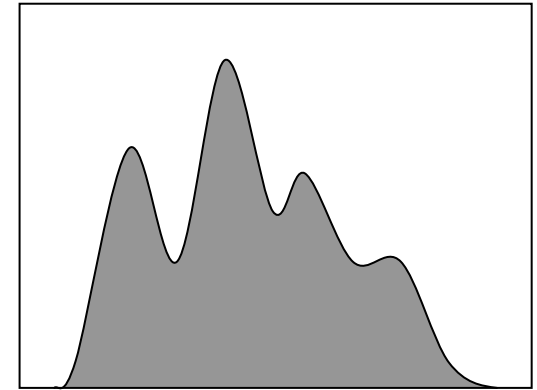
画像の統計モデル



K 個の正規分布の混合分布
として画像が存在 (仮定)

多変量混合正規分布モデル

$$p(\mathbf{x} | \theta) = \sum_{i=1}^K w_i N(\mathbf{x}; \mu_i, \Sigma_i)$$



観測データ: y

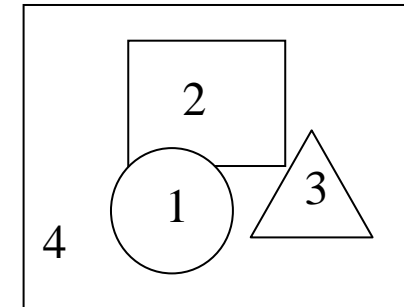
(色, テクスチャなど)

マルコフ確率場における
画像の文脈情報表現

ラベル x の生起確率

$$P(\mathbf{x} | \phi) = \frac{1}{Z} e^{-U(\mathbf{x} | \phi)}$$

ラベル配置: \mathbf{x}



事前情報:

$$U(\mathbf{x}) = \sum_{i \in S} V_1(x_i) + \sum_{i \in S} \sum_{j \in N_i} V_2(x_i, x_j)$$

画像分割モデル

ベイズ推定によるラベル x の事後（周辺）分布最大化

$$p(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\varphi} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}) \underbrace{P(\mathbf{x} | \boldsymbol{\varphi}) p(\boldsymbol{\theta}) p(\boldsymbol{\varphi})}_{\text{パラメータ}}$$

パラメータ：

$$\begin{cases} \text{多変量正規分布} : \boldsymbol{\theta} \cdots \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i \\ \text{マルコフ確率場} : \boldsymbol{\varphi} \cdots \text{外場係数 } \alpha_k, \text{相互作用係数 } \beta \end{cases}$$

$$P(\mathbf{x} | \{\alpha_\lambda, \lambda \in L\}, \beta) \approx \prod_{s \in S} P(x_s | x_{N \setminus s}, \{\alpha_\lambda, \lambda \in L\}, \beta)$$

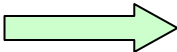
$$P(x_s | x_{N \setminus s}, \{\alpha_\lambda, \lambda \in L\}, \beta) = \frac{\exp\{-\alpha_{x_s} - \sum_{r \in N_s} \beta I(x_s, x_r)\}}{\sum_{\lambda \in L} \exp\{-\alpha_\lambda - \sum_{r \in N_s} \beta I(\lambda, x_r)\}}$$

を仮定.

画像分割問題は、前述の事後分布に基づいて、各画素のラベルとモデルパラメータに関する最適化問題となる。

$$\mathbf{x}^*, \boldsymbol{\theta}^*, \boldsymbol{\varphi}^* = \arg \max_{\mathbf{x} \in \Omega, \boldsymbol{\theta} \in \Theta, \boldsymbol{\varphi} \in \Phi} p(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\varphi} | \mathbf{y}) \quad (\text{MAP推定})$$

ラベルの配置空間 Ω は、解析的・数値計算的に評価困難。

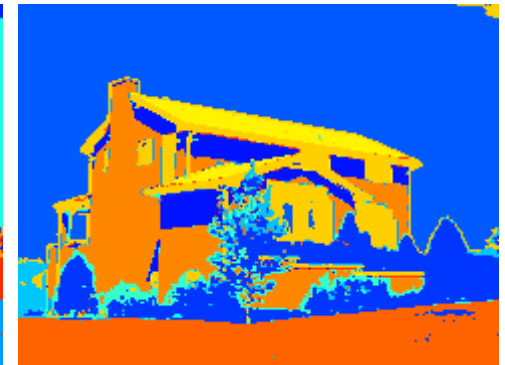
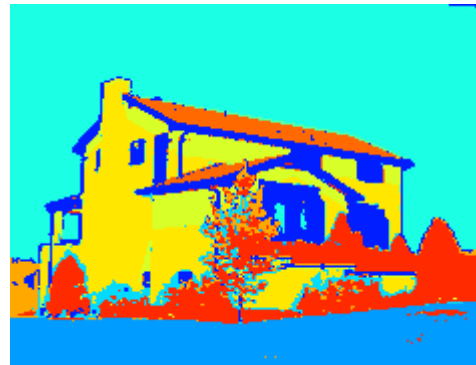
近似解法が必要  ギブスサンブラ

[青木、長橋]

ギブスサンブラによる画像の分割例 (1)



原画像



サイトラベル

線ラベル

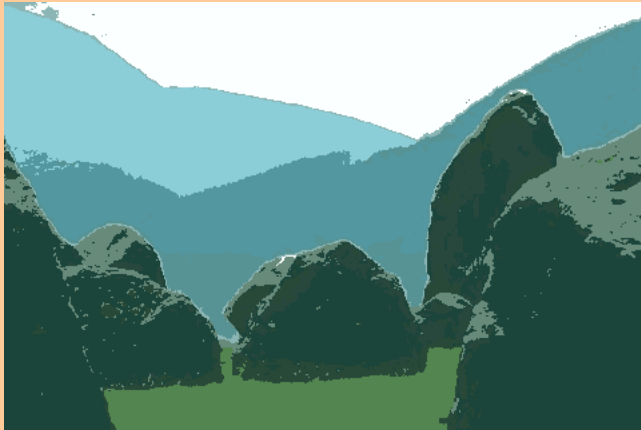
指定ラベル数
(使用ラベル数)

6(6)

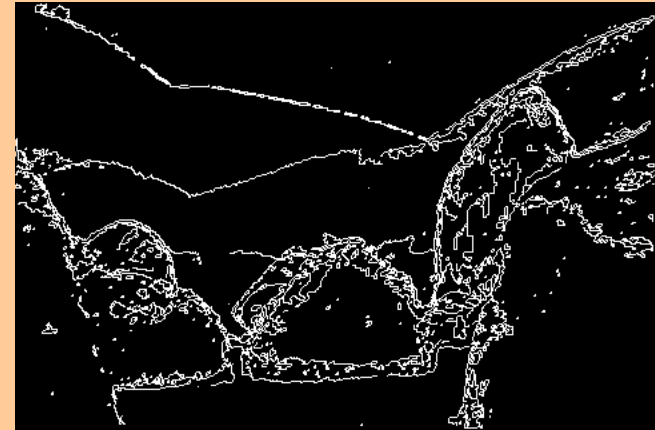
12(12)

20(12)

ギブスサンプラによる画像の分割例 (2)



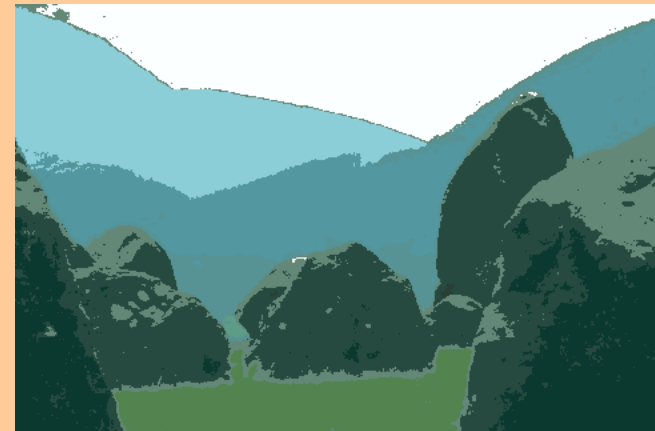
分割ラベル
(MAP推定, 線過程あり)



線ラベル
(MAP推定, 線過程あり)



分割ラベル
(MAP推定)



分割ラベル
(MPM推定)

テキストチャ生成 (1)

Gibbs Sampler法

Step1. f の初期化 $f_i \in L, f \in L^S$

Step2. for $i \in S$ do

2-1) $p_l = P(f_i = l | f_{N_i})$ for all $l \in L$ の計算. ←

2-2) f_i を p_l の確率でラベル l に設定. ←

Step3. Step 2 を N 回繰り返す.

条件付生起確率：ギブス分布から計算

$$P(f_i = l | f_{N_i}) = \frac{\exp\{\sum_{i' \in N_i} V_2(f_i, f_{i'})\}}{\sum_{f_i \in L} \exp\{\sum_{i' \in N_i} V_2(f_i, f_{i'})\}}$$

ルーレット選択によるラベル決定

テクスチャ生成 (2)

Metropolis法

Step1. f の初期化 $f_i \in L, f \in L^S$

Step2. for $i \in S$ do

2-1) f_i をランダムに $l \in L$ で置換した配置 f'

2-2) $p = \min\{1, \underline{P(f') / P(f)}\}$ の計算.

2-3) p の確率で f を f' とする.

Step3. Step2を N 回繰り返す.

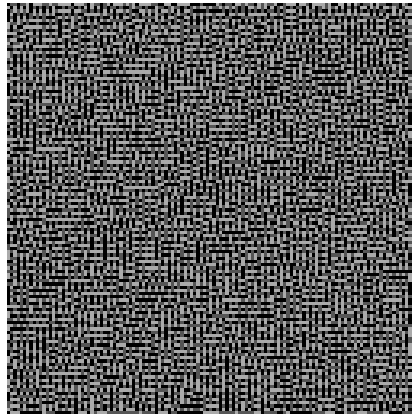
$$\begin{aligned}
 P(f') / P(f) &= \frac{\exp\{-\sum_{i' \in N_i} V_2(l, f_{i'})\}}{\exp\{-\sum_{i' \in N_i} V_2(f_i, f_{i'})\}} \\
 &= \exp\{\sum_{i' \in N_i} V_2(f_i, f_{i'}) - \sum_{i' \in N_i} V_2(l, f_{i'})\}
 \end{aligned}$$

テクスチャ生成の例 (Gibbs Sampler法)

$$M = 4,$$

$$\beta_1 = \beta_2 = -1$$

$$\beta_3 = \beta_4 = -1$$

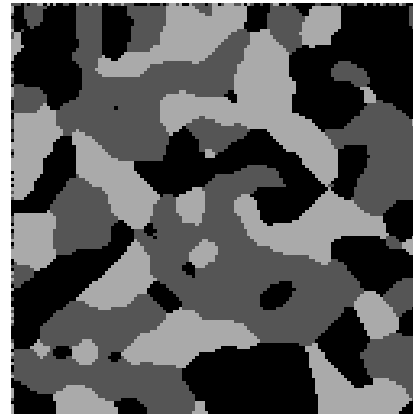


等方性

$$M = 3,$$

$$\beta_1 = \beta_2 = 1$$

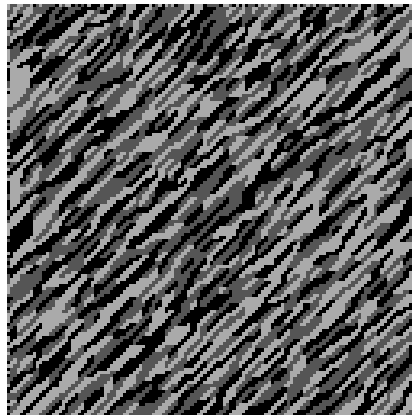
$$\beta_3 = \beta_4 = 1$$



$$M = 4,$$

$$\beta_1 = \beta_2 = \beta_3 = 1$$

$$\beta_4 = -1$$



非等方性

$$M = 4,$$

$$\beta_1 = \beta_2 = -1$$

$$\beta_3 = \beta_4 = 2$$



主な参考文献

S.Z.Li, Markov Random Field Modeling in Computer Vision, Springer-Verlag, (1995)

P.M.Lee, Bayesian Statistics, an introduction, third edition Hodder Arnold, (2004)

間瀬茂著, 空間データモデリング ―空間統計学の応用―, 共立出版, (2001)

上田修功, ベイズ学習[I] ―統計的学習の基礎―, 電子情報通信学会誌, vol.85,4, pp.265-271, (2002)

伊庭幸人, 種村正美著, 計算統計II マルコフ連鎖モンテカルロ法とその周辺, 岩波書店, (2005)

Pedro F. Felzenszwalb and Daniel P. Huttenlocher, "Efficient Belief Propagation for Early Vision", IEEE Conf. on Computer Vision and Pattern Recognition, (2004).

青木工太, 長橋宏: "ベイズ推定に基づく階層的事前分布とマルコフ確率場による画像領域分割", 信学論, vol.J88-D-II, no.7, pp.1318-1322 (2005)