# 階層的領域分割法に基づく木構造条件付確率場による一般物体認識

奥村　健志†　　滝口　哲也††　　有木　康雄††

† 神戸大学大学院工学研究科　〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
†† 神戸大学自然科学系先端融合研究環　〒 657–8501 兵庫県神戸市灘区六甲台町 1–1
E-mail: †okumura@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

**あらまし**　コンピュータによる一般物体認識は，近年，ロボットビジョンや画像検索といった分野で，実現が強く求められている．従来手法には，局所領域から抽出される特徴と局所領域間のクラス共起関係を用いて，各局所領域のクラスを認識する，条件付確率場と呼ばれる手法が広く用いられている．しかしながら，この手法には，局所領域から抽出される特徴の識別能力が不十分であり，また，物体のスケール変化に頑健でないといった問題が存在する．この問題を解決するため，我々は，階層的領域分割法に基づく木構造条件付確率場によって，複数スケールにおける認識結果を統合する手法を提案する．7 クラスの画像データセットを用いた実験の結果，提案手法により認識率が 2.2%向上した．

**キーワード**　一般物体認識，画像セグメンテーション，条件付確率場，階層化

# Generic Object Recognition by Tree Conditional Random Field based on Hierarchical Segmentation

Takeshi OKUMURA†, Tetsuya TAKIGUCHI††, and Yasuo ARIKI††

† Graduate School of Engineering, Kobe University　1–1, Rokkodai, Nada, Kobe, 657–8501 Japan
†† Organization of Advanced Science and Technology, Kobe University　1–1, Rokkodai, Nada, Kobe, 657–8501 Japan
E-mail: †okumura@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

**Abstract**　Generic object recognition by a computer is strongly required in various fields like robot vision and image retrieval in recent years. Conventional methods use Conditional Random Field (CRF) that recognizes the class of each region using the features extracted from the local regions and the class co-occurrence between the adjoining regions. However, there is a problem that the discriminative ability of the features extracted from local regions is insufficient, and these methods are not robust to the scale variance. To solve this problem, we propose a method that integrates the recognition results in multi-scales by tree conditional random field based on hierarchical segmentation. As an experimental result to the image dataset of 7 classes, the proposed method has improved the recognition rate by 2.2%.

**Key words**　generic object recognition, image segmentation, conditional random field, hierarchization

## 1. はじめに

一般物体認識とは，実世界の画像における物体を，一般的な名称で認識することを指す．これはコンピュータビジョンの分野における最もチャレンジングな課題の一つである．しかしながら，計算機による人間の視覚機能の実現という観点から，これはロボットビジョンへの応用が期待されている．また，近年，デジタルカメラや大容量のハードディスクが広く普及したことにより，膨大な動画像を人手で分類し，検索することが困難になりつつある．そこで，一般物体認識により，計算機が自動で動画像を分類することで，人手による検索を容易にすることが期待されており，ますます一般物体認識の重要性が高まっている．

一般物体認識には，大きく分けて 2 種類の従来手法が存在する．一つは，画像全体のクラスを認識する手法である．この手法では，局所特徴の集合で画像を特徴付ける Bag of Features (BoF) [1] がよく用いられる．この大域特徴を用いて，サポートベクターマシンや潜在的意味解析で画像のクラスを認識する．

もう一方は，画像の各画素のクラスを認識する手法である．まず，領域分割法により，画素を，色やテクス

チャが類似した局所領域ごとにまとめる．そして，色特徴やテクスチャ特徴といった低次特徴を局所領域から抽出し，それらに基づいて，機械学習アルゴリズムなどにより局所領域のクラスを認識する．このアプローチの一つに，ガウス混合分布を用いた手法 [2] がある．しかし，この手法では，各局所領域に対して，独立に認識を行うため，曖昧な特徴しか抽出出来ない領域の認識が困難となる．その上，認識結果が，全体として一貫性のない不安定なものになりやすいという問題が生じる．

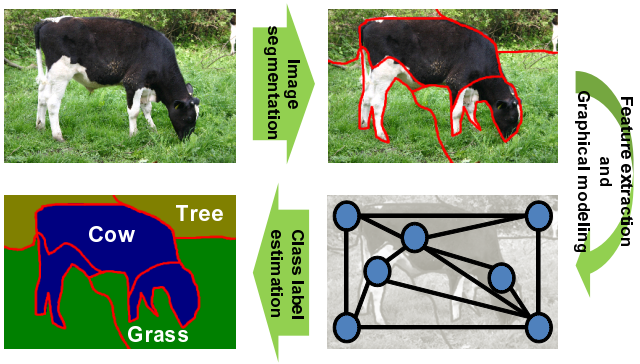


図 1　条件付確率場による画像の画素ごとのクラス認識

より一貫性のある安定した認識を実現するため，画像内の物体間には共起の関係が存在する，という考えに基づき，グラフィカルモデルである条件付確率場 [3] を用いた手法 [4] [5] が注目を集めつつある．これらの手法は，局所領域から抽出される低次特徴からだけでなく，隣接する領域間の共起関係まで考慮して，各局所領域のクラスを認識する．曖昧な特徴のみから認識が行われた局所領域の認識結果が，隣接する局所領域との関係を考慮することにより，改善される（図 1 を参照）．ここで，クラス共起とはコンテキスト情報の一種である．例えば，「牛」というクラスと「草」というクラスは同時に存在しやすいが，一方,「牛」というクラスは「車」というクラスとは同時に存在しにくい，といった情報のことである．


図 2　従来手法の問題点

しかし，従来手法の多くには共通の問題がある．それは，局所領域から抽出される特徴の識別性が不十分であり，また，これらの特徴は物体のスケール変化に頑健でない，という問題である．我々は，この問題は前処理として行う画像の領域分割において，スケールが単一であることに帰因していると考える．例として，図 2 を見てみると，これは実際に画像を領域分割したものであるが，2 種類の画像それぞれにおいて，領域分割された一つの局所領域が，画像内において，何に相当するかが異なっている．このように，小さな局所領域から得られる特徴は識別のための情報に乏しく，また，画像内に写る物体の大きさの違いを捉えきることが出来ない．そこで，この問題を解決するため，我々は，図 3 のように，細かい領域分割から粗い領域分割まで，複数のスケールで領域分割を行い，階層的領域分割法に基づく木構造条件付確率場 [3] を用いて，複数スケールにおける認識結果を統合する手法を提案する．


図 3　複数スケールにおける画像の領域分割

本論文の構成は以下の通りである．2 章では，提案手法について述べる．3 章では，提案手法の有効性を 7 クラスの画像データセットで評価した結果について述べる．4 章では，本論文のまとめと今後の方針について述べる．

## 2. 提 案 手 法

図 4 に提案手法の流れを示す．まず，入力画像に対し，Segmentation by Weighted Aggregation (SWA) [6] を適用する．SWA とは画像の階層的領域分割手法である．階層においては，より下層になるにつれて，細かく領域分割が行われ，また，より上層になるにつれて，粗い領域分割が行われる．最上層では領域分割は行われない．ある層における一つの局所領域は，一つ下の層の複数の局所領域に対応する．

それから，色特徴やテクスチャ特徴などの低次特徴を，最上層を除く各層の各局所領域から抽出する．最上層においてのみ，画像全体を特徴付けるのに適している Bag of Features (BoF) [1] を抽出する．それらに基づき，Gentle Adaboost [8] で全ての層の全ての局所領域に対して，認識対象のクラスの信頼度を算出する．このクラス信頼度の分布が，提案手法における局所特徴に相当する．

最後に，階層的領域分割の各層間の局所領域の対応関係に従い，木構造条件付確率場を構築する．これは，各局所領域におけるクラス信頼度の分布と局所領域間のク

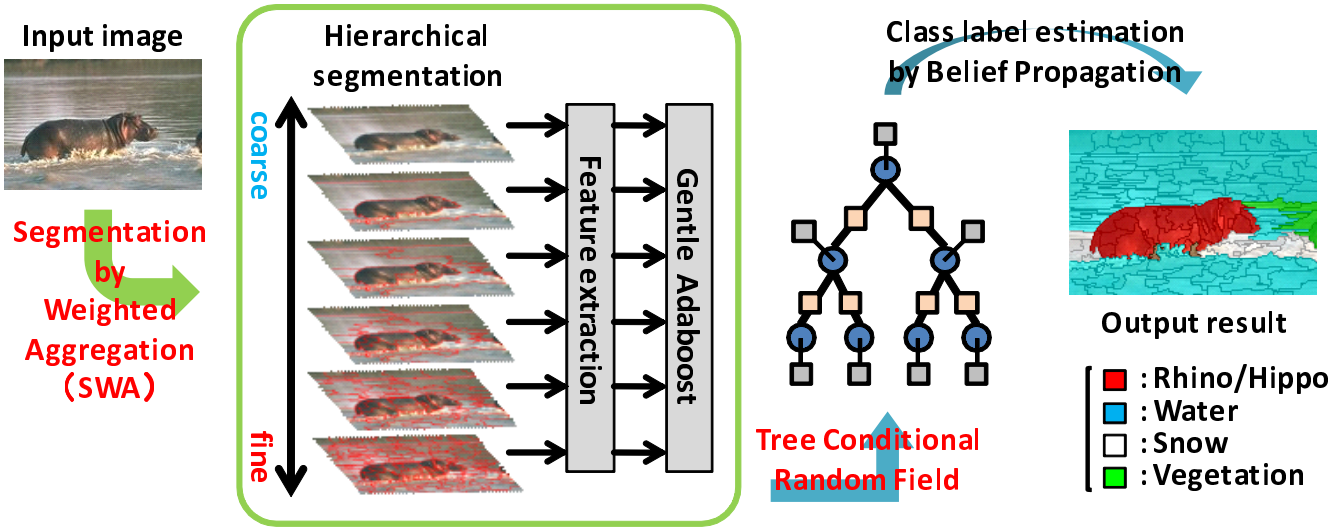


図 4　提案手法の流れ

ラス共起情報の総和である条件付分布として定義される．この分布関数を最大にするクラスの割り当てを，確率伝搬法 [7] により推定する．

次に，提案手法で用いた手法について，それぞれ詳細に述べる．

### 2.1　SWA による階層的領域分割

階層的領域分割のために，我々は Segmentation by Weighted Aggregation [6] を用いる．この手法では，画像を重み付きグラフとして捉える．グラフの節点が画素に対応し，グラフの辺は隣接する節点を結ぶものとする．グラフを切断，つまり，あるセグメント $S$ を切り出すための評価関数 $\Gamma(\mathbf{u})$ は以下のように定義される．

$$\Gamma(\mathbf{u}) = \frac{\sum_{i>j} w_{ij}(u_i - u_j)^2}{\sum_{i>j} w_{ij} u_i u_j} = \frac{\mathbf{u}^T L \mathbf{u}}{\frac{1}{2}\mathbf{u}^T W \mathbf{u}} \qquad (1)$$

ここで，$\mathbf{u} = \{u_1, u_2, \cdots, u_n\}$ は状態変数 $u_i$ のベクトルであり，あるノード $i$ に対して，$u_i = 1$ ならば，ノード $i$ はあるセグメント $S$ に属し，$u_i = 0$ ならば，ノード $i$ はあるセグメント $S$ に属さないことを意味する．また，$L$ はグラフのラプラシアン行列であり，$W$ は重み行列である．重みとは隣接ノード間の類似度のことであり，類似度が大きいほどその辺の切断が難しいといえるので重みが大きくなる．式 (1) の分子は，グラフの切断コストの関数となっており，分母はあるセグメント $S$ のサイズを表している．この評価関数を最小化するようなセグメントを求めることが目的となるが，これが意味することは，セグメント内のノード間の重みの総和は出来る限り大きくした上で，切り出される各セグメントの大きさが，なるべく均一になるようにすることである．

この最適なセグメンテーションを行うため，評価関数 $\Gamma(\mathbf{u})$ の最小化問題を，正の最小固有値 $\lambda$ に関する固有値問題 $L\mathbf{u} = \lambda W \mathbf{u}$ として解く．このとき，Algebraic MultiGrid (AMG) ソルバー [9] を用いて，再帰的粗大化 $\mathbf{u} \approx P\mathbf{U}$ を行い，この固有値問題を近似的に解く．$P$ は疎補間行列で，$\mathbf{U}$ は粗大化された状態ベクトルである．再帰的に繰り返すことで，状態変数が徐々に粗大化され，グラフ，つまり，入力画像は階層構造として表現される．粗大化を，一度行えば，状態変数の数はおよそ半分になる．提案手法では，階層構造の各層と層間の対応関係を用いる．

### 2.2　認識に用いる特徴

階層的領域分割の後，我々は階層の最上層を除く各層の各局所領域から，以下の低次特徴を抽出する．

- RGB，HSV，Lab，YCbCr 色空間の各要素
- ガボールフィルタと LoG のフィルタ応答
- 局所領域の重心座標
- 局所領域の面積

色特徴とテクスチャ特徴については，各画素からの特徴抽出後，平均，標準偏差，歪度，尖度といった統計量を，各局所領域ごとに計算する．

また，階層の最上層においては，画像から Bag of Features (BoF) [1] を抽出する．BoF はアピアランスベースの特徴表現手法である．まず，SIFT (Scale-Invariant Feature Transform) [10] のような局所特徴を画像から抽出し，それらを k-means 法によって，W 個のクラスタに分割する．各クラスタの重心ベクトルは Visual Words と呼ばれ，単語数 $W$ は実験的に決められる．このようにして，画像を Visual Words の頻度ヒストグラムとして表現する．BoF による画像の特徴付けには，主に 2 つの利点が挙げられる．一つは，局所特徴の集合として表現されているため，オクルージョンに頑健である．もう一つは，k-means 法によって，ベクトル量子化が行われているため，アピアランスの変化に頑健であることが挙げられる．

これらの特徴が，階層の各層の各局所領域をそれぞれ特徴付け，それらに基づき，認識対象の全クラスに

対する信頼度が Gentle Adasboost [8] によって計算される．Gentle Adaboost とは，多数の弱識別器の重み付き投票によって出力が決定される，ブースティング学習の一種である Adaboost から派生したものである．Gentle Adaboost は二値判別の識別器であり，各層ごとに学習を行うため，認識対象のクラス数×階層の層数の数だけ識別器を用意し，学習を行う．

### 2.3 木構造条件付確率場による認識

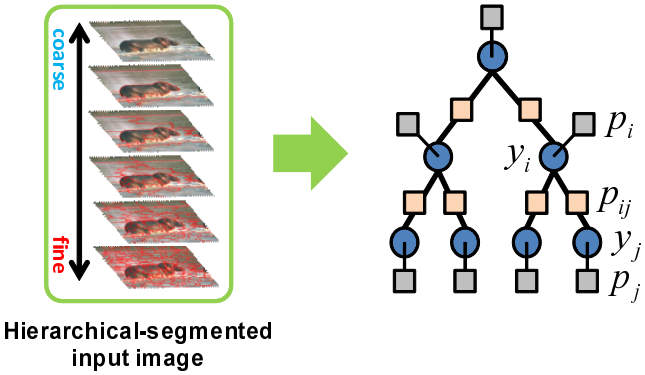


図 5　木構造条件付確率場による画像のグラフ表現

条件付確率場 [3] とは，元々，自然言語処理の分野で提案された，識別的なグラフィカルモデルである．これは，観測データに基づいて，構造を持ったデータのクラス（ラベル）を推定するために有効なモデルである．このモデルを階層的領域分割を行った画像に適用すると，各局所領域は頂点に，層間に関係がある局所領域の間は辺によって結ばれる．そのため，画像は図 5 のようなグラフ表現として表される．我々は，グラフ構造がループのない木構造であることから，このモデルを木構造条件付確率場と呼ぶ．

$i \in N$ は，階層的領域分割が行われた画像の各局所領域，$\pi(i)$ は，ある親の頂点 $i$ に対する子の頂点の集合，$\mathbf{X} = \{\mathbf{x}_i\}_{i \in N}$ は，Gentle Adaboost によって算出された，各局所領域におけるクラス信頼度の分布，そして，$\mathbf{y} = \{y_i\}_{i \in N}$ は各頂点において推定されたクラスを表す．このとき，木構造条件付確率場のモデル式は，以下の条件付分布 $P(\mathbf{y}|\mathbf{X};\boldsymbol{\theta})$ によって記述される．

$$P(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}) = \frac{1}{Z}\exp\left\{\sum_{i \in N} p_i(y_i|\mathbf{x}_i;\boldsymbol{\alpha}) + \sum_{i \in N}\sum_{j \in \pi(i)} p_{ij}(y_i, y_j;\boldsymbol{\beta})\right\} \quad (2)$$

ここで，$Z$ は正規化のための分配関数と呼ばれる．$\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ は，木構造条件付確率場のモデルパラメータであり，正解クラスの付与された学習画像全てを用いて，以下のような最大事後確率推定によって決められる．

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}}\left\{\sum_{t=1}^{T} \log P(\mathbf{y}^t|\mathbf{X}^t;\boldsymbol{\theta}) - \frac{R}{2}||\boldsymbol{\theta}||^2\right\} \quad (3)$$

ここで，$T$ は学習画像の枚数，$R$ は過学習を防ぐためのパラメータである．$\boldsymbol{\theta}^*$ は，L-BFGS 法 [11] によって，解析的に求められる．

そして，$p_i(y_i|\mathbf{x}_i;\boldsymbol{\alpha})$ は，Gentle Adaboost の出力に基づく，各頂点におけるクラス信頼度分布である．また，$p_{ij}(y_i, y_j;\boldsymbol{\beta})$ は，エッジで結ばれた頂点間のクラス共起情報である．最終的なクラス推定において，我々は式 (2) に示された条件付分布を最大にする，つまり，グラフ全体で最適になるように頂点間のクラス共起まで考慮した上で，各頂点へのクラスの割り当てを推定する．

この目的のために，我々は最大事後周辺確率推定を用いる．

$$y_i^* = \arg\max_{y_i} \sum_{\mathbf{y} \setminus y_i} P(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}) \quad (4)$$

ここで，$y_i^*$ は，事後周辺分布を最大にするクラスである．グラフ構造が木構造であるので，大域最適な推定が確率伝搬法 [7] によって可能である．

領域分割時の誤りを減らすため，我々は階層の最下層における認識結果を，最終的な認識結果と見なす．この認識結果は，階層の全ての層における認識結果を考慮した上での結果となる．その上，この認識結果は大域最適が保証されており，提案手法は物体のスケール変化に頑健であるといえる．

## 3. 評 価 実 験

### 3.1 実 験 概 要

実験のための画像データセットとして，Corel 7 Dataset を用いた．これは，7 クラス 100 枚の画像データセットで，画像のサイズは 180 × 120 画素である．各画像には，正解クラスが画素単位で与えられている．

認識率は，認識対象のクラスごとの認識率を平均した，クラス平均認識率を用いる．我々は，クロスバリデーション法（一つ抜き法）で学習とテストを行った．また，階層的領域分割において，我々は階層の最下層における局所領域の数を 200 個とした．これは，super-pixel 表現 [12] と呼ばれる．そして，階層の層数を 6 とし，また，BoF における Visual Words の数は 500 単語とした．各階層における局所領域の数は，下層から順に，およそ 200,100,50,25,12,1 個（最上層は画像全体とする）となっている．

ここで，提案手法の有無による認識精度の変化を調べた．

### 3.2 実験結果と考察

表 1 から，我々の提案手法によって，認識率が 2.2%改善されたことがわかる．ここで，表中の NH とは，No Hierarchization の略で，階層化を行っていない従来手法 [13] のことを指す．

考察のために，いくつかの認識結果の例を図 6 に示

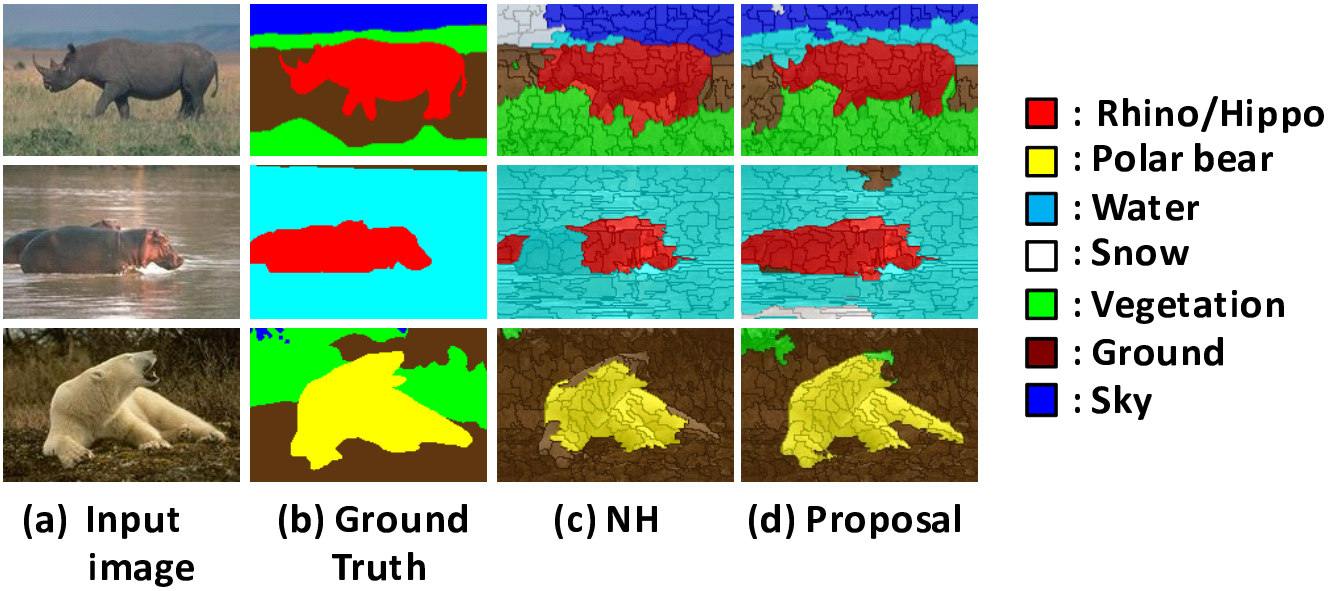


図 6　認識結果の例

す．従来手法である (c) と比較すると，提案手法である (d) は，特に，異なるクラス間の境界付近の誤認識が改善されている．これは，提案手法が，階層的領域分割に基づく木構造条件付確率場を構築することにより，複数スケールを考慮しているためだと考えられる．

表 1　認 識 結 果

| Class | Rhino | Polar bear | Water | Snow |
|---|---|---|---|---|
| NH [13] | 71.8% | 71.0% | 82.6% | 70.6% |
| **Proposal** | 75.7% | 72.7% | 78.9% | 73.8% |
| Class | Vegetation | Ground | Sky | **Average** |
| NH [13] | 78.9% | 74.7% | 41.7% | 70.2% |
| **Proposal** | 79.4% | 76.5% | 49.6% | 72.4% |

## 4. ま と め

本論文では，我々は，階層的領域分割法である Segmentation by Weighted Aggregation に基づく階層的構造を，条件付確率場による一般物体認識のフレームワークに導入した新しい手法を提案した．我々は，これを木構造条件付確率場と呼んだ．階層化による複数スケールの考慮により，認識結果はスケールの変化に頑健となった．結果として，認識精度が 2.2%向上した．今後の方針としては，クラス共起情報以外の有用なコンテキスト情報や，3 次元情報といった新しい特徴について検討していく予定である．

## 文　献

[1] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, “Visual categorization with bags of keypoints, ” Proc. ECCV Workshop on Statistical Learning in Computer Vision, pp. 1-22, 2004.
[2] K. Barnard, and D. Forsyth, “Learning the Semantics of Words and Pictures, ” Proc. IEEE International Conference on Computer Vision, pp. 408–415, 2001.
[3] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data, ” Proc. International Conference on Machine Learning, pp. 282-289, 2001.
[4] J. Shotton, J. Winn, C. Rother, and A. Criminisi, “Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation, ” Proc. IEEE European Conference on Computer Vision, pp. 1-15, 2006.
[5] S. Gould, J. Rodgers, D. Cohen, G. Elidan and D. Koller, “Multi-Class segmentation with relative location prior, ” International Journal of Computer Vision, pp. 300-316, 2008.
[6] E. Sharon, A. Brandt, and R. Basri, “Fast multiscale image segmentation, ” Proc. IEEE Computer Vision and Pattern Recognition, pp. 70-77, 2000.
[7] C. M. Bishop, “Pattern Recognition and Machine Learning, ” Springer, Chapter. 8, 2006.
[8] J. Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: a statistical view of boosting, ” Technical Report, 1998.
[9] A. Brandt, S. McCormick, and J. Ruge, “Algebraic multigrid (AMG) for automatic multigrid solution with application to geodetic computations, ” Inst. for Computational Studies, POB 1852, Fort Colins, Colorado, 1982.
[10] D. G. Lowe, “Object recognition from local scale-invariant features, ” Proc. IEEE International Conference on Computer Vision, pp. 1150-1157, 1999.
[11] J. Nocedal, “Updating Quasi-Newton Matrices With Limited Storage, ” Mathematics of Computation, pp. 773-782, 1980.
[12] X. Ren, J. Malik, “Learning a classification model for segmentation, ” Proc. IEEE International Conference on Computer Vision, pp. 10-17, 2003.
[13] 奥村健志，滝口哲也，有木康雄，“大域的特徴として BoF を導入した CRF による一般物体認識，” 画像の認識・理解シンポジウム，MIRU2009, OS4-2, pp. 95-102, 2009.