論　文

局所情報と大域情報を統合する条件付確率場による画像ラベリング

豊田　崇弘[†a)]　　田上　啓介[†]　　長谷川　修[††]

Conditional Random Fields: Integration of Local Information and Global Information for Image Labeling

Takahiro TOYODA[†a)], Keisuke TAGAMI[†], and Osamu HASEGAWA[††]

あらまし　本研究では画像中のすべての画素にカテゴリーラベルを割り当てる画像ラベリングを行う．提案手法では従来手法のように局所的な特徴のみを用いるのではなく，大域的な特徴も用いてラベルの推定を行う．また，カテゴリー間の同時に起こりやすい関係や起こりにくい関係もモデル化し，ラベルの推定に利用する．局所的な推定と大域的な推定の統合には条件付確率場（Conditional Random Field: CRF）を利用する．従来のCRF は局所的な関係をモデル化するのには優れているが，大域的な関係は十分にモデル化できない．そこで複数の確率場を用いたモデル化が行われているが，モデル構造が複雑となるという欠点がある．これに対し提案手法では一つの確率場において推定を統合するためモデル構造が簡単である．また，従来手法では局所的な特徴への依存が大きいという問題があったが，提案手法では大域的な特徴も利用するためこの問題にも対応している．提案手法の有効性は 2 種類の屋外シーン画像のラベリング実験により確認した．大域的な特徴を用いることで認識精度は 10%程度向上し，大域的な視点から明らかな誤りは大きく減少した．

キーワード　画像ラベリング，シーン画像，条件付確率場，局所情報，大域情報

1. まえがき

現在，画像認識技術は広く実用化され，特定の条件下において特定の対象を認識する精度は非常に高くなっている．今後も更に画像認識の応用は広がり，それとともに一般的な条件下で多様な対象の認識が必要とされるようになる．現在，認識手法の多くは主に色やテクスチャなど局所的な画像特徴を利用している．しかし，対象とする画像が多様となるにつれ，大域的な画像特徴の重要性も増す．例えば，図 1 の例では局所的なパッチの情報のみではその領域の認識は困難である．しかし，そのパッチが下の画像の一部であることが分かれば認識が可能となる．この例のように一般的な画像の認識では大域的な特徴から得られる情報が大きな役割を果たす．本論文では局所的な特徴から得られる情報と大域的な特徴から得られる情報を効果的に統合するための新しい枠組みを提案する．

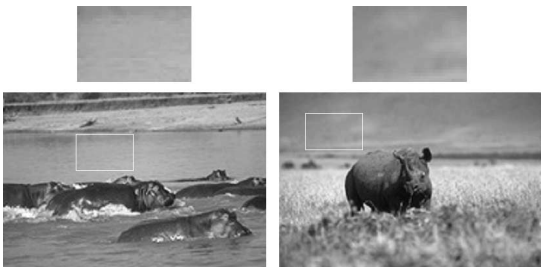

図 1　局所的な情報のみではパッチを認識するのが困難であるが，大域的な情報を用いることで容易に認識できるようになる

Fig. 1　Global information helps to recognize the patches that are difficult to recognize from local information.

本論文では提案手法を応用範囲の広いシーン画像のラベリングに適用し，その有効性を評価する．画像ラベリングは画像中のすべての画素に意味のあるカテゴリー（空，建物，車等）を割り当てる処理で，画像の領域分割と分割した領域についてのカテゴリー認識を

† 東京工業大学大学院総合理工学研究科，横浜市
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, R2–52, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, 226–8503 Japan
†† 東京工業大学像情報工学研究施設，横浜市
Imaging Science and Engineering Laboratory, Tokyo Institute of Technology, R2–52, 4259 Nagatsuta-cho, Midori-ku, Yokohama-shi, 226–8503 Japan
a) E-mail: toyoda@isl.titech.ac.jp

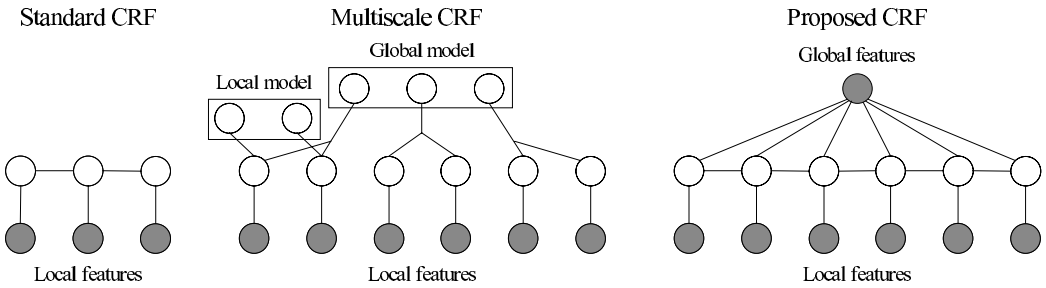

図 2　標準の CRF，従来手法の多重スケール CRF [1]，提案手法の CRF．丸は変数を表すノードで，灰色ノードは観測データを表す．
Fig. 2　Standard CRF, Multiscale CRF [1] (Conventional method) and Proposed CRF. Nodes represent variables and gray nodes represent observations.

行う．これにより画像中のどの位置に何があるかが記述される．画像ラベリングの応用としては物体の検出や認識，シーン理解などが挙げられる．

画像ラベリングではマルコフ確率場（Markov Random Field：MRF）[2] がしばしば利用される．MRF を利用する手法では入力画像 X に対応するラベリング $L = \{l_i\}_{i \in S}$ の事後確率 $P(L|X)$ を，同時確率 $P(L, X)$ を用いて表す．

$$P(L|X) \propto P(L, X) = P(L) \prod_{i \in S} P(\mathbf{x}_i | l_i). \qquad (1)$$

ここで $\mathbf{x}_i$ は格子点集合 S 内の格子点 i における入力画像の特徴，l_i は i のラベルを示す．

MRF では各格子点における特徴 $\mathbf{x}_i$ とラベル l_i の関係 $P(\mathbf{x}_i|l_i)$ とともに，近傍に割り当てられるラベル間の関係 $P(L)$ も考慮する．近傍格子点同士はなるべく同じラベルになるようにモデル化することで滑らかなラベルの分布が得られる．MRF に基づいた手法には木構造を利用した手法もある [3]．木構造を利用することで局所的なラベル間の関係のみでなく，大域的な関係もモデル化される．

画像ラベリングでは生成モデルと呼ばれる MRF のほか，識別モデルと呼ばれる条件付確率場（Conditional Random Field：CRF）[4] も利用される．CRF はポテンシャル関数 ϕ_i，ψ_{ij} を用いて事後確率 $P(L|X)$ を直接モデル化する．

$$P(L|X) = \frac{1}{Z} \prod_{i \in S} \phi_i(l_i|X) \prod_{i \in S} \prod_{j \in N_i} \psi_{ij}(l_i, l_j|X). \qquad (2)$$

ここで j は i の近傍 N_i の格子点，Z は正規化項である．ϕ_i は i におけるラベルの推定を行うもので，任意の関数が利用される．ψ_{ij} も同様に任意の関数が用いられ，i と j におけるラベルの関係を表す．MRF ではラベル間の関係 $P(L)$（式 (1)）は入力画像に依存しないのに対し，CRF では依存した形で表現される点が異なる．

CRF は図 2 左（Standard CRF）のように表現できる．灰色のノードは観測データを表し，ここでは各格子点における画像特徴を表す．灰色ノードと連結している白色ノードは各格子点に対応するラベルを表しており，ノード間のエッジは二つの関係を表す．

MRF は同時確率 $P(L, X)$ をモデル化するのに対し，CRF は事後確率 $P(L|X)$ をモデル化する．一般に同時確率 $P(L, X)$ は複雑な分布となる可能性が高いため，求めたい事後確率 $P(L|X)$ を直接モデル化する CRF の方が識別精度は高くなる．本論文でも CRF を利用した手法を提案する．

一般の画像では近傍ラベル間の相関が強い．そこで CRF では，各ラベルはその近傍のラベルのみに依存するというマルコフ性を仮定し，$P(L|X)$ をモデル化する．このとき局所的なラベル間の関係はモデル化されるものの，大域的なラベル間の関係はモデル化されない．そこで画像ラベリングでは，大域的な関係も表現できるように標準の CRF [4] を改良した手法 [1], [5] が利用される．[1] では低解像度のラベル分布と高解像度のラベル分布のモデルを用いて，局所的・大域的ラベル間の関係をモデル化する（図 2 中央）．また，[5] では CRF を階層化し，1 階層目で各画素から抽出した画像特徴を観測データとする CRF，2 階層目で 1 階層目の出力を観測データに対応させた CRF を用いてモデル化を行う．

従来の CRF を利用したラベリング手法 [1], [5] は，局所的なラベル間の関係と大域的なラベル間の関係を

表現するのに，複数のモデルや確率場を用いる．そのためモデルの構造が複雑となり，ラベリング処理に伴う計算コストも大きくなる．これに対し，本論文で提案する手法は一つの確率場において局所的な関係と大域的な関係の両方をモデル化するため，モデル構造が簡単である．その結果，ラベリング処理に伴う計算コストが抑えられる．

従来手法の問題点として，観測データに局所的な画像特徴しか使用しない点が挙げられる．このために，大域的なラベルの分布を表現するのに複雑なモデル構造が必要とされる．またモデル全体が局所的な特徴に大きく依存しており，局所的な特徴がもつ誤りやあいまいさがモデル全体に影響を与えるという問題点もある．更に，「建物領域中に空がある」というように大域的な視点からは明らかな誤り方をすることも問題である．

提案手法では以上のような問題点を解決するために，局所的な画像特徴と大域的な画像特徴の両方を抽出し，二つの特徴を観測データとして明示的に含んだモデルを設計する（図 2 右）．提案手法は一方の特徴のみに依存することがなく，また一つの確率場でモデルを表現するため，全体のモデル構造も簡単である．更に大域的な画像特徴の抽出処理や，それらから各画素のラベルを推定する処理を簡便にしたため，改良に伴う計算コストの増加はわずかである．

提案手法のモデルでは，局所的な特徴から例えば「この画素は青だから空だろう」とか「このようなテクスチャだから道路だろう」という推定を行う．一方，大域的な特徴からは，例えば「このような構図のシーンだから，画像の下部は道路でその両脇は建物だろう」という推定を行う．これらの推定は局所的・大域的という異なる視点からのもので，相補的な役割を果たす．更に近傍のラベル間の整合性や 1 枚の画像に同時に存在するラベル間の整合性を考慮し，最終的なラベリング結果を決定する．これにより局所的にも大域的にも整合性あるラベリングが実現される．

関連研究に，局所的な物体検出器とともに大域的な画像特徴を利用した物体検出がある [6]．この手法では，大域的な画像特徴をもとに目的の物体の画像上での位置と大きさを推定する．ただし推定する位置は垂直方向のみである．[6] では着目する物体は一つだけなので，検出器による推定と大域的な特徴からの推定は容易に統合できる．具体的には，二つの推定の位置と大きさの隔たりに応じてペナルティを科して統合している．これに対し，提案手法では画像ラベリングを行うため，複数のカテゴリーを同時に扱う．そして大域的な特徴からは各カテゴリーについて画像上での二次元の分布を推定する．また推定の統合は，局所的・大域的整合性，カテゴリー間の整合性，二次元分布としての整合性を考慮して行うため，物体検出と比べて難易度が高くなっている．一方，手法の一般性も高くなるため，物体検出を含め応用範囲は広くなる．

以下，**2.** で提案手法のモデルを述べ，**3.** で局所レベル，**4.** で大域レベルのモデル表現を述べる．**5.** では大域的なラベル間の関係，**6.** では局所的なラベル間の関係について述べ，**7.** ではラベリング処理の方法を述べる．**8.** でラベリング実験を示し，最後 **9.** で結ぶ．

2. 提案手法のモデル

提案手法では入力画像 X に対するラベリング L のエネルギー $E(L|X)$ を用いて事後確率 $P(L|X)$ を定義する．$E(L|X)$ が小さいほど整合性のある安定したラベリングであることを示し，$P(L|X)$ の値は大きくなる．$E(L|X)$ は四つの関数 f_i，g_i，g_L，h_{ij} を用いて計算する．

$$P(L|X) = \frac{1}{Z}\exp\{-E(L|X)/T\} \tag{3}$$

$$= \frac{1}{Z}\exp\left\{\sum_{i\in S} f_i(l_i|X) + \alpha\sum_{i\in S} g_i(l_i|X) + \beta\sum_{i\in S} g_L(l_i|X) + \gamma\sum_{i\in S}\sum_{j\in N_i} h_{ij}(l_i,l_j|X)\right\}, \tag{4}$$

$$Z = \sum_L \exp\{-E(L|X)/T\}. \tag{5}$$

ここで Z は正規化項，T は温度（本研究では $T=1$ に固定）を表す．l_i，l_j は画素 i とその近傍領域 N_i 内の画素 j に割り当てられているカテゴリーラベルを表す．α，β，γ は関数の重み付けをする．

f_i は局所レベルの関数，g_i は大域レベルの関数を表し，それぞれ局所的な画像特徴，大域的な画像特徴をもとに画素 i のラベルを推定する．g_L は大域的な特徴をもとにカテゴリーの共起確率を推定し，画像全体のカテゴリーの分布にバイアスをかける関数である．すなわち，1 枚の画像で同時に生じやすいカテゴリーの関係（例えば車と道路）や，同時に生じにくいカテゴ

リーの関係（例えばカバと雪）をモデル化する．

h_{ij} は近傍ラベル l_i，l_j の相互作用を表す関数で，二つのラベルの整合性を評価する．なるべく近傍のカテゴリーラベルが同じになるように h_{ij} は作用し，結果としてカテゴリーラベルの分布を平滑化する役割を果たす．ただし，提案手法のモデルでは h_{ij} は入力画像 X に依存し，画素間の特徴の類似度が高いほど相互作用が強く働く．その結果，カテゴリー分布の境界は画素間の特徴の類似度の低いエッジ部分に重なりやすくなる．

提案手法のモデルは図 2 右のように表現される．他の CRF との大きな違いは，観測データに局所的な画像特徴のみでなく，大域的な画像特徴も含めている点である．これにより一つの確率場による簡単な構造のモデル化が可能となっている．また，2 種類の相補的な特徴からの独立な推定経路をもつため，一方の特徴のみに依存することもなくなっている．

3. 局所レベルの表現

局所レベルでは画像中の各画素を色とテクスチャにより特徴づけ，それらをもとにカテゴリーラベルを推定する．色特徴は「空」や「草木」のように特徴的な色をもつカテゴリーを識別するのに有効である．しかし一方で光源の変化や視点の変化による影響を受けやすいという問題もある．また，同じカテゴリーの対象であっても，「白い雪」や「灰色の雪」のように色に多様性がある場合も多い．そこで提案手法では色特徴のほかに，照明条件の変化に頑健なグレースケールのテクスチャ特徴も利用する．

局所的な特徴からカテゴリーラベルを推定するには識別器を用いる．ここで種々の識別器が利用できるが，本研究では最近傍法に基づいた識別器を利用する．最近傍法に基づいた識別器は学習・認識処理が簡単で，また扱うカテゴリー数を容易に増やすこともできる．識別器を構成するにはテスト用データとは別に用意した学習用データを用いる．学習用データは学習用画像とその各画素に手動でカテゴリーラベルが付けられたデータのセットである．

3.1 局所レベルの特徴

実験では RGB カラー画像を用いるが，一般に RGB 値は照明条件の変化に敏感で，明るさの変化でも RGB 値は大きく変化する．そこで提案手法では各画素の RGB 値を CIE L*a*b*色空間に変換した 3 値を色特徴 $\mathbf{v}_{\mathrm{color}}(i)$ として利用する．L*a*b*は均等色空間であり，知覚的な色の違いが空間上での距離に相当するため，カラー画像を扱うのに適している．

テクスチャ特徴は，色の変化やカテゴリー内の色の多様性に対応するために，RGB 値をグレースケールに変換した画像から抽出する．提案手法では照明条件の変化に頑健な local binary pattern（LBP）[7] を用いてテクスチャ特徴を抽出する．[7] では LBP の回転不変への拡張が提案されているが，本研究では回転不変性を必要としないので回転不変ではない LBP を用いる．

半径 R, 近傍画素数 P の LBP（$LBP_{P,R}$）は以下のように定義される（詳細は [7] の式 (6)，(7) を参照）．

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \tag{6}$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0. \end{cases} \tag{7}$$

ここで g_c は着目している画素 c の輝度値，g_p は c を中心とする半径 R の円周上に等間隔に並んだ近傍画素 p の輝度値である．

LBP は着目画素と P 個の近傍画素との輝度値の大小関係により，2^P 通りの局所的なパターンに識別する．[7] では ‘uniform’ パターンと呼ばれる代表的な局所パターンを導入し，‘uniform’ パターン以外はすべて ‘その他’ のパターンとして一つにまとめている．具体的には以下のように $U(LBP_{P,R})$ を定義し，$U(LBP_{P,R}) \leq 2$ である $LBP_{P,R}$ を ‘uniform’ パターンとしている．

$$\begin{aligned} U(LBP_{P,R}) = {} & |s(g_{P-1} - g_c) - s(g_0 - g_c)| \\ & + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|. \end{aligned} \tag{8}$$

本研究では $P = 8$，$R = 1$ の LBP を用いるが，その場合，‘uniform’ パターンは 58 個存在する．

LBP は輝度値の大小関係のみに着目しているため，輝度値の階調変化に対して不変である．しかし一方で，わずかな輝度値の変化でも大小関係が変化すると局所パターンも変わってしまう．これではノイズなど外乱の多い実環境中で利用するには敏感すぎると考えられる．そこで提案手法では LBP に改良を加え，輝度値の変化に対し，より頑健な ‘改良 LBP’ を定義して用

いることにする．

改良 LBP では式 (7) の $s(x)$ の代わりに，しきい値 θ_{LBP} を導入した以下の $s'(x)$ を用いる．

$$s'(x) = \begin{cases} 1, & x \geq \theta_{\mathrm{LBP}} \\ 0, & x < \theta_{\mathrm{LBP}}. \end{cases} \tag{9}$$

改良 LBP では輝度値の差がしきい値 θ_{LBP} 以上のものに対して大小関係を定める．この改良により輝度値の階調変化に対する不変性はなくなるが，ノイズなどによる輝度値の変動に対しては頑健となる．

テクスチャ特徴は着目している画素を中心とする周囲の正方形領域 $\mathbf{R}_{\mathrm{LBP}}$（$= 9 \times 9$ 画素）に，改良 LBP を適用することで得る．改良 LBP が出力する領域内の局所パターンの頻度（ヒストグラム）がテクスチャ特徴となる．本研究では実験に用いる画像の質などを考慮し，$\theta_{\mathrm{LBP}} = 5$ に固定した．改良 LBP は 58 個の 'uniform' パターンと一つの 'その他' のパターンを出力するため，ビン数 59 のヒストグラムが各領域から得られる．ただし，テクスチャ特徴を抽出する正方形領域が画像の端となった場合は，画像上の領域のみに改良 LBP を適用することにする．

以上より，各画素は 3 次元の色特徴 $\mathbf{v}_{\mathrm{color}}(i)$ と 59 次元のテクスチャ特徴 $\mathbf{v}_{\mathrm{texture}}(i)$ により表現される．これが図 2 の各画素における観測データとなる．

3.2 代表ベクトルによるカテゴリー表現

最近傍法に基づいた識別器を用いて，局所的な特徴から各画素のカテゴリーラベルを推定する．このとき学習用データから抽出される特徴数は膨大なので，それらをすべて用いるのは効率が悪い．そこでカテゴリーごとに特徴をクラスタリングし，得られる代表ベクトルの集合により各カテゴリーを表現する．クラスタリングには Self-Organizing Incremental Neural Network（SOINN）[8], [9] を利用する．SOINN はノイズに対して頑健で，またデータの分布を表現するのに適当なクラスタ数を適応的に決定するなどの利点がある．

最初に，すべての学習用画像の各画素から **3.1** で述べた色特徴 $\mathbf{v}_{\mathrm{color}}(i)$ とテクスチャ特徴 $\mathbf{v}_{\mathrm{texture}}(i)$ を抽出する．そのうち同じカテゴリーラベルが付けられている画素の特徴を，テクスチャと色それぞれ別に SOINN によってクラスタリングする．SOINN の出力として得られる代表ノードの集合を各カテゴリーの代表ベクトルとする．

SOINN に特徴集合 V を入力したときの出力を SOINN(V) で表すと，カテゴリー c の色特徴とテクスチャ特徴の代表ベクトルの集合 U^c_{color}，U^c_{texture} は，

$$U^c_{\mathrm{color}} = \mathrm{SOINN}(V^c_{\mathrm{color}}), \tag{10}$$

$$U^c_{\mathrm{texture}} = \mathrm{SOINN}(V^c_{\mathrm{texture}}). \tag{11}$$

ただし，

$$V^c_{\mathrm{color}} = \{\mathbf{v}_{\mathrm{color}}(i) \mid \mathrm{HandLabel}(i) = c\}, \tag{12}$$

$$V^c_{\mathrm{texture}} = \{\mathbf{v}_{\mathrm{texture}}(i) \mid \mathrm{HandLabel}(i) = c\}. \tag{13}$$

ここで HandLabel(i) は手動で付けられた画素 i の正しいカテゴリーラベルである．

3.3 局所レベルのカテゴリー推定

式 (4) の f_i は以下のように定義される局所レベルの関数で，局所レベルの色特徴 $\mathbf{v}_{\mathrm{color}}(i)$ とテクスチャ特徴 $\mathbf{v}_{\mathrm{texture}}(i)$ を基に画素 i のラベルを推定する．

$$f_i(l_i|X) = \log(P(l_i|\mathbf{v}_{\mathrm{color}}(i))P(l_i|\mathbf{v}_{\mathrm{texture}}(i))). \tag{14}$$

ここでラベル l_i の確率分布 $P(l_i|\mathbf{v}_{\mathrm{color}}(i))$，$P(l_i|\mathbf{v}_{\mathrm{texture}}(i))$ は **3.2** で得た代表ベクトルの集合 U^c_{color}，U^c_{texture} を用いて計算する．

$P(l_i|\mathbf{v}_{\mathrm{color}}(i))$ は色特徴 $\mathbf{v}_{\mathrm{color}}(i)$ と各カテゴリーの代表ベクトルのうち，最近傍のものとの距離 $d^c(i)$ をもとに計算する．このとき距離尺度にはユークリッド距離（$\|\cdot\|$）を用いる．

$$d^c_{\mathrm{color}}(i) = \min_{\mathbf{u}} \|\mathbf{v}_{\mathrm{color}}(i) - \mathbf{u}\|, \ \mathbf{u} \in U^c_{\mathrm{color}}, \tag{15}$$

$$P(l_i = c|\mathbf{v}_{\mathrm{color}}(i)) = \frac{1}{Z_{\mathrm{color}}(i)} \exp\left(-\frac{d^c_{\mathrm{color}}(i)^2}{\sigma^2_{\mathrm{color}}}\right), \tag{16}$$

$$Z_{\mathrm{color}}(i) = \sum_c \exp\left(-\frac{d^c_{\mathrm{color}}(i)^2}{\sigma^2_{\mathrm{color}}}\right). \tag{17}$$

ここで $Z_{\mathrm{color}}(i)$ は正規化項，σ_{color} は識別能力を制御するパラメータである．$P(l_i|\mathbf{v}_{\mathrm{texture}}(i))$ も同様にして $Z_{\mathrm{texture}}(i)$，$\sigma_{\mathrm{texture}}$ を用いて計算される．

以上より，式 (14) は次のように書き換えられる．

$$f_i(l_i = c|X) = -\frac{d^c_{\mathrm{color}}(i)^2}{\sigma^2_{\mathrm{color}}} - \frac{d^c_{\mathrm{texture}}(i)^2}{\sigma^2_{\mathrm{texture}}} - \log\left(Z_{\mathrm{color}}(i)Z_{\mathrm{texture}}(i)\right). \tag{18}$$

ここで正規化項を含む右辺第 3 項目は，式 (4) の Z にまとめることができるので明示的に扱う必要はない．

4. 大域レベルの表現

シーン画像ではカテゴリーの分布の仕方にパターンや傾向がある．例えば，「地面は画像の下部」，「空は画像の上部」，「車がある画像には道路もある」，「カバのいる画像には雪はない」などである．特に類似シーンの画像ではカテゴリーの分布の仕方に共通点が多い．例えば，道路シーンの画像では「画像の下部は道路，上部は空」という分布が多い．このような性質を利用することで，未知の入力画像上のカテゴリーの分布を類似シーン画像から推定することが可能となる．すなわち，「この画像は道路シーンの画像に似ているので，画像の下部は道路だろう」のような推定ができる．提案手法ではカテゴリーの分布の傾向を学習用データから学習し，未知画像上のカテゴリーの分布の推定に利用する．

シーンの類似性は画像の局所的な特徴よりもむしろ大域的な特徴に反映されやすいと考える．例えば，「草木の多いシーン画像では画像全体に緑色が多く分布している」，「屋外シーンでは画像上部に青い空領域が広がっている」といった特徴がある．そこで提案手法では複数の大域的な特徴を利用して，学習用画像から入力画像に類似したシーン画像を検索し，それをもとにカテゴリーの分布を推定する．

4.1 大域レベルの特徴

大域的な画像特徴には，シーンの性質を記述するのに有効と考えられる 4 種類の異なる特徴を利用する．(1) RGB ヒストグラム，(2) 輝度値のこう配方向のヒストグラム，(3) ラプラシアンのヒストグラム，(4) CIE L*a*b*値．

ヒストグラムは領域内の位置の変化に不変なため，大域的な特徴を表現するのに適してる．提案手法では色とテクスチャの特徴を表すのにヒストグラムを用いる．ここで，テクスチャ特徴に局所レベルと同じ LBP を利用することもできる．その場合，局所レベルと特徴を共有できるため処理の効率化につながる．しかし提案手法では一つの特徴に大きく依存するのを避けるため，局所レベルと大域レベルとで異なる特徴を利用することにした．これにより多様な特徴からの推定が行えるようになる．

大域的な特徴には画像の構図も重要である．しかしヒストグラムでは位置情報が失われるため構図は表現されない．そこで入力画像の縮小画像を作成し，そのL*a*b*値も特徴として利用する．局所レベルでは画素単位の類似度を測るのに L*a*b*値を用いたが，ここでは縮小画像全体の類似度を測るのに用いる．

一般に学習用画像の数は限られているため，画像全体が類似しているシーン画像を検索するのは困難なことがある．そこで提案手法では各画像を左上，右上，左下，右下の四つの領域に等分し，それぞれの領域に着目して特徴の記述，類似シーン画像の検索を行う．このとき各領域間の特徴には相関があるため，着目している領域外の特徴も重要な情報源となる．そこで各領域の類似シーン画像を検索する際に，着目領域の特徴を重視するように重み付けを行い，同時にすべての領域の特徴も利用する．これは四つの領域のヒストグラムを連結し，着目している領域のヒストグラムにのみ重み付けをすることで実現できる．このとき各領域の位置情報も保存されるため，もとの画像の大まかな構図も特徴として表現される．

（1） RGB ヒストグラム：画像の左上，右上，左下，右下の四つの各領域で RGB の三つのヒストグラムを作成する．各ヒストグラムのビン数を 16 とし，それぞれの領域で 48 $(= 16 \times 3)$ 次元の特徴を抽出する．

（2） 輝度値のこう配方向のヒストグラム：4 等分した画像のそれぞれ領域内において，各画素を中心とする 3×3 画素領域を考える．ただし最外側の画素は除く．各画素について，周囲 8 画素のうちグレースケールの輝度値の差が最大である画素に対してこう配方向を特定し，その方向の頻度をヒストグラムで表す．こう配なしも含め，9 ビンのヒストグラム（9 次元の特徴）を各領域について作成する．

（3） ラプラシアンのヒストグラム：等分した 4 領域内の最外側を除く各画素において，グレースケールの輝度値の 8 近傍ラプラシアンを計算する．ラプラシアン値 $[-255^8, 255^8]$ を 256 ビンに分けてヒストグラムを作成する．よって各領域で 256 次元の特徴が抽出される．

（4） CIE L*a*b*値：画像サイズをもとの 10 分の 1 に縮小した画像を作成し，その CIE L*a*b*値を特徴とする．縮小画像には 10×10 画素領域の平均 L*a*b*値を用いる．もとの画像サイズが $W \times H$ 画素では $\lfloor W/10 \rfloor \times \lfloor H/10 \rfloor \times 3$ 次元の特徴となる．

4.2 類似シーン画像の検索

4.1 で述べた大域レベルの特徴をもとに学習用画像から類似シーン画像を検索する．このとき画像の左上，右上，左下，右下の四つの領域についてそれぞれ別に類似シーン画像を検索する．

図 3　左上の画像は入力画像，右の四つの画像は四つの各領域について検索された類似シーン画像，左下は大域レベルの特徴のみを用いて推定したカテゴリーの分布
Fig. 3　Upper left: Input image, Right four images: Collected similar images, Lower left: Inference of the category distribution using the global features.

まず画像全体の特徴を得るために，四つの領域から抽出した（1）〜（3）の 3 種類のヒストグラムをすべて連結する．各領域からは 313 (= 48 + 9 + 256) 次元のヒストグラム特徴が抽出されるので，連結すると全体で 1252 (= 313 × 4 次元のヒストグラム特徴となる．ただし画像を検索する際は，着目している領域のヒストグラムに重み w を付け，その部分を重視するようにする．提案手法では $w = 3$ とし，着目領域とそれ以外の三つの領域の和が同じ重みになるようにした．

二つの画像 I_1，I_2 の領域 t (= { 左上, 右上, 左下, 右下 }) における類似度 $S(I_1^t, I_2^t)$ を測るのに，ヒストグラム部分ではインタセクション（min）を，L*a*b* 部分には二乗誤差を用いる．

$$S(I_1^t, I_2^t) = \sum_{t'} w_t(t') \sum_{k=1}^{K} \min\left\{ \mathrm{hist}_1^{t'}(k), \mathrm{hist}_2^{t'}(k) \right\} - \sum_{i \in S'} \frac{\|\mathrm{Lab}_1(i) - \mathrm{Lab}_2(i)\|^2}{\sigma_g^2}, \tag{19}$$

$$w_t(t') = \begin{cases} w, & \text{if} \quad t' = t \\ 1, & \text{otherwise.} \end{cases} \tag{20}$$

ここで $t' =$ { 左上, 右上, 左下, 右下 }，$\mathrm{hist}_1^{t'}(k)$ は画像 I_1 の領域 t' における k ビン目のヒストグラム値，K (= 313) はヒストグラム全体の長さを表す．S' は縮小画像の領域，$\|\mathrm{Lab}_1(i) - \mathrm{Lab}_2(i)\|^2$ は画素 i における L*a*b*値の二乗誤差の和を表す．$w_{t'}(t)$ は重み付け関数，σ_g はパラメータである．

式 (19) で算出される類似度 $S(I_1^t, I_2^t)$ の高い順に各領域 N 枚の類似シーン画像を検索する．N が小さいと少しの誤りでも大きな影響が生じるため，カテゴリーの分布の推定の信頼度は低くなる．反対に N が大きいと類似度の低い画像も検索結果に含まれてしまう．そこで本研究では学習用データの質と量を考慮し，$N = 3$ に固定した．

図 3 に類似シーン画像を検索した結果の例を示す．左上の画像がテスト用の入力画像で，右の四つの画像が各領域で最も類似度の高かった学習用画像である．入力画像と検索された画像の各着目領域を比較すると類似していることが確認される．

4.3　大域レベルのカテゴリー推定

4.2 で述べたように，画像の四つの領域 t (= { 左上, 右上, 左下, 右下 }) のそれぞれについて N 枚の類似シーン画像を検索し，それらを用いて未知の入力画像上におけるカテゴリーの分布を推定する．検索した学習用画像には手動でラベル付けされたデータ $B^c(i)$ がある．$B^c(i)$ は画素 i のラベルがカテゴリー c であるか '1'，ないか '0' を示す．

$$B^c(i) = \begin{cases} 1, & \text{if HandLabel}(i) = c \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

大域レベルの特徴 $\mathbf{v}_g$ が得られたときに画素 i のラベル l_i がカテゴリー c である確率 $P(l_i = c|\mathbf{v}_g)$ は，類似シーン画像のラベル付きデータ $B^c(i)$ の平均を求めることで得られる．しかし類似シーン画像の枚数 N

図 4　大域レベルの特徴によるカテゴリー分布の推定．明るい領域はカテゴリーの分布している可能性が高いことを示す．
Fig. 4　Inference from the global features. Bright pixels indicate high probability for the category.

が少ないため，推定される分布は一般性に欠けるかもしれない．そこで一般性を高めるために最初に平均値 0，標準偏差 σ_G の 2 次元の等方的なガウス関数 $G(i)$ を用いてカテゴリーの分布を平滑化する．

$$P(l_i = c|\mathbf{v}_\mathrm{g}) = \frac{1}{N}\sum_t \sum_{n=1}^{N} \left\{ \left(B^c_{t,n}(i) * G(i) \right) * M_t(i) \right\}, \tag{22}$$

$$M_t(i) = \begin{cases} 1, & \text{if } i \in R_t \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

ここで $*$ は畳込み積分を表し，$M_t(i)$ は領域 t の画像部分 R_t のみを通すマスクである．$B^c_{t,n}(i)$ は領域 t の n 枚目の類似シーン画像のラベル付きデータである．

式 (4) の大域レベルの関数 g_i は大域的な画像特徴 $\mathbf{v}_\mathrm{g}$ をもとに画素 i のラベルを推定する関数で，$P(l_i = c|\mathbf{v}_\mathrm{g})$ を用いて以下のように定義する．

$$g_i(l_i|X) = \log\left(P(l_i|\mathbf{v}_\mathrm{g}) + \varepsilon\right) \tag{24}$$

ε は真数が 0 となるのを避けるための小さい値の定数で，本研究では $\varepsilon = 10^{-4}$ とした．

図 3 で検索した類似シーン画像をもとに，カテゴリーの分布を推定した結果が図 4 である．ここに示すのは以下の実験で用いた七つのカテゴリーについてである．図の明るい領域がカテゴリーの分布している可能性の高いことを表しており，この例では，空は上部に道路は下部に，草木や建物はその周りに広がっている可能性が高いことが推定されている．これらの推定をまとめると図 3 の左下のようになる．このとき各画素の局所的な特徴は用いておらず，画像の大域的な特徴のみから推定している．それにもかかわらず，良い精度で推定が行えているのが確認される．

5. カテゴリーの共起モデル

カテゴリーの中には画像上に同時に生じやすい組合せや，反対に生じにくい組合せがある．例えば，車と道路は同時に生じる可能性が高いが，カバと雪は同時に生じる可能性は低い．このような性質を利用することで，画像全体として明らかな誤りを減少させることができる．また，推定があいまいなときなどにも有効である．提案手法ではこのようなカテゴリーの共起モデルを学習用データから学習し，ラベリング処理に活用する．

共起モデルは，**4. 2** で検索した四つの領域の類似シーン画像を基に学習する．まず，延べ $N \times 4$ 枚の画像について，すべてのカテゴリーの組 (c_1, c_2) の共起確率 $P(c_1, c_2|\mathbf{v}_\mathrm{g})$ を求める．これは $N \times 4$ 枚の画像のうち，(c_1, c_2) が同時に生じた画像の割合を計算すればよい．ここで $\mathbf{v}_\mathrm{g}$ は画像検索に用いた大域レベルの特徴を表している．

ラベリング処理では，学習した共起確率 $P(c_1, c_2|\mathbf{v}_\mathrm{g})$ を用いて，ラベル l_i のカテゴリーを推定する．

$$P(l_i = c|\mathbf{v}_\mathrm{g}) = \prod_{i' \in S} \delta(l_{i'}, c) \prod_{c'} P(c, c'|\mathbf{v}_\mathrm{g}) \tag{25}$$

$$= \prod_{c'} P(c, c'|\mathbf{v}_\mathrm{g})^{n^c_L}, \tag{26}$$

$$\delta(x, y) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

ここで n^c_L はラベリング L において，カテゴリー c が割り当てられている画素数である．

カテゴリーの共起モデルを表す式 (4) の g_L は $P(l_i = c|\mathbf{v}_\mathrm{g})$ を用いて定義される．しかし，式 (27) で計算される $P(l_i = c|\mathbf{v}_\mathrm{g})$ は非常に小さい値となるため，代わりに $P'(l_i = c|\mathbf{v}_\mathrm{g})$ を用いて以下のように定義する．

$$g_L(l_i = c|X) = \log\left\{P'(l_i = c|\mathbf{v}_\mathrm{g})\right\} \tag{28}$$

$$= \log\left\{\prod_{c'} \left(P(c, c'|\mathbf{v}_\mathrm{g}) + \frac{1}{2N} \right)^{\frac{n^c_L}{|S|}}\right\}. \tag{29}$$

ここで $|S|$ は全画素数を表す．また，$1/2N$ を加えるのは少数の検索画像から学習したモデルがオーバフィッティングするのを避けるためである．

6. 近傍ラベルの相互作用

自然画像上では普通，カテゴリーは連続に分布し，同一カテゴリーラベルの領域を形成する．このため近傍画素のラベルは同じカテゴリーであることが多い．一般的な MRF や CRF ではこの性質を利用し，近傍画素のラベルが同じになる確率が高くなるようにモデルを設計する．具体的には式 (1) の $P(L)$ や式 (2) の ψ_{ij} で，近傍ラベルがなるべく同じカテゴリーになるように相互作用をモデル化する．この相互作用は画像上でのカテゴリーラベルの分布を平滑化する役割を果たし，異なるカテゴリーラベルが散乱しないようにする．

一般の画像では複数のカテゴリーが分布するため，カテゴリーラベルの分布間には境界が存在する．カテゴリーの境界を正しく認識するためにはラベルの分布の平滑化のみでは不十分である．そこで CRF では入力画像に依存した適応的な平滑化を行う．これは，CRF では入力画像が与えられた条件下でモデルを設計するためにできることである．適応的な平滑化の例は，入力画像上でエッジが存在した場合，その画素間でラベルの相互作用が弱くなるようなモデル化である．

提案手法のモデルでは，近傍画素間の色特徴とテクスチャ特徴の距離，画像上の座標間の距離に応じてラベルの相互作用の強さを決定する．これにより近傍の画素で，特に特徴の類似度の高い画素と同じカテゴリーになりやすくなる．このような入力画像に依存した適応的な平滑化を行うことで，画像上でのエッジとカテゴリー分布の境界を重ねやすくできる．

近傍ラベル l_i，l_j の相互作用は式 (4) の h_{ij} で表現され，以下で計算される．

$$h_{ij}(l_i,l_j|X) = \phi(l_i,l_j)\exp\left(-\frac{d_{\text{color}}(i,j)^2}{\sigma_{\text{color}}^2}-\frac{d_{\text{texture}}(i,j)^2}{\sigma_{\text{texture}}^2}-\frac{d(i,j)^2}{\sigma_{\text{r}}^2}\right), \tag{30}$$

$$d_{\text{color}}(i,j) = \|\mathbf{v}_{\text{color}}(i)-\mathbf{v}_{\text{color}}(j)\|, \tag{31}$$

$$d_{\text{texture}}(i,j) = \|\mathbf{v}_{\text{texture}}(i)-\mathbf{v}_{\text{texture}}(j)\|, \tag{32}$$

$$d(i,j)^2 = \|x_i-x_j\|^2+\|y_i-y_j\|^2, \tag{33}$$

$$\phi(l_i,l_j) = \begin{cases} 1, & \text{if } l_i=l_j \\ -1, & \text{otherwise.} \end{cases} \tag{34}$$

d_{color}，d_{texture} は色特徴間，テクスチャ特徴間のユークリッド距離（$\|\cdot\|$）を表し，d は画像上の座標間のユークリッド距離を表す．ここで画素 i の座標は (x_i,y_i) としている．

ϕ は近傍ラベル l_i，l_j の整合性を判定する関数で，同じカテゴリーならば整合していると判断し，式 (3) のエネルギー $E(L|X)$ を小さくする．反対に異なるカテゴリーならば，$E(L|X)$ を大きくし，ラベリング状態が不安定であることを示す．σ_{color}，σ_{texture} は **3. 3** で用いたパラメータで，σ_{r} については **8. 1** で述べる．これらのパラメータは異なる特徴の距離を結び付ける働きをする．

一般的な MRF では式 (30) の指数関数部がモデル化されていない．そのため相互作用の強さが一定で，カテゴリーラベルの分布の平滑化のみが行われる．

7. ラベリング処理

ラベリング処理は，入力画像 X が与えられたときに，式 (4) の事後確率 $P(L|X)$ を最大にするラベリング L を求める処理である．一般に L のとり得る状態数は膨大で，最適解を求めるのは困難である．そこで近似解を求めることになる．近似解を推定する手法は種々あるが，本研究では推定処理が簡単なギブスサンプラー [2] を利用する．ギブスサンプラーでは式 (3) の温度 T の設定により推定処理を調整することもでき，提案手法のモデルの性能を評価するのに都合がよい．

ギブスサンプラーでは反復処理を通じて，$P(L|X)$ が大きくなるように逐次的にラベリング状態を更新し，MAP 推定を行う．1 回の処理で更新するのは着目している一つの画素のラベルで，全体のラベリング状態が安定するまで処理を繰り返す．最終的なラベリング結果は繰返し処理を停止した時点におけるラベリング状態とする．

ギブスサンプラーにおけるラベルの更新は，着目している画素以外のすべてのラベルを固定した条件で更新する．すなわち着目している画素のラベル l_i を条件付確率 $P(l_i|L^-,X)$ で更新する．ここで $L^-=\{l_{k(\neq i)}\}$ は l_i 以外のすべてのラベルを表す．

$$P(l_i|L^-,X) = \frac{P(l_i,L^-|X)}{P(L^-|X)} \propto P(l_i,L^-|X). \tag{35}$$

次に $P(l_i, L^-|X)$ を，エネルギー $E'(l_i)$ と $E'(l_i^-)$ を用いて書き換える．ただし Z' は正規化項，$E'(l_i)$ は，ラベル l_i を更新した後のエネルギー $E(l_i, L^-|X)$ のうち，l_i に依存する部分，$E'(l_i^-)$ は l_i に依存しない部分のエネルギーを表す．

$$P(l_i, L^-|X) = \frac{1}{Z'} \exp\left\{-E(l_i, L^-|X)\right\} \quad (36)$$

$$= \frac{1}{Z'} \exp\left\{-E'(l_i) - E'(l_i^-)\right\} \quad (37)$$

$$\propto \exp\left\{-E'(l_i)\right\}. \quad (38)$$

更新するラベル l_i に依存する部分のエネルギー $E'(l_i)$ は以下のように書ける．

$$-E'(l_i) = f_i(l_i|X) + \alpha g_i(l_i|X) + \beta g_L(l_i|X) + 2\gamma \sum_{j \in N_i} h_{ij}(l_i, l_j|X) \quad (39)$$

ここで第 4 項目の係数 '2' は，i の近傍画素 j $(\in N_i)$ と i $(\in N_j)$ を近傍とする画素 j の 2 通りの相互作用を表している．

初期状態のラベリング $L^0 = \left\{l_i^0\right\}_{i \in S}$ は式 (4) の f_i と g_i を考慮して決定する．すなわち各画素における初期ラベル l_i^0 は，

$$l_i^0 = \max_c \left\{f_i(l_i = c|X) + \alpha g_i(l_i = c|X)\right\}. \quad (40)$$

8. ラベリング実験

提案手法によるモデルの有効性をシーン画像のラベリング実験により検証した．実験データには他手法との性能比較のため，[1], [3], [5] で使用されている Sowerby 画像と，[1] で使用されている Corel 画像を利用した（http://www.cs.toronto.edu/~hexm/research.htm）．認識精度は正しいラベルが割り当てられた画素の割合で評価する．また本実験で使用した CPU は Intel Xeon 3.2 GHz である．

8.1 実験設定

実験に用いた Sowerby 画像は道路を中心とした屋外シーンの写真である．データには 96×64 画素サイズの画像とそれに対応するラベル付きデータが 104 枚ずつ含まれている．ラベル付きデータの各画素には七つのカテゴリーラベル（'sky', 'vegetation', 'road marking', 'road surface', 'building', 'street object', 'car'）か 'unlabeled' のいずれかが手動で割り当てられている．

Corel 画像はカバや北極グマなど野生動物の写った自然のシーン画像である．画像サイズは 180×120 画素で，画像とラベル付きデータが 100 枚ずつある．カテゴリーラベルは 7 種類，'rhino/hippo', 'polar bear', 'water', 'snow', 'vegetation', 'ground', 'sky' である．

学習用画像には無作為に選択した画像のセットを使用し，残りをテスト用画像とする．Sowerby 画像では 60 枚を学習用画像，44 枚をテスト用画像とする．Corel 画像では 60 枚を学習用画像，40 枚をテスト用画像とする．テスト用画像のラベル付きデータはラベリング結果の正解率を評価するためにのみ使用する．また実験では 'unlabeled' が割り当てられている画素は扱わないことにする．

パラメータは以下のように設定した．式 (4) $\alpha = 0.3$, $\beta = 10.0$, $\gamma = 1.0$．これらは各関数によるラベルの推定を統合するための重みである．本論文では実験的に値を決定しており，最適値とは限らない．しかしモデルの性能を評価するのには有効である．式 (4) $N_i = 17 \times 17$ 画素，式 (30) $\sigma_r = N_i$．提案手法のモデルでは式 (30) で，近傍画素との距離に応じて相互作用の強さを決定している．そこで近傍領域を広くとることにした．

式 (18) $\sigma_{\text{color}} = 5.0$, $\sigma_{\text{texture}} = 0.1$．これらは色特徴とテクスチャ特徴の識別能力を制御する．本実験では二つの識別能力が同程度になるように，標準偏差を基に設定した．式 (19) $\sigma_g = 100.0$．大域的な特徴の類似度を測る際，ヒストグラムと L*a*b*値が同程度の働きをするように設定した．

提案手法では **3.2** で述べたように，最近傍法に用いる特徴数を削減するために SOINN を利用する．実験では各カテゴリーの代表ベクトルの数が数十から数百になるように，SOINN のパラメータを $\lambda = 100$, $\text{age}_{\text{dead}} = 20$ とした [8]．ただし SOINN のパラメータは提案手法の処理速度には影響を与えるが，認識精度には大きな影響を与えなかった．

ギブスサンプラーによるラベルの更新処理は全体のラベリング状態が安定するまで繰り返す．実験では，Sowerby 画像で 5 万回，Corel 画像で 15 万回処理を繰り返した（図 6 参照）．ここで二つの繰返し処理の回数に差があるのは画像サイズが異なるからである．

8.2 ラベリング結果

図 5 に Sowerby 画像（上）と Corel 画像（下）のラベリング結果の例を示す．局所的な特徴のみをもとに各画素を独立に識別した場合（Pixel-wise classifier），カテゴリーラベルが散らばって分布しているのが確認

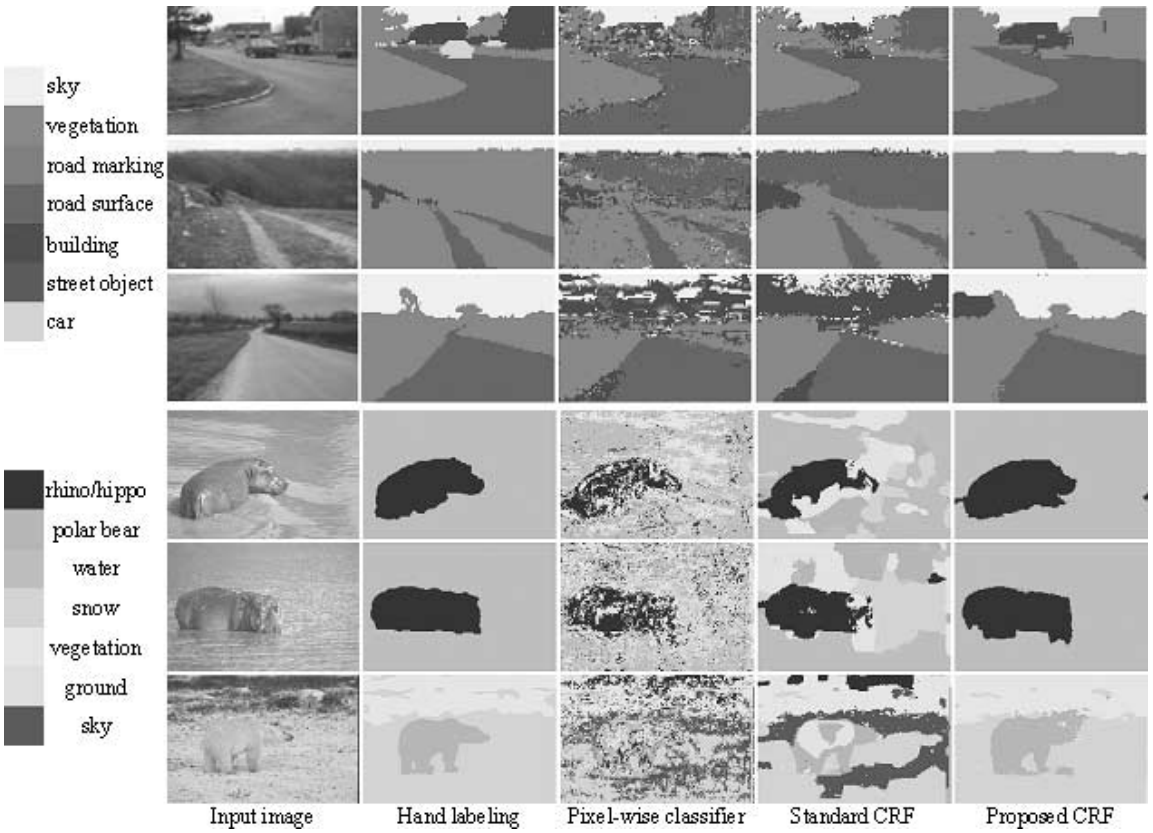

図 5　Sowerby 画像（上三つ）と Corel 画像（下三つ）のラベリング結果．局所特徴のみで各画素を独立に識別（Pixel-wise classifier），局所特徴のみの標準の CRF（Standard CRF），局所特徴と大域特徴を用いた提案手法（Proposed CRF）によるラベリング結果

Fig. 5　Labeling results by the pixel-wise classifier, the standard CRF with local features, and the proposed CRF with local and global features.

される．これは近傍ラベル間の関係を考慮していないためである．ここで標準の CRF（Standard CRF）により近傍カテゴリー間の相互作用をモデル化すると，カテゴリーの分布が平滑化され，各カテゴリーはある程度大きさをもった領域となる．

標準の CRF ではラベル間の大域的な関係がモデル化されていないため，画像上部の空付近に道路領域が形成されるというように，大域的な視点から明らかな誤り方をする．また局所的な特徴による識別（Pixel-wise classifier）をもとにラベリングを行うため，識別器の段階での誤りが大きな影響を与えている．例えば Sowerby 画像の三つ目の例では，初期ラベリングで雲を建物を識別したために，その誤りが平滑化によって広がっている．更に Corel 画像では，「カバと北極グマ」や「カバと雪」のように起こり得ない（学習用データでは生じていない）関係も生じている．

提案手法（Proposed CRF）では局所的な推定のみでなく，大域的な視点からも推定を行うことで，局所特徴によって生じる誤りを修正することができる．例えば，「空領域中の道路」や「北極グマ中の地面」のように大域的に明らかな誤りが修正された．更に提案手法ではカテゴリーの共起モデルを学習しているため，同時に起こり得ないような関係はなるべく出力しないようになっている．例えば，標準の CRF で誤った「カバ，北極グマ，雪」の関係も，整合性あるラベリング結果に修正している．このように，提案手法を用いることで画像全体として妥当なラベリング結果が得られる．

表 1 にラベリング正解率と，同じ実験データを使った他手法 [1], [3], [5] による正解率を示す．Sowerby 画

表 1　提案手法と各手法によるラベリング正解率の比較
Table 1　Labeling accuracies.

	Pixel-wise classifier	Standard CRF	Proposed CRF	Others [3]	[1]	[5]
Sowerby	78.2%	84.8%	**91.5 %**	90.7%*	89.5%	89.3%
Corel	53.3%	67.4%	**79.3 %**		80.0%	

* [3] では 192 × 128 画素の大きな画像が使用されており，学習用画像も 1 枚多い 61 枚である（テスト用画像は 43 枚）．このため利用できる情報量が多く，本実験よりも有利な実験設定となっている．

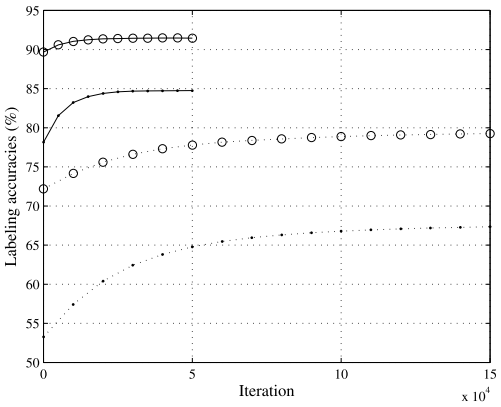

図 6　ラベリング処理の繰返し回数と正解率の関係．‘o’ は提案手法，‘·’ は局所特徴のみの標準 CRF の結果を表し，実線 ‘−’ は Sowerby 画像，点線 ‘···’ は Corel 画像への適用結果を表す．

Fig. 6　Labeling iteration and accuracies. ‘o’: Proposed method, ‘·’: standard CRF with local features, ‘−’: Sowerby images, and ‘···’: Corel images.

像では，正解率は標準の CRF（Standard CRF）で 84.8%だったのに対し，提案手法では 91.5%まで向上した．この結果はこれまでに報告されている他手法による正解率（90.7% [3]，89.5% [1]，89.3% [5]）よりも良好となっている．

Corel 画像では，画素単位の識別（Pixel-wise classifier）で 53.3%，標準の CRF で 67.4%だった正解率が，提案手法では 79.3%まで向上した．提案手法は簡単なモデル構造からなっており，複数のモデルを用いる [1] と比較してラベリング処理が簡便である．それにもかかわらず良好な精度が得られた点は重要である．

図 6 のグラフは，ラベリング処理の繰返し回数と正解率の関係を示している．‘o’ は提案手法，‘·’ は標準 CRF による結果で，実線 ‘−’ は Sowerby 画像，点線 ‘···’ は Corel 画像に対する結果を表している．提案手法による処理時間は，Sowerby 画像と Corel 画像それぞれ画像 1 枚当り 90 秒と 500 秒程度であった．このときの処理時間のほとんどは近傍ラベルとの相互作用の計算に要したものである．提案手法で導入した大域的な特徴の計算は容易で，また大域レベルの推定も，最初に計算した値を繰り返し用いるため計算コストの増加はほとんどなかった．

図 6 のグラフの左端 0 における正解率は初期ラベリングの精度を示している．提案手法ではこの時点において高い正解率を得ていることから，大域的な特徴による推定が初期ラベリングにおいて大きな役割を果たしていることが分かる．初期ラベリングを大域的な特徴のみを用いて行うと，Sowerby 画像で 77.2%，Corel 画像で 63.0%となった．このとき処理速度は画像 1 枚当り 1 秒以下と非常に高速である．提案手法はこのような優れた性質をもつ大域的な特徴を活用することで性能を発揮している．

9. むすび

本論文では条件付確率場（Conditional Random Field：CRF）を用いて，画像の局所的な特徴と大域的な特徴の両方から得られる推定を効果的に統合する手法を提案した．提案手法は二つの異なる視点からの特徴を用いることで一方の特徴に依存することをなくし，局所的にも大域的にも整合性あるラベリングを可能とした．

提案手法の有効性は 2 種類のシーン画像のラベリングに適用することよって示した．大域的な特徴を用いることで，大域的な視点から明らかな誤りを減少させることができ，また画像全体として不適当なラベリングを修正することもできた．提案手法は簡単なモデル構造を保持しており，計算コストをほとんど増加させずに認識精度を大幅に向上させることができた．

提案手法は一般的なモデルの枠組みであり，拡張性が高い．例えば，従来手法のように複数の確率場を用いるモデルに拡張することも可能である．この場合，モデル構造は複雑となるが認識精度の向上が見込まれる．提案手法の課題としてはパラメータの設定が挙げられる．本論文で実験的に決定した値は今後，最ゆう

推定の枠組みで扱うことを考える．

謝辞 本研究の実施にあたり NEDO 産業技術研究助成事業から支援を頂きました．記して感謝致します．

文　献

[1] X. He, R.S. Zemel, and M.Á. Carreira-Perpiñán, “Multiscale conditional random fields for image labeling,” Proc. Computer Vision and Pattern Recognition, vol.2, pp.695–702, 2004.

[2] S. Geman and D. Geman, “Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intell., vol.6, no.6, pp.721–741, 1984.

[3] X. Feng, C. Williams, and S. Felderhof, “Combining belief networks and neural networks for scene segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.4, pp.467–483, 2002.

[4] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” Proc. 18th International Conference on Machine Learning, pp.282–289, 2001.

[5] S. Kumar and M. Hebert, “A hierarchical field framework for unified context-based classification,” Proc. Int. Conf. on Computer Vision, vol.2, pp.1284–1291, 2005.

[6] K. Murphy, A. Torralba, and W. Freeman, “Using the forest to see the tree: A graphical model relating features, objects and the scenes, and the scenes,” Advances in Neural Information Processing Systems 16, 2003.

[7] T. Ojala, M. Pietikäinen, and T. Mänpää, “Multiresolution gray-scale and rotation-invariant texture classification with local binary patterns,” IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.7, pp.971–987, 2002.

[8] S. Furao and O. Hasegawa, “An incremental network for on-line unsupervised classification and topology learning,” Neural Netw., vol.19, no.1, pp.90–106, 2006.

[9] S. Furao, An algorithm for incremental unsupervised learning and topology representation, PhD Thesis, Tokyo Institute of Technology, 2006.

（平成 18 年 4 月 7 日受付，9 月 25 日再受付）

豊田　崇弘　（学生員）

平 16 東工大・情報工学卒．平 17 同大大学院総合理工学研究科知能システム科学専攻修士課程了．現在，同大学院博士課程在学中．画像解析，パターン認識などの研究に従事．

田上　啓介

平 16 都立大・理・物理卒．平 18 東工大大学院総合理工学研究科知能システム科学専攻修士課程了．現在，（株）NTT データ．

長谷川　修　（正員）

平 5 東京大学大学院博士課程了．博士（工学）．同年電子技術総合研究所入所．平 11 カーネギーメロン大学ロボティクス研究所滞在研究員．平 13 産業技術総合研究所主任研究員．平 14 東京工業大学大学院理工学研究科付属像情報工学研究施設助教授．同年科技団さきがけ研究 21 に兼任．画像処理，パターン認識，ニューラルネットワークなどの研究に従事．人工知能学会，日本認知科学会，日本顔学会，IEEE CS 等各会員．